SUPPORTING NEW PRODUCT DEVELOPMENT USING CUSTOMERS' ONLINE DATA AND COMPUTATIONAL INTELLIGENCE METHODS

HANAN YAKUBU

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Industrial and Systems Engineering

SUPPORTING NEW PRODUCT DEVELOPMENT USING CUSTOMERS' ONLINE DATA AND COMPUTATIONAL INTELLIGENCE METHODS

Hanan Yakubu

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

(Signed)

Hanan Yakubu (Name of student)

Abstract

In recent times, the availability, and the proliferation of the generation of online data derived from social media and e-commerce platforms have been capitalised by firms in different ways to influence, promote, enhance, and develop new products. In the area of the new product development (NPD) process, online data can be applied differently in each stage of the NPD process. Within the NPD process, enhancing customer satisfaction and making product demand forecasts are two areas that require the extensive use of data. Previously, to obtain data for NPD, surveys were mostly conducted by product manufacturers to seek information’s from customers before designing a new or improving a new product. However, the nature of conducting surveys tends to be cumbersome and respondents can easily misinterpret the questionnaires. Surveys also have the limitation of being incomplete and, usually, the ratings used in surveys do not convey the real needs of respondents. Thus, developing customer satisfaction models from surveys presents many complexities since customers’ responses' fuzziness is usually not considered. Similarly, the identification and predicting of the most important product attributes have not been explored in past studies in addressing the dynamic needs of consumers. This is due to the over reliance on surveys that fails to provide reliable data for manufacturers, thus preventing them from producing products that meet the rapid changes in customer needs due to technological advancement.

As part of product development activities, the demand for products is usually forecasted to prevent revenue loss. However, most of these forecasts require large amount of historical data to develop a demand forecast model. With the advent of the internet, manufacturers can integrate constantly updated user generated online data in forecasting models in order to forecast the adoption of products. To overcome the above limitations, the objectives of this research are presented in three phases: i) To propose a novel customers satisfaction model that address the fuzziness and nonlinearity of customer satisfaction models using multigene genetic programming based fuzzy regression (MGGP-FR) ii) To formulate a methodology for determining and predicting the importance of product attributes. The Shapely Value and Choquet integral are employed to estimate the importance of product attributes and based on the importance values, a fuzzy rough set times series method is proposed to forecast the future importance of product attributes. iii) To propose a new market share model and demand forecasting model that addresses uncertainties in forecasting. A market share model is developed from the multinomial logit (MNL) model and the fuzzy regression (FR) approach while the demand model is developed from a modified Bass model integrated with sentiment scores from online reviews.

A case study on modelling customer satisfaction for electronic hairdryers using MGGP-FR is presented in this study. To validate the proposed methodology, the results of the MGGP-FR are compared with previously proposed methods mainly FR, genetic programming (GP), and genetic programming-based fuzzy regression (GP-FR). Based on the mean relative errors and the variance of errors of the MGGP-FR and previous methods, the proposed MGGP-FR showed a better performance when compared with the previous methods. Next, forecasting the future importance of the product attributes of an electronic hairdryer is illustrated using the fuzzy rough set time series method. The proposed fuzzy rough set time series forecasting accuracy outperformed the fuzzy time series method. Lastly, a case study on forecasting the adoption of a Tablet P.C is used to illustrate the applicability of the proposed fuzzy modelling and discrete choice analysis method for forecasting product adoption using online reviews. The proposed method was compared with the fuzzy time series forecasting and the original Bass model and was found to be better as it provided different scenarios for the forecast and acceptable forecasting results.

Publications

Journal papers

Hanan Yakubu Hanan Yakubu*, C.K.M. Lee, C.K. Kwong A fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews. *Complex and Intelligent Systems*(submitted)

Hanan Yakubu C.K. Kwong Forecasting the importance of product attributes using online customer reviews and google trends, *Technological Forecasting and Social Change,* (2021)

Yakubu, H., Kwong, C.K. & Lee, C.K.M. A multigene genetic programming-based fuzzy regression approach for modelling customer satisfaction based on online reviews. *Soft Comput* (2021)

Conference papers.

H. Yakubu and C. K. Kwong, "Multigene Genetic Programming Based Fuzzy Regression for Modelling Customer Satisfaction Based on Online Reviews," *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Macao, China, 2019, pp. 1541-1545

H. Yakubu and C. K. Kwong, "Using Online Big Data for Determining the Importance of Product Attributes," *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore, Singapore, 2020, pp. 691-695

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my former supervisor Dr C.K. Kwong who guided me through the initial stages of my studies. His patience, kindness and understanding encouraged me to do better in the initial stages of my studies. I appreciate his constructive criticism, comments and suggestions that enabled me to put together sound research. Without his encouragement, I would not have made it this far.

My second appreciation goes to my supervisor Dr C.K.M Lee who took the responsibility from my former supervisor and gave me guidance and different perspectives for my studies that enabled me to approach some aspects of my research with a fresh approach and continue my research to the end. I appreciate her useful suggestions, comments, and constructive criticism.

I would also like to thank my mother, my brother and my friends who supported me on this intellectual journey. I appreciate their love, support, and prayers.

Finally, I wish to express my sincere gratitude to my lovely husband Abdul-Manan Iddrisu, for encouraging me to undertake this journey in the first place. I appreciate his support and prayers. Without him, my journey would certainly have been harder.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANFIS	Adaptive Fuzzy Inference System
ANN	Artificial Neural Network
ANP	Analytical Network Process
BoW	Bag of Words
C-Kano	Creativity based Kano Model
DCA	Discrete Choice Analysis
DENFIS:	Dynamic Evolving Neural-Fuzzy Inference System
FAHP:	Fuzzy Analytical Hierarchy Process
FIS	Fuzzy Inference System
FLR	Fuzzy Logical Relations
FQFD	Quality Function Deployment
GA:	Genetic Algorithm
GP:	Genetic Programming
GP-FR:	Genetic Programming based Fuzzy Regression
HMM	Hidden Markov Model
HOQ	House of Quality
IGA	Interactive Genetic Algorithms
LEM2	Learning from Examples version 2
MCDM	Multi-Criteria Decision Making
MGGP	Multigene Genetic Programming

MGGP-FR: Multigene Genetic Programming based Fuzzy Regression

ML Machine learning

NLP Use Natural Language Processing

PCA Principal component analysis

POS Part-Of-Speech

QFD Quality Function Deployment

SCAMPER Substitute, Combine, Adapt, Modify, Put to another use, Eliminate, Reverse

SVM Support Vector Machine

TRIZ Theory of Inventive Problem Solving

Chapter 1 Introduction

As the world of engineering and manufacturing is rapidly evolving, product manufacturers are increasing their efforts to continue to enhance their products and services to improve customer satisfaction and beat their competitors in the global market. In the process of improving the overall customer experience, data analytics has been adopted in the current era of the digital revolution to enhance product design and services. It is predicted that by 2025, the global data will reach 175 zettabytes with 51% of the data expected to be in data centres and 49% is expected to be in the public cloud (Tom Coughlin, 2018). With such a massive increase in data, such growth comes along with an increase in the demands of customers who turn to improved products and services at a faster rate. This can be challenging for product manufacturers who have to work with less resources and within a limited budget.

Thus, most product manufacturers are shifting from product-centric based designs to customer-centric based product designs whiles leveraging the benefit of data generated from the digital revolution. It is imperative for product manufacturers to consider this shift by understanding human the behaviour generated from consumers' user experience in digital mediums, companies can understand the context of consumers needs and evolve with consumers to create unique ideas for their product. Sabir (2020) examined the impact of product design dimensions on customer satisfaction. A survey was conducted to confirm the relationship between the symbolic dimension and satisfaction. The results confirmed how affective responses indicted and influenced customer satisfaction.

Anderson et al. (2004) developed a framework to show how customer satisfaction in products and services affects future customer behaviour and the risk of future cash flows. The authors integrated three theoretical components namely i) the arguments that customer satisfaction influences customer behaviour, ii) the new linkage between customer satisfaction and the bargaining power of the firm and finally, iii) how customer behaviour and the firm's bargaining influence shareholder value. Their study empirically showed how important it is for businesses to satisfy the needs of consumers in order to remain competitive.

The study shows the relevance of how customer satisfaction affected the long-term financial performance of a company. Ye et al. (2014) also conducted a study that showed how customer satisfaction was is one of the most important critical factors in NPD. On the 2nd of August 2016, Samsung launched, its flagship smartphone Galaxy Note 7. It was announced as the “best smartphone money can buy”. However, on the 24th of August 2016, the first explosion of the Samsung Note 7 was reported in South Korea. This was followed by a series of complaints all over the world about the explosion of the Galaxy Note 7. Eventually Samsung had to halt the production and sales of the new smartphone due to safety concerns (Yun et al., 2018).

Nowadays, consumers are adopting ways to convey their experiences through social media, e-commerce platform, blogs etc. Data generated from these digital mediums are mainly comments or reviews. Other activities such as clicks on websites, search data generated from a search engine, data from connected devices and purchasing habits all generate data encompassing the volume, variety and velocity of the data generated from digital space. By employing big data analytics methods, product manufacturers can mine and analyse data in real-time. Hidden patterns can also be extracted from consumer reviews and purchasing decisions. Also, customers find value in diversified products that have the potential to satisfy certain needs at any point in time. Thus, the determination of key aspects of a product that can satisfy different market segments is a critical area during market research that requires industries and businesses to pay close attention to. Moreover, the decisions on determining the target aspects of products required by consumers are usually considered in the early stages of product design in the new product development (NPD) process. All these efforts are made in an attempt to make an informed decision before a product prototype is developed.

Product attributes distinguish one product from another, and these product attributes influence consumers purchasing decisions. Thus, the relevance of identifying critical and the most important customers for a product cannot be underestimated. In identifying which product attributes consumers need the most in a product, product manufacturers use surveys and traditional assessment like the Likert scale to represent customers perception to rank consumer’s needs. Also, by identifying the most important product attributes, manufacturers can be able to make reasonable decisions on the best engineering characteristics in designing the product. Li et al., (2019) stated that the rise of usage of the internet among consumers lead to an increase in the expression of consumers requirements for a product. These large amounts of customer requirements expressed on online platforms in the form of reviews usually lack the guidance of purposive questions as established in the traditional method of conducting surveys. This makes it challenging to identify critical engineering characteristics. Moreover, with many attributes to consider during the development of a product, manufacturers faced the problems of identifying the most relevant attributes that will meet the demands and satisfaction of consumers with less resources.

Meanwhile, the past product attributes may not be relevant for today or future products whereas some product attributes from the past still exist in current products. This dynamic process of adding, removing, and chaining product attributes stems from the dynamic needs of consumers’ time. Developing products with little or slow improvement that does not keep up with the dynamic needs of customers can be detrimental to the success of the product as seen in the cellular phone market. It is also known that the cellular phone market is characterized by short product life cycles (Chong and Chen, 2010). Moreover, the time for developing the product could be longer and as such, by the time the product is launched consumers would have changed their needs regarding the product. For such products with time-based customer requirement, such as fashionable items, a longer time to launch the product could lead to loss of revenue when the market for the product is not critically assessed.

1.2 Problem statement

The advent of web 3.0 has provided numerous opportunities for product manufacturers to integrate their product manufacturing process through cloud computing, especially during the early phase of new product development. This has also facilitated many studies in the existing literature to investigate and pursue research on how businesses and product manufacturers can integrate the advantages the cyberspace has to offer in their daily operations. These studies however do have some gaps that need to be addressed to add to the growing works in the existing literature. This problem and gaps in the existing literature addressed by this thesis include:

1. There is lack of integration of data from the web for new product designs during the early phases of new product development. During the early phases of new product development, idea generation is considered to be an integral part of product design. The idea generation phase comprises identification of all the necessary and possible factors needed for the development of a new product, such as the product attributes, the

functionality of the product, the design of the product etc. that will meet the demand of the consumers. However, for most existing practices in the idea generation phase of new product development, there is an over-reliance on the conducting of surveys in order to identify the needs of potential consumers or the market. The problem of conducting surveys is that they can be time consuming and cumbersome. Surveys require predefined questionnaires and it may also make many assumptions about the needs of customers (Suef et al., 2017). Also, the findings of surveys conducted today may be irrelevant for future products as customers' needs are dynamic and what they claim to want today may be different from what they want tomorrow. Moreover, the accuracy of a survey can be compromised since the sampling of the respondents, the response to the questionnaires, the inability to have the real response of respondents can lead to poor survey results that do not represent the needs of the market. To address this problem, data from the internet needs to be leveraged and integrated into new product designs.

2. There is less research to develop customer satisfaction models that address the fuzziness and non-linearity in customer needs. The customer satisfaction model tends to be mathematical that can relate the product design attributes to the customer's satisfaction. However, product design attributes in the customer satisfaction model have the potential to be highly nonlinear due to the complex relationships that are developed to explain

how consumers are affected by product attributes. Also, for most existing studies on customer satisfaction models, the conventional survey has used to identify the product design needs of consumers to generate these models. Survey responses from respondents do not address the fuzziness in the replies. These responses, which are highly subjective, can be highly inconsistent among all the respondents. Moreover, most surveys are conducted to develop customer satisfaction models with the rating scale such as the Likert scale used to capture subjective opinion with numbers (Huang et al., 2019). The rating methods do not capture the opinions of consumers since these ratings do not address the fuzziness in the opinions of respondents. Thus, insufficient studies have been made to addresses the fuzziness and nonlinearity needed to develop robust customer satisfaction models.

3. In the existing literature regarding product design, the determination of relevant product design attributes for new product development has been made through various multi-criteria decision making (MCDM) so as to come up with the best product that will meet the needs of the market (Galetto et al., 2018; Wang, 2020). However, customer needs are dynamic and constantly changing. Thus, there is a need to establish frameworks that can assist product manufacturers to predict the needs of customers without wasting resources on product designs that cannot meet the dynamic needs of consumers. Modern products that rely on technology usually have

shorter lifecycles due to rapid changes in technology. Lastly, product demand forecasting methods have been explored in the existing literature with fewer studies with not much research focus on a new product that could be forecasted with little available data. The existing literature has heavily relied on forecasting models with historical datasets. The existing forecasting methods have not leveraged the advantages that data from the internet has to offer to forecast the demand of the product. Moreover, fuzziness and uncertainty exist in the forecast and have received less attention in demand forecasting. This research intends to formulate different forecast strategies in different scenarios that consider the uncertainty in products demand in the era of rapid technological advancements.

1.3 Research aim and objectives.

This research aims to develop frameworks to integrate data from the internet during the product design and development stage. A novel computational intelligence technique is presented to develop customer satisfaction models that relate product attributes to customer satisfaction. We also aim to present a methodology to determine and predict the future importance of product attributes to reduce the uncertainty in the acceptance of a new product by customers. Lastly, we aim to develop new product demand forecasting model that considers big data is presented to allow product manufacturers to make a more reliable forecast. The objectives of this research are as follows;

1. To develop a customer satisfaction model that will relate product attributes to customers' satisfaction using a MGGP-FR.
2. To develop a methodology for determining the importance and future importance of product design attributes using a Shapley value, Choquet integral and a fuzzy rough set time series method.
3. To develop a fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews.
4. To evaluate the effectiveness of the proposed methods by comparing them using a case study.

1.4 Research scope and significance

This research focuses on the formulation of computational intelligence methods and methodologies that will integrate customers' online data in the early phases of product development and product design. This study will contribute to the existing body of knowledge by addressing the research gaps in existing academic studies. This study will also provide insight for the researchers and practitioners in the R&D department of manufacturing firms in considering and leveraging the proposed methods and integrated them into their businesses development plans.

1.5 Structure of the thesis

The structure of the thesis is outlined as follows:

- Chapter 2 presents a comprehensive literature review on existing methods that integrates customers' needs in product design and product development. Existing studies on the methods of computational intelligence techniques for modelling customer satisfaction are presented in this chapter. Previous Studies on competitive intelligence using online reviews and google trends for determining the importance of product attributes and the forecasting of product demand are presented in this chapter as well. Lastly, relevant previous studies on product demand forecasting models will also be presented.
- Chapter 3 describes the methodology for modelling customer satisfaction using a MGG-PR. This chapter also describes the methodology for determining and predicting the importance of product attributes. Also, the framework for forecasting the adoption of the product demand is described.
- Chapter 4 presents the theoretical background and implementation of MGGP-FR for modelling customer satisfaction. The effectiveness of the proposed method is evaluated, and validation tests are described. The results of the validation tests are compared with other methods and discussed.
- Chapter 5 describes the methodology and implementation for determining the importance and future importance of product attributes using the Shapley value, Choquet integral and the fuzzy rough set time

series method. The effectiveness of the proposed method is evaluated by comparing with existing methods.

- Chapter 6 presents the methodology and implementation for fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews. The Bass model in different scenarios is evaluated for its effectiveness and compared with existing methods.
- Chapter 7 concludes the thesis by presenting a summary of the study and the academic contributions. Also, the limitation of this study and future works are presented in this chapter.

Chapter 2 Literature review

In this chapter a literature review on modelling customer satisfaction, competitive intelligence for product design with data and demand forecasting methods is presented. Section 2.1 describes the process of the literature review. Section 2.2 presents a literature review on consumer research in product development and in section 2.3, the extant study on integrating customers' needs in product design is described. Section 2.3.1 gives an account of past studies on methods of computational intelligence techniques for modelling customer satisfaction. Section 2.3.2 outlines the methods for the determination of the importance of customer needs product design. Natural language processing and opinion mining is reviewed in section 2.4. Google Trends is reviewed and Google Trends for forecasting is described in section 2.5 and 2.5.1 respectively. Demand forecasting methods are reviewed in section 2.6. In section 2.6.1, the statistical approach to forecasting is described and artificial intelligence methods for demand forecasting are described in section 2.6.2. The Bass model is described in 2.7 and the Discrete choice models and their application are reviewed in Section 2.8. Finally, in section 2.9, the summary of the literature and the research gaps are discussed. A summary of the literature review is shown in in Figure 2.1.

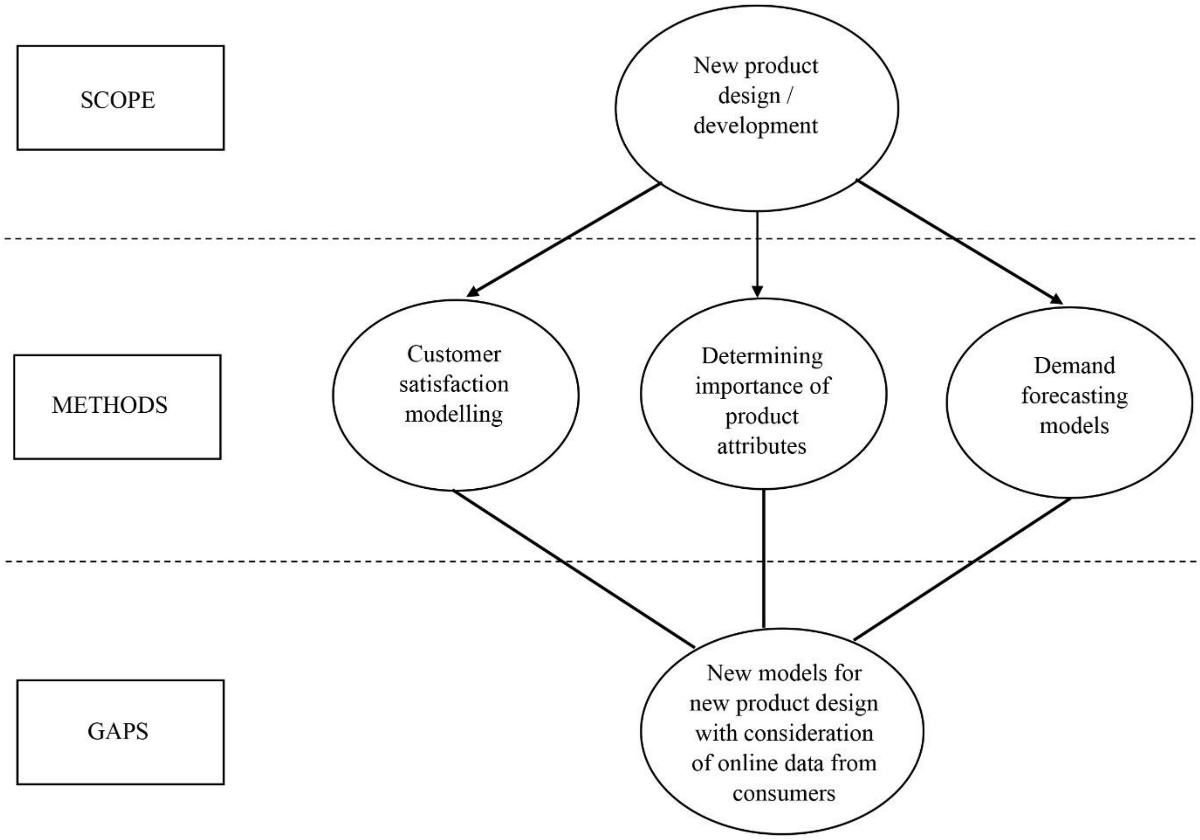

Figure 0.1 A summary of the literature review

2.1 Structure of literature review process

The literature review was categorised into two areas: product design/ development and demand forecasting. Research papers related to customer satisfaction modelling/ product design and demand forecasting were identified based on the concepts, methods, and models were extracted and analysed. The academic papers were obtained from the following online database: Scopus web of science, Google scholar, IEEE explore, and Springer link. The structure for conducting the process of the literature review is shown in Figure 2.2

With the digital databases chosen for this study, the major keywords and alternative words based on the two areas of this studies were used as search strings. The Boolean operators 'AND' and 'OR' were used to develop search strings for use in the digital database. Table 2.1 shows the major keywords and the associated alternative keywords used in the search process.

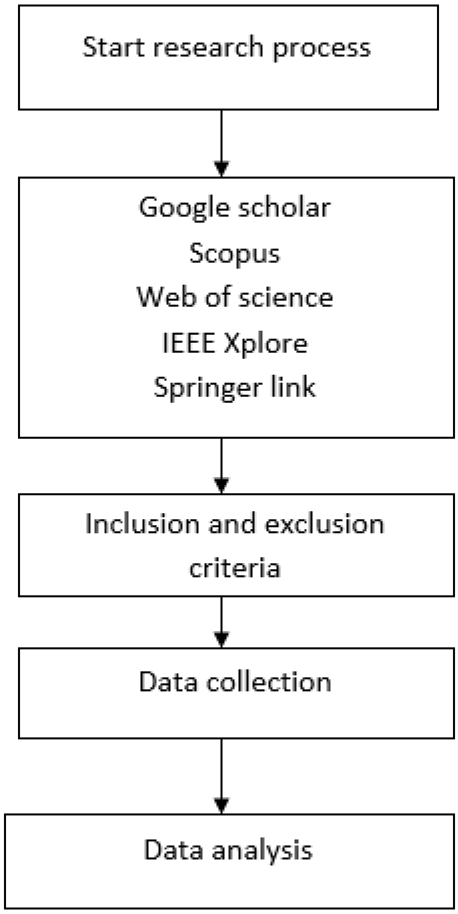

Figure 0.2 Literature review process

Table 0.1 Major search words and alternative words

Major search words	Alternative search words
Product design	Product development, Product model, product attributes, product characteristics
Customer satisfaction modelling	Customer preference modelling, Customer requirements modelling
Demand forecasting	Sales forecasting, Demand modelling, Sales prediction, Product adoption forecast

For select relevant publications, a quality assessment score was defined and developed to select relevant papers in this study. Table 2.2 shows the quality assessment checklist for the articles obtained from the databases by following the protocol describe in (Khan et al., 2013). Over eighty papers were analysed during the literature review process.

Table 0.2 Quality assessment checklist

QA questions	score
Is scope of study clearly described	1
Does the study explicitly focus on product design	0.5
Does the study explicitly focus on customer satisfaction modelling?	0.5
Does the study explicitly focus on demand forecasting?	0.5

2.2 Consumer research in product development process

The initial phase of new product development is fraught with uncertainties and fuzziness as it is difficult to determine what exactly customer will need or what will meet consumer needs. Surveys are usually used to seek consumers' needs. However, it has been argued that the approach in asking what consumers want is fruitless as, for the most part, most consumers don’t have any idea on what they want (Ulwick, 2002). To avoid the failure of new products, it is still relevant to understand how consumers perceive a product even though seeking their views on what they want can be challenging. This is because this initial stage for developing a product tends be crucial in identifying potential products that might be successful on the market (Kohli and Jaworski, 1990; Rochford, 1991). Many new product development processes exist however a large number of these techniques are not adopted in companies (Nijssen and Lieshout, 1995; Nijssen and Frambach, 2000). Focus groups and surveys and the like have been the dominant characteristics in most of the research on new product development. This might account for the higher failure rates of new product (Wind and Mahajan, 1997). In Table 2.3, some works on customer satisfaction in new product development are shown. Similarly, some techniques used in customer satisfaction modelling are given in Table 2.4 and, in Table 2.5, the literature on identification of important product attributes are summarized.

Table 0.3 Overview of some of the works of customer satisfaction in new product development

Author	Problem	Method	Finding	Limitation
(Galetto, et al, 2018)	In Independent scoring method cardinal properties are arbitrarily attributed to data collected on ordinal scales ordinal scales. There is always a debatable ordinal to cardinal conversion.	Decision support method (QFD)	Prioritised customer needs and translated them into engineering characteristics	Did not address the fuzziness in customers responses
(Kano et al., 1984)	Inability to retain customers due to poor methods for identifying customers' needs	Decision support Method (Kano model)	Developed a model to categorise product attributes and how they affect customer satisfaction	No explicit models were developed and the fuzziness were not addressed.
(Xi et al., 2020)	The original kano model and subject and qualitative.	Decision support Method (Kano model) Fuzzy theory	A more comprehensive Kano model was developed. Fuzziness in the Kano model was also addressed.	No explicit customer satisfaction model was developed.
(Barone et al., 2007)	Lack of integration of customer emotions and perceptions in product design.	Decision support Method (Kano model) Conjoint analysis	The proposed method was able to identify product attributes that could induce specific emotions and feelings in customers.	No explicit models were developed and the fuzziness were not addressed
(Yang et al., 2019)	QFD lack innovative tools for back up design cases.	QFD, theory of inventive problem solving, Kansei engineering	Important design zones and critical innovative points were identified in from the house of quality OF QFD	Relied on surveys Did not establish an explicit model
(Kim and Yoo, 2018)	Difficulty extracting delighter in products and services from the Kano model.	Data mining and decision support method (Kano model)	The method extracted delighter from big data	Limited data sources, no explicit model and no fuzziness addressed.

(Raharjo et al., 2008)	The problem of subjectivity in early product design.	QFD and ANP	Reduced human judgement error. Adaptable to constantly changing environment	No explicit developed model Fuzziness not addressed
(Wu and Liao, 2021)	QFD lack innovative product design approach	QFD, TRIZ AND Fuzzy theory and Statistical approach	Fuzziness in customer preferences addressed Introduced an ergonomic product innovative design and evaluation design stage Reduced time to decision making for new product design	No explicit model addressed
(Wang and Wu, 2014)	Highly diverse market results in poor market segmentation	Conjoint analysis, Kano model and VIKOR	Recognized important product features to crate ad hoc market segment and systematic approach to prioritize product varieties	Fuzziness not addressed, and no explicit models developed

Table 0.4 Overview of some of the techniques for modelling customer satisfaction

Author	Problem	Method	Finding	Limitation
(Kwong et al., 2010)	Fuzziness and randomness addressed independently in in modelling relationships in in QFD.	Fuzzy least square regression	Fuzziness and nonlinearity addressed in modelling relationships in QFD	Relied on surveys
(Jiang et al.,2019)	Customer surveys do not adapt to rapidly to changing customer preferences.	Dynamic evolving neuro fuzzy inference system	Fuzziness on customer preferences were addressed	The models developed were not explicit only first order linear models were developed
(Fang et al., 2020)	Uncertainty in multisegmented market not addressed in QFD	Fuzzy chanced constrained programming	The proposed method was developed for multisegmented market fuzziness was addressed	Relied on surveys

(Wang and Zhou, 2020)	Customer preferences are usually ambiguous	Genetic algorithm fuzzy Kano model fuzzy AHP	The proposed method could determine the customer needs from images.	Fuzziness not addressed
(Biet al., 2019)	Limitations of surveys determining the customer satisfaction	Support vector machine Ensemble neural network Kano model	Customer requirements were identified from the online review and the proposed model. No prior assumption about the distribution of the customer satisfaction.	Fuzziness was not addressed, customer segments not considered. No explicit model developed
(Li et al., 2018)	Surveys tend to be time consuming. Difficulty in identification of affective needs	Kansei engineering machine learning	The proposed method established a relationship between design elements to affective responses	Fuzziness was not addressed
(Nazari-Shirkouhi and Keramati, 2017)	Many fuzzy regression (FR) approaches for modelling customer satisfaction exist. Marketing mix has not been considered in customer satisfaction modelling.	FR- data envelopment analysis algorithm	FR model that considers marketing mix was developed.	Relied on surveys
Jiang et al., (2013).	Customer satisfaction models did not capture fuzziness and nonlinear relations between customers satisfaction and design attributes.	Chaos based fuzzy regression	Chaos based fuzzy regression outperformed fuzzy least squares regression, statistical regression, and FR	Relied on surveys
Cherif et al., (2010).	Impreciseness in QFD	Goal programming and QFD	Fuzziness addresses in QFD	Relied on surveys

Table 0.5 Overview of some of the literature on determining the importance of product attributes

Author	Problem	Method	Finding	Limitation
Liu S.-T (2005)	The membership functions for fuzzy weighted average are not known in prioritization of design requirement	Fuzzy weighted average method nonlinear programming	Fuzziness in ranking address	Relied on surveys
Yan et al.(2013)	Vagueness and impreciseness in QFD Approximation in fuzzy linguistic approach causes loss of information.	Fuzzy theory	The proposed method quantifies qualitative assessment of customer preferences.	The dynamic preferences of customers were not address the proposed method relied on surveys
Battistoni et al., (2013)	Inconsistencies in using surveys for prioritising product attributes	AHP	Prioritised product attributes using AHP	Does not address fuzziness. Relied on surveys Does not address the dynamic needs of customers
Wang et al.,(2014)	Fuzzy QFD limited due to inability to perform multiplication of fuzzy numbers	Fuzzy QFD	Multiplication of fuzzy numbers were replaced with an alternative	The dynamic prioritised customers' needs not addressed.
Wang, and Tseng (2011)	Long list of product attributes affect customers assessment ability during product ranking	Shapley value	The process reduced product configuration process to avoid trivial selection of attributes.	The dynamic needs of customers were not addressed needs not addressed
Jia et al. (2016)	Uncertainty and fuzziness associated with evaluation process of customer requirements	Fuzzy measure and fuzzy discrete Choquet integral	The interaction among engineering attribute of a product were addressed	The dynamic customer requirements were not addressed relied on surveys

Wang et al. (2020)	High priori information requirement is needed to address the subjectivity and randomness in in linguistics evaluations	Grey relational analysis Cloud computing	Small sample size of product attributes evaluations were addressed The cloud model addressed randomness and fuzziness technical product attributes were	The dynamic product attributes requirements were not addressed the correlations among customers' requirements were ignored
Ying et al. (2018	Uncertainty in selecting attributes for new product development concept	Simple additive weighing. Hybrid information. Multiple attribute decision making and linguistic term.	The concepts for the new product development were ranked based on the proposed method. Fuzziness in the linguistics assessment was addressed.	The trends of the new product concept was not addressed
Li et al. (2018)	Limitation of identifying and understanding customer requirements	Interval linguistic weighted arithmetic averaging operator	Vagueness addressed	Relied on survey dynamic customer requirements not addressed

2.3 Integrating Customer's needs in product design.

As globalization tends to increase competition in the marketplace, product manufacturers are increasingly making efforts to develop products, services, and systems to satisfy the needs of their customers. Understanding the needs of customers and fulfilling those needs tends to drive customer willingness to purchase a product from a manufacturer. To enhance customer satisfaction, product attributes and their associated engineering characteristics are researched not only to generate a product of higher quality but also to search for product attributes that address the need of consumers (Galetto et al., 2018).

One of the main methods adopted by product manufacturers to elicit customer needs and integrate them in product design is the use of the Kano model. In 1984, Professor Noriaki Kano proposed a theory on product development and customer satisfaction with a model known subsequently as the Kano model. The Kano model presents categories of product attributes according to how consumers perceive them. The categories of the product attributes according to consumers perception proposed were, ‘one-dimensional attributes’, ‘attractive attributes’, ‘must-be attributes’, ‘indifferent attributes’ and ‘reverse attributes’. Thus, an opportunity to improve and select product attributes during product development could be identified and enhanced with the Kano model (Kano et al.,1984). The Kano model, after its development, has been used by many businesses to save time and resources in choosing the best product attributes. The model was also further developed from interviews with target customers, focus groups and other indirect methods such as the Delphi or the expert brainstorming approach (Xi et al., 2020). The questionnaires used in the Kano model enables researchers to categorise customer’s needs. The categories show the level of customers’ satisfaction according to how the needs of the customer are met. The level of customer satisfaction then indicates the category of a product attribute. The Kano model’s underlying assumption is that high-quality products tend to satisfy all customer requirements. The categories of the Kano model in a detailed description are explained as follows; the ‘Must-be attributes’: the must-be attributes refer to the aspects of products that are a necessity for the product to function. These attributes are not mentioned as requirements by consumers and meeting those needs does not lead to an increase in customer satisfaction. Such attributes form the basis of the product. In ‘one-dimensional attributes’, customers can explicitly state their requirements for a product. They are attributes to be included in a product to increase customers’ satisfaction, as customers’ satisfaction decreases when the attributes are omitted from the product. The term ‘attractive attributes’, refers to the aspects of the product that exponentially increases the satisfaction of customers, though usually latent. ‘Indifferent attributes’ are those that result in no change in customer satisfaction when included in the product. ‘Reverse attributes’ refer to product attributes that decrease customer satisfaction when added to the product (Barone *et al*., 2007). The categorisation of the customer requirements can be identified after a survey has been conducted for the Kano model. The customer (respondent) is made to state his/her level of satisfaction in two scenarios; when the requirements are met (functional question), and when the requirements are not met (dysfunctional question). Figure 2.3 shows the graph of the kano model.

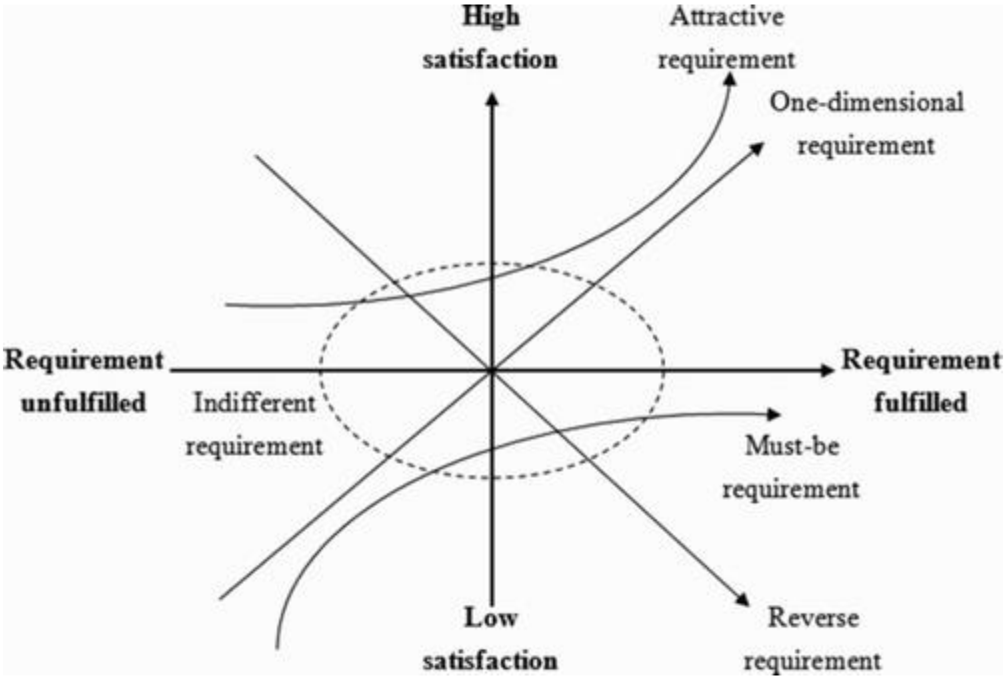

Figure 0.3 A graph of the Kano model of customer satisfaction (Kano et al., 1984)

Customer satisfaction studies have for a long time been concentrated on unravelling areas that have the potential to increase customer satisfaction (Busacca and Padula, 2005) These areas, when identified, need to have their particular products requirements quantified along with the examination of the effect on the overall customer satisfaction. The results from these processes are key for identifying what aspect of a product needs improvements, when necessary (Madzík, 2016). With the Kano model, a lack of a deeper insight into customer satisfaction modelling is encountered.

Moreover, the Kano model procedure is time-consuming and has potentially expensive traits concerning the cost involved in conducting interviews to develop the model. To address some deficiencies of the Kano model, Xu *et al*. (2009)

proposed an analytical Kano model for customer needs analysis. In the study, two approaches to support product design were presented; the development of a Kano classifier to be used as a tangible criterion for categorising customer needs and a configuration index, introduced as a decision factor for product configuration design. In another study, a Kano evaluator was used to justify the importance of the product configurations. A fuzzy questionnaire was suggested to alter the subjective two-dimensional Kano questionnaires. Moreover, the authors proposed a fuzzy statistic to help interpret the results of the Kano model. The model developed in this study provided a method to assist companies in gaining insights into different consumers concerns(Lee and Huang, 2009). Lin et al., (2017)) presented a framework to quantitatively calculate the likelihood that customers will be satisfied with a product, based on Kano classification, and examined the asymmetrical nonlinear relationship between attribute-level performance and customer satisfaction. A logistic regression method that models the probability of customer satisfaction was presented in the authors' work. Also, a hybrid framework that incorporates customers' preferences and customers' perceptions in product configuration was proposed by Wang and Wu (2014). The hybrid three-phase framework suggested comprises; (1) capturing customer preferences followed by segmentation of the market, (2) conducting conjoint analysis to determine the utilities of core attributes of a product and employing a Kano's model to extract customer perceptions of optional attributes (3) assessing product varieties based on maximizing the overall customer satisfaction. In another study, regression analysis was employed by Chen (2012) to classify the quality attributes of the Kano model; mainly the ‘must be’, ‘one-dimensional’, ‘attractive’ and ‘indifferent’ product attributes. Since the Kano model categories were considered individually during the analysis, the author further proposed a new division class known as the mixed class distribution categories. The proposed approach simplified the Kano model process of product attributes categorisation. Thus, the Kano survey involving respondents to evaluate functional and dysfunctional requirements was eliminated by developing a robust and efficient way of classification. Another hybrid framework that combined fuzzy analytical hierarchy process, the fuzzy Kano model with zero-one integer programming, was also presented by Wang and Wang (2014) to incorporate customer preferences and customer perceptions into the decision-making process of the product configuration. In their study, an FAHP was used to elicit customer preferences for core attributes of a product while additional product attributes were not considered as the core attributes were obtained from the Kano model. Thus, maximisation of the overall customer utility was presented to determine the optimal product variety. Chang et al. (2009) proposed a method to train an artificial neural network (ANN) to group customers on the internet into clusters. The Kano method was used to extract the implicit needs from users in different clusters to enhance web personalisation. Rashid *et al*. (2011) applied a linguistic approach for identifying customers’ preferences which were not tangible. In this study, reverse attributes were considered based on a semantic approach. The threshold of the numbers was used to represent consumers' opinions which were later transformed into the probability needed for a Monte Carlo simulation system to produce virtual customers. In a study by Chen and Chuang (2008), a hybrid design combining grey relational analysis with the Taguchi method was used to optimise quality attributes with multi-criteria characteristics.

In an attempt to address the challenges in identifying one-dimensional and attractive characteristics, Chen *et al*. (2010) proposed a new creativity based-Kano model, the (C-Kano model). The C-Kano model combined innovative techniques such as theory of inventive problem solving (TRIZ) and the methods such as substitute, combine, adapt, modify, put to another use, eliminate, reverse (SCAMPER) were included in the into the Kano model to discover customer needs and identify attractive attributes. Principal component analysis (PCA) and the Kano model were employed to explore customer satisfaction by determining customers perception about a product using the composite decision criterion of eigenvalues (Dou *et al*., 2016). As most studies were focused on identifying customers' needs with Kano's model, Dou *et al*. (2016) proposed a combined Kano model and interactive genetic algorithm for effective product customisation, to develop customer-driven product design. The application of the Kano model in this approach was used to identify different customisation attributes according to their influence on customer satisfaction. Product attributes were then adjusted for customisation in IGA-based product design, to find the most optimal design profile. The assessment of the original Kano model questionnaires did not consider the vagueness and fuzziness in the answers of respondents. Thus, a fuzzy Kano model integrated with importance-performance analysis was proposed by Wang and Fong (2016), to capture customer perceptions of service attributes with the potential of increasing customer satisfaction. Multiple regression and logistic regression analysis were used to elicit the weight of the service attributes. Relevant service attributes were then used to forecast customer retention. Kim and Yoo (2018) presented a methodology for extracting customers' delighters (attributes that excite customers) from the Kano model using big data analytics. Opinions from online forums were extracted to obtain the 'delighter attributes'. Importance analysis based on a volume concept was presented using similarity analysis, volume size analysis and a potential delighter index.

Another method adopted by product manufacturers to integrate consumer needs in product design is that of quality function deployment (QFD). The QFD originated in Japan in the 1960s as a form of cause-and-effect analysis to transform the voice of the customer into an engineering characteristic. The QFD was also intended to develop quality assurance methods that would integrate customer satisfaction into a product before it was manufactured. The QFD was initially adopted to improve new automobiles that were being developed during the rapid growth in the Japanese automobile industry (Akao and Mazur, 2003).

The QFD is made up of four matrices whereby the output of each matrix is the input for the next matrices. The four matrices depict four phases: product planning, part deployment, process planning, and product planning (Liu and Wang, 2010). In the classical QFD, the input of the first matrix, also known as the house of quality (HOQ) representing the customer requirements, is translated to engineering characteristics. A correlation matrix is generated in the QFD between customer requirements and engineering characteristics, forming the roof or the top of the HOQ and a relationship matrix is developed between the customer requirements and engineering characteristics forming the body of the HOQ. A correlation matrix is also developed between the engineering characteristics and the importance of the engineering characteristics are also embedded in HOQ. Figure 2.4 shows the components of the HOQ of the QFD. Numbers are used to prioritising the importance of the engineering requirements and customer requirements. In order to acquire the data, HOQ surveys are usually conducted to elicit the needs of consumers. A fair number of studies have applied QFD in developing new products. The QFD has been used to determine optimal engineering specifications using a nonlinear mathematical program (Dawson and Askin, 1999). González et al. (2003) also presented a dynamic hierarchy process for a model to improve school furniture using the QFD.

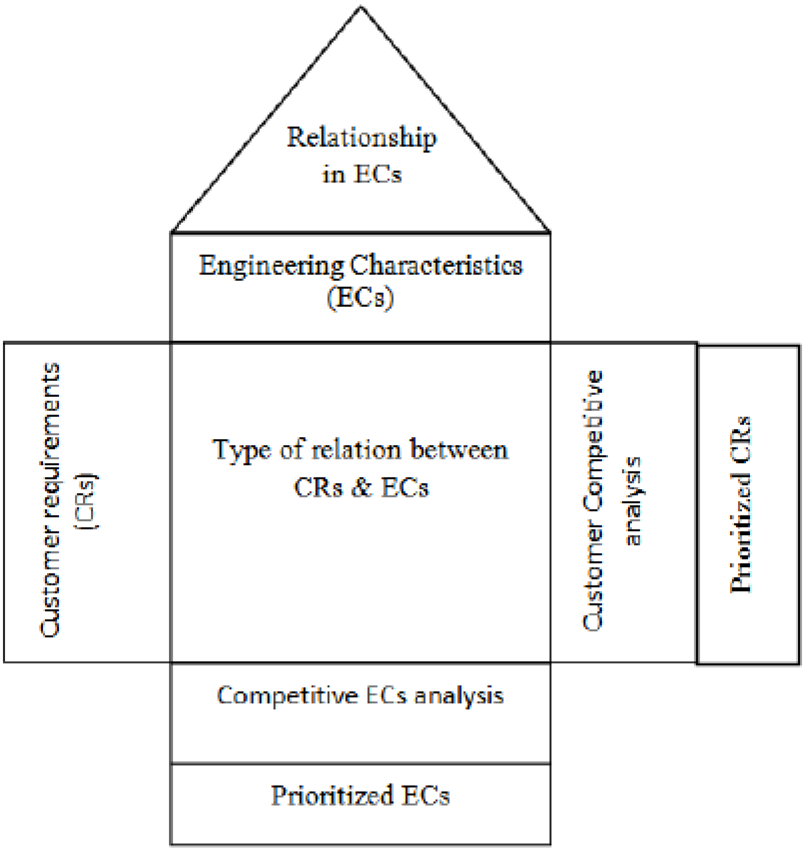

Figure 0.4 Components of HOQ

Lai et al. (2007) also proposed to optimize product design using Kano's model and goal programming. Due to the high subjectivity of the QFD, the interrelationship between and within the components of a QFD and their relative contributions in terms of the value were then determined using the analytical network process (ANP) (Raharjo et al., 2008). Yang et al. (2019) also presented a hybrid method of the theory of inventive problem solving and Kansei evaluation in QFD to facilitate new product design and evaluation in the early design phase. Similarly, the integrated approach of QFD and the theory of innovative problem solving and a fuzzy group decision-making theory for ergonomic product innovative design and evaluation in the early design stages was developed Zhang et al.,(2014) also proposed a QFD Optimizer based on an integrated mathematical programming formulation. The QFD Optimizer enables a design team to develop a HOQ to assist them to analyse and understand the complex relationship between customer needs and engineering characteristics and to also determine the most optimal target engineering requirements. In solving the uncertainties of qualitative information obtained from surveys for QFD, a continuous interval-valued linguistic term set was developed to capture customers preferences to improve aviation services dependence (Wu and Liao, 2021). The QFD has also been used to achieve a collaborative product design and optimal module mix in the multi-market segment using linear integer programming to maximize product functionality. This enabled manufacturers to balance the trade-offs between improving product varieties and optimising manufacturing complexity (Wang and Chen, 2012). A fuzzy nonlinear programming model based on Kano's concept was also used to determine the fulfilment levels of customer requirements.

2.3.1 Computational intelligence techniques for modelling customer satisfaction

Understanding how customers tend to be satisfied with a product in the real world can be complex and complicated for product manufacturers in the fast-paced globalised world market. To understand how this issue, data mining techniques have been adopted in recent years over the statistical methods due to the ability of data mining methods to deal with the incomplete data; identify patterns in data;

optimise processes and handle ambiguous data that past methods have failed to achieve. Surveys for the most part have been used in the extant literature to develop customer satisfaction models. Some of such studies can be seen in Kwong *et al*. (2010), who proposed a generalised fuzzy least squares regression to account for randomness and fuzziness in modelling relationships in QFD. Chen *et al*. (2005) proposed a fuzzy expected value operator approach to model customer satisfaction in the QFD process. The underlying principle of the study was similar to a stochastic programming method, but it addressed the inherent fuzziness in customers' requirements and engineering characteristics relationships. The target values of the engineering characteristics with maximum satisfaction and returns were selected for further development. An imprecise goal programming method was proposed by Cherif *et al*. (2010) to determine the optimal targets of the engineering characteristics in QFD, to maximise customer satisfaction. Zhong *et al*. (2014) proposed a fuzzy chance-constrained modelling approach using a fuzzy expected value operator and fuzzy chance-constrained programming. The objective of the proposed model was to minimise the fuzzy expected cost to determine the target value of the engineering characteristics. The complexities in modelling customer satisfaction to depict real-life scenarios imply a non-linear relationship that might exist in the model developed. The complexity in modelling the relationship between the customer requirements and technical attributes was addressed when a non-linear fuzzy model was proposed to provide an approach to incorporate resource factors in QFD (Fung *et al*., 2002).

Also, Chen and Chen (2014) proposed a normalisation model that satisfies Lyman's normalisation requirement in the house of quality (HoQ) of QFD with consideration of the correlation between customer requirement. The proposed model identified the essential design requirements by aggregation and a normalising function and addressed the shortcomings of the Wassermann, normalisation model. Further, Chen *et al*., (2004) presented a fuzzy regression-based mathematical programming method for product design. They considered in their model, the fuzziness, financial factors, and customer expectations. The model addressed both symmetrical and non-symmetric cases of fuzziness. Thus, a trade-off among different levels of customer satisfaction and obtaining a relevant specification of engineering characteristics for a new product was presented in the study. Dawson and Askin (1999) proposed a non-linear mathematical program for eliciting optimal engineering settings. The model was developed as a function of customer value functions, engineering development, production cost and development time constraints. Another study also produced mixed-integer linear programming that incorporated discrete data. The model was designed with multi-objective decision making to optimise cost, technical difficulty and customer satisfaction (Delice and Güngör, 2011).

Liu et al., (2014) presented fuzzy nonlinear regression to minimise the fuzziness in a model that relates customer requirements and engineering specifications. They considered the degree of compensation among the customer requirements during the trade-off of design attributes. Further, the authors explained that the prioritisation of products is influenced by the degree of compensation among the characteristics. Jiang *et al*. (2013) proposed a chaos-based FR approach to address nonlinearity and uncertainties in the customer satisfaction model from QFD. In this study, the authors generated polynomial structures from chaos optimisation algorithms, and FR analysis was then used to obtain the fuzzy coefficients of the evolved polynomial structures. GP was also used to generate models for relating customer satisfaction and design attributes (Chan *et al*., 2011). The model addressed higher-order and interaction terms in relating customer satisfaction to design attributes. Kwong *et al*. (2009) proposed a neuro-fuzzy approach to address nonlinear and explicit customer satisfaction models. Fuzzy rules were developed using an adaptive fuzzy inference system (ANFIS) from which significant rules were generated for the model. To aggregate customers' preference for product attributes, Grigoroudis and Siskos (2002) presented a multi-criteria satisfaction analysis method. The method aggregates a variety of customers' preferences in a particular satisfaction function.

Also, the disaggregation and aggregation procedure in the proposed method considers the judgement of customers as well as their choices. The model follows the underlying principles of ordinal regression to measure and analyse customer satisfaction. A FR data envelopment analysis algorithm was also designed by employing the four marketing mixes (product, price, place and promotion) and developing a relationship between customer satisfaction and new product design in a fuzzy context (Nazari-Shirkouhi and Keramati, 2017). Li et al., (2018) also integrated four machine learning algorithms, namely the support vector regression, the classification and regression tree, the ridge regression, and a multilayer perceptron to capture new affective responses of customers for a new product design through an online questionnaire. This was done to address the small scale, time-consuming and updating conventional modes of conducting surveys. Bi et al., (2019) also proposed a model of customer satisfaction using an ensemble neural network and an effect-based Kano model. The model function however could not be explained due to its black-box nature. An interactive genetic algorithm with an interval hesitation time and a fuzzy Kano model was also used to determine customer needs for a new product from morphological characteristics and perceptual images factors of different products (Wang and Zhou, 2020). This method heavily relied on expert and consumers interviews to assess predefined questions.

A fuzzy chance-constrained programming was also used in QFD to develop a new product design in a multi-segmented market. This method was used to search for optimal engineering characteristics by maximising overall customer satisfaction, and study also employed the classical survey method to elicit customers' needs (Fang et al., 2020). Jiang et al. (2019) suggested a dynamic evolving neural-fuzzy inference system to model customer preferences.

The limitation of the suggested method was that it was dependent on the type of membership function, the limitation of a high number of inputs affecting the accuracy of the model and the inability to reduce inputs to or select the best inputs to enhance the accuracy of the model.

2.3.2 Determining the importance of customer needs for product design.

Identification of product features according to relevance is one of the approaches that manufacturers use to manage scarce resources. This process of identifying and ranking product features allows them to prioritise goods according to the importance of products attributes. Liu (2005), proposed a nonlinear programming method to calculate the technical importance of design requirements by applying a more general form of the fuzzy weighted average. A fuzzy outranking approach was also proposed by Wang (1999) to prioritise design requirements in quality function deployment and determine the degree of outranking from two types of relation; concordance relation and discordance relation representing the degree of agreement and disagreement. Chan et al. (1999) proposed fuzzy and entropy rating methods to rate the importance of customer needs. They converted customer assessment into fuzzy numbers and used fuzzy arithmetic to obtain the relative weight of each customer needs. The entropy method of information was applied to analyse customers' assessment of the performance of a company's product attributes and competitors. An alternative method to fuzzy QFD was proposed by Yan *et al*. (2013) to prioritise design requirements in QFD. The proposed approach relied on an order based semantic of linguistics information and fuzzy preference relations of the linguistics profiles. Battistoni *et al*. (2013) employed the analytical hierarchy process (AHP) to find weights for customers' preferences. The AHP was of paramount importance in helping to determine the main features that products assume once released to consumer markets, providing evidence of which characteristics the final user is looking for, and the priority. Kwong and Bai (2002) argued that although the AHP is has been extensively used in weighing the importance of customer requirements, the importance judgements attributed to customer requirements are vague and uncertain. They proceeded to propose a fuzzy AHP to improve the imprecise ranking of customer requirements.

Further, Wang (2014) proposed an approach known as fuzzy quality function deployment (FQFD) with relative weight preferences and associated FQFD with a relative preference relation to identify adjusted criteria weights in a fuzzy MCDM model. Through the relative preference relation, the priorities of customer demanded qualities represented by relative preference degrees of customer demanded qualities over the related average are quickly derived. An adaptive attribute selection approach for configurator design using the Shapley value was proposed by Wang and Tseng (2011) to estimate the usefulness of each attribute in a configurator design. The method iteratively selects the most relevant attributes that contribute much information. A probability ranking principle is then applied to give a product recommendation. Kwong et al. (2007) proposed a methodology for aggregating the importance of engineering characteristics. They developed fuzzy relation measures between customer requirements and engineering characteristics as well as fuzzy correlation measures between engineering characteristic that were determined using a fuzzy expert system.

Quantification for the importance degree of engineering characteristics or prioritising engineering characteristics plays a central role in QFD. However, quantifying the importance of engineering characteristics can sometimes be challenging when using the quality function deployment tool. Jia et al. (2015) investigated the quantification problem for the importance degree of such target ECs by addressing the uncertainties, fuzziness and incompleteness involved during the evaluation process of quantifying engineering characteristics. They proposed a method based on the integration of the fuzzy evidential reasoning algorithm-based approach and the fuzzy discrete Choquet integral-based approach to derive the final importance degree of the target engineering characteristics. Furthermore, Chen et al. (2006) proposed a fuzzy weighted average method in a fuzzy expected value operator to rank technical attributes. The fuzzy importance value of each technical attributes was determined using a nonlinear programming method. The ranking of technical attributes followed the method according to the corresponding fuzzy expected values of their fuzzy importance. The fuzzy expected value operator was then defuzzified to prioritise the technical attributes without knowledge of the exact membership function.

A gradient weighted class activation mapping algorithm, a kind of neural network frames, has also been used to determine the relative importance of product attributes (Lee et al., 2020). In dealing with small sample sized problems and the subjectivity and randomness, Wang et al. (2020) suggested a hybrid cloud model and grey relational analysis method to prioritise technical product attributes in QFD. Ying et al. (2018) developed a hybrid-information multiple attribute decision-making problem approach based on cumulative prospect theory to select attributes of new product development concepts. Surveys and the Delphi method were used to obtain an assessment of the determined product attributes for the new product development concept alternatives. A new integrated design concept evaluation using a weighted least squares model based on vague set theory was also proposed to rank alternatives of product concepts in supporting product development under uncertain environments (Geng et al., 2010). In the same fashion, Aydoğan et al. (2020) proposed to incorporate Z-number and axiomatic design concepts for concept design evaluation. The proposed method however depended on predefined questions to elicit designed concepts hence did not address the fuzziness associated with consumers' needs.

2.4 Natural language processing and opinion mining

Consumer review sites usually have a large number of reviewers leaving feedback and comments in large volumes. Going through each review could present a daunting task for manufacturers in an attempt to get insights into consumers' thoughts about a particular product. An efficient means to process reviews is to use natural language processing (NLP) techniques. NLP is one of the areas that break down complex documents of words into forms that can further be processed by a computer. Thus, computers are programmed to process a significant number of natural language corpora. To make use of these reviews, NLP has been used to analyse and synthesise texts in reviews. In NLP, an algorithm is employed to identify and extract the natural language rules existing in the unstructured raw textual data. These rules are converted into an easy format for computers to interpret. In the NLP, two main techniques are utilised: syntactic analysis and semantic analysis. In syntactic analysis, words are arranged into sentences to make meaningful grammatical sense. The syntactic methods involve lemmatisation; where various forms of words in a text are reduced into a single form for easy analysis, morphological segmentation; where words are divided into individual units called morphemes. Also, other methods involved in the NLP are the part-of-speech- tagging, where parts of speech are assigned to every word, parsing; where grammatical analysis is provided for each sentence and finally stemming; where inflected words are reduced to their root form (Hirschberg and Manning, 2015; Martinez, 2010b; Sun et al., 2017).

In NLP, a corpus characterises a set of documents. Lexicons, which describe words, refer to different words existing in the corpus. Syntax defines the structure of sentences as well as the rules used in developing the sentences. One of the basics required in undertaking NLP is the efficient identification of the semantics (meaning) of the text. Texts in NLP are parsed by grouping the components of the text into syntactic structures. Humans communicate in ways that are different from the way we interact with computers via programming methods. Also, human communication is complicated and interwoven with different words and meanings. In word sense disambiguation, words present different semantics or meanings depending on the context and this is one of the main challenges in NLP (Martinez, 2010). The theories and mathematical models built-in NLP can process different languages but not in the same manner, as languages differ in rules, grammar, and construction. Some studies have also attempted to model NLP statistically. This is because the occurrence and frequency of certain words or phrases occur more than others (Krapivin *et al*., 2010). To make the processing of natural language faster, machine learning (ML) techniques have been used to allow computers to become trained on how to process texts, without requiring any explicit programming (Prusa and Khoshgoftaar, 2017). ML enables computers to work on many different documents that are presented. In existing studies, ML has proven to be efficient in handling large volumes of data. In the ML process of NLP, data is obtained, and this is followed by building a classifier. Relevant features are extracted from the data using a classifier which has to be trained and tuned to increase the accuracy of the classification (Kanakaraj and Guddeti, 2015).

A step further in NLP is the development of sentiment analysis and opinion mining tools. These tools not only give the meaning of the data for the computer to understand but also offer an insight into the thoughts and feelings of the online reviewers. In the world of e-commerce, sentiment analysis or opinion mining are used interchangeably to describe the computational process of classifying opinions from text data in order to understand a customer's attitude towards a product. It comprises extracting the polarity of the text to determine the sentiments as either positive, negative, or neutral. It also involves identifying and extracting product features or aspects that are of concern to the customer. Sentiment analysis follows the taxonomy displayed in Figure 2.5.

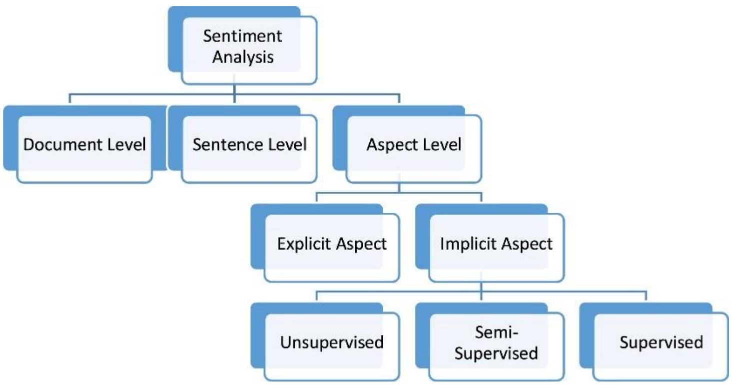

Figure 0.5 Taxonomy of sentiment analysis (Tubishat et al., 2018)

At the document level, sentiments in a whole document are extracted to determine the polarity of the whole document. Sentiment analysis is employed then to classify the entire document as either positive, negative or neutral (Chen *et al*., 2018a). At the sentence-level, of sentiment analysis involves, analysing all sentences in a document without considering the features or aspects in the document. Usually, the sentences are classified as either objective or subjective before classifying the document according to their polarity (Tubishat et al., 2018). The aspect level analysis gives a fine-grained analysis of the whole document. Sentiment on specific features regarding a product, service, or any entity in a review is extracted and classified according to their polarity (Chen *et al*., 2018a). One of the main tasks in sentiment analysis is polarity identification. That is, getting to understand how most reviewers feel towards a product. To determine the polarity of a text, opinions or sentiments are described by adjectives or adverbs appearing in the texts (Ding and Liu, 2007; Kobayakawa *et al*., 2009). One of the effective ways to get an insight into reviews posted by reviewers is to summarise a document into a list of phrases, which appear to convey the polarity orientation of a document's score algorithm. Hu and Wu (2009) for instance, presented a sentence weight classifier to classify sentences as either positive or negative by summarising the sentences. However, in some sentences, the polarity of the document depends on the context in which certain words are used. Linguistic rules and opinion aggregation functions have been used to determine the polarity of aspects of products in a document in such circumstances (Ding and Liu, 2007).

The linguistics rules applied include an intra-sentence conjunction rule, which compares different sentences to make a final decision on the polarity of a particular sentence. Next is the pseudo-intra-sentence conjunction rule, where convictions are compared to determine the polarity of other sentences when no explicit mentioning of the polarity of a product aspect is made in other reviews. In the last rule, an inter-sentence conjunction rule is followed, where the polarity of consecutive and neighbouring sentences is used to determine the polarity of other adjacent sentences. Besides establishing grammatical rules, semi-supervised learning methods have also been exploited to construct a domain-specific lexicon. Such methods combine shallow semantic parsing and corpus base statistical learning to allow a scalable and domain-specific sentiment extraction Lau *et al*., (2011). An ontology-based sentiment classification was also proposed by (Polpinij and Ghose, 2008) to classify the polarity of the online reviews using a support vector machine (SVM). The classifier employed used a lexical variation and synonyms in the ontology, and the words used in training the classifier were represented in a structured bag of words (BoW). This approach addressed the discrepancies in classes that we are unable to distinguish different varieties of words, such as 'block' and 'blocks', 'grow', 'grew' etc. Optimisation approaches were adopted by Li *et al*. (2015) to address the feature selection and sentiment classification problem. In this study, a particle swarm optimisation algorithm was employed to obtain an optimal global mix of features and parameters to use in SVM for the classification of a document. Some researchers have also attempted to merge classifiers in sentiment analysis to enhance the classification of textual data. For instance, in (Ankit and Saleena, 2018), the authors proposed an ensemble classifier that undergoes a pre-processing phase followed by the extraction of features from the refined data.

A base classifier is used for sentiment classification, and an ensemble classifier is used for sentiment classification and analysis. Likewise, Gang *et al*. (2014) also presented an ensemble approach based on the Bayesian model. Their model evaluated the contribution of each model represented in the ensemble. The authors also compared three other ensemble methods, bagging, boosting, and random subspace based on Naive Bayes, maximum entropy, decision tree, K-nearest neighbour and an SVM for opinion classification, to determine the effects of ensemble learning methods on sentiment classification. As human language can be fraught with different semantics such as metaphors, proverbs and the like, the true sentiments in a text could be difficult for a computer algorithm to decipher. As such, noise sensitivity as a result of language ambiguity leading to errors in classifying the polarity of textual data was addressed (Fersini *et al*., 2014). Kim and Lee (2014) used a semi-supervised Laplacian eigen map to reduce the dimensions of textual data, to make sentiment classification more efficient. Redundancy was omitted in this work to enable visualization of the document before any further processing. Interrelation among documents, topics and words were applied in a three-layer sentiment propagation model. Fuzzy theory was introduced in the model to make use of a fuzzy membership function representing the weight of textual data used in fuzzy SVM (Kim and Lee, 2014).

Deep learning is one of the artificial intelligence techniques that is becoming popular in big data analysis. Mahendhiran and Kannimuthu (2018) used a deep learning algorithm for polarity classification of large textual documents. Their algorithm is used is used to determine the polarity of text, videos, and even acoustic information. In addition, Jianqiang *et al*. (2018), introduced a word embedding method using unsupervised learning from textual data. Word embedding is incorporated in N-grams features with a sentiment polarity score of aspects. The feature set is then merged into a deep convolution neural network. The proposed model attempted to capture contextual information with the recurrent structure and form a representation of the text using the convolution neural network.

Moreover, sequential models have also been explored by Kang *et al*. (2018). In their study, a Hidden Markov Model (HMM) was used for text classification using a sequence of words. Sentiments of text patterns were observed using an ensemble text HMM classifier. Hidden variables were defined in semantic cluster information by conserving the co-occurrence of words. The polarity of the sentences arranged by the Text HMMs was then calculated. Another study by Fernández *et al*. (2016) showed a framework that determined the sentiment of textual data based on the determination of the relationships between tagged words. By using a sentiment propagation algorithm that considers linguistic characteristics, the sentiment of a text was determined. The features of the framework comprised intensification, modification, negation, and adversative and concessive relations. In their method, the negation played a relevant role in sentiment analysis. Other studies on hybrid classification algorithm were studied to develop a classifier that could perform sentiment analysis. An example of such a hybrid classifier is the ensemble technique made up of a naive Bayes and the genetic algorithm (GA) for sentiment classification. The genetic algorithm was formulated to predict the classification scores needed to predict the sentiments category in a text (Chen *et al*., 2018).

Unfortunately, some review sites are flooded with spam reviews to create a more positive image of a product. Fake reviews not only promote certain targeted products but also discredit other products, services, or an organisation. Three types of reviews were identified according to Jindal and Liu (2008). They include fake reviews, non-reviews and read reviews on a product. To mitigate some of these challenges, Guan *et al*. (2012) analysed online reviews for spam detection using a social review graph. The method of developing relationships between three main players: reviewers, reviews and online stores as a heterogeneous graph was introduced. The other factors from which the model was developed were the reviewers' trustiness, reviewers' honesty, and stores' reliability. An iterative reinforcement method was then employed to make use of the influence of the reviews on products that other reviewers had reviewed. Behavioural characteristics of users generating spam reviews have also been explored to assist in the understanding of what constitutes a spam review which differs from other reviews in their ratings. Specific products are also targeted to increase the effects on the reviews. Xie et al. (2012), observed that standard review patterns appeared stable and uncorrelated to their rating pattern. Spam reviews, on the other hand, presented an enormous amount of information that correlated with their rating. Accordingly, a time series model was constructed to represent the correlation.

Apart from the need to identify spams online reviews from e-commerce websites, the quality of the online reviews in terms of how helpful they are to other potential consumers was also studied. Customers usually have to struggle to identify which reviews might be beneficial to them. Moreover, useless reviews result in the need to developing efficient methods to select the most useful reviews for further processing. Kim (2006) proposed support vector machine (SVM) regression to assess the helpfulness of online reviews and rank the reviews according to the level of usefulness. A helpfulness function was then developed as a function of the rating from people who found the review helpful. Zhang and Tran (2008) presented an entropy-based approach to score and rank online reviews according to usefulness. A semi-supervised system called online review quality mining was proposed by Zheng et al. (2013) to identify useful reviews. The method proposed and used the characteristics of e-commerce communities to identify useful reviews and label instances and unlabelled instances to enhance the classification of these opinions. A study by Hong *et al*. (2012) showed a method of user preference-based features for implementing a binary usefulness review classification framework to separate helpful reviews from useless reviews. A ranking method was incorporated in an automatic voting system to rank reviews according to their helpfulness.

2.3.1. Opinion mining for product design

In product design, understanding customers' requirements is paramount as it safeguards against losses and the recalling of products. Customers' requirements are often described by the features or aspects of products that customers' desire. In opinion mining, feature extraction is one of the activities that is carried out by opinion mining classification models. An unsupervised approach to product feature extraction was proposed by Quan and Ren (2014) to extract domain-specific features by using similarity distance measures. A domain feature vector was obtained by comparing the features in reviews with a domain corpus. The study used a similarity measure called pointwise mutual information-term frequency-inverse document frequency to analyse the relationship between candidate features and domain features. Liu et al. (2018) described a method that extended the hyper induced topic search (HITS) algorithm on a bipartite network for feature extraction from online reviews. The proposed model consists of a directed bipartite features-sentiment relation network, bearing a candidate feature sentiment pairs obtained from dependency syntax analysis. The other part of the proposed model is the mutual HITS which integrates a pointwise mutual information weighted and HITS algorithm to rank candidate features to extract real product features. Similarly, Manek et al. (2017) proposed a Gini Index based feature selection method with an SVM classifier for aspect term extraction. Other machine learning algorithms employed in product feature aspect extraction are the integrated PageRank algorithm, synonym expansion and implicit feature deductions for feature extraction. Reviews or textual data are split into segments where part-of-speech (POS) tags are generated from individual sentences. A node rank algorithm ranks the process to identify the product features in review texts (Yan et al., 2015).

Khan et al. (2016) are presented a method on sentiment analysis and feature extraction by using a domain-specific lexicon. To assign weights on features in this study, a multi-objective model selection procedure was introduced to evaluate the feature weight by integrating SentiWordNet, a sentiment lexicon in the sentiment analysis process. The feature weights were then used in training an SVM to enhance the classification of the texts. Finally, classification of the texts was performed by the multi-objective model selection. In an attempt to increase the accuracy of the classification of the texts, Jin et al. (2016) also presented a method of identifying customer requirements from online product reviews, intending to extract opinionated sentences that bore specific features from a product in online reviews. The framework to achieve this purpose involved selecting pairs of comparative opinionated texts bearing specific product features and analysing them for similar product features. An optimisation problem was formulated by using a greedy algorithm and, a suboptimal solution was obtained for classification. Ambiguous expressions in texts, such as sarcasm, make it difficult for computers to classify a document or extract a certain feature. Some reviewers also use certain descriptions for products features which are usually not explicit for easy feature extraction. The features in these kinds of texts are known as implicit features. Wei et al. (2013) extracted implicit features from texts using a hybrid association rule mining algorithm. Association rules are mined through a different complementary algorithm. Candidate features are extracted from word segmentation, followed by POS tagging and feature clustering. The model makes use of a five-collocation extraction algorithm, and the co-occurrence degree in between the candidate features is computed. Fusion relation embedded representation learning framework was proposed by Govindarajan (2013) for an opinion object-attribute extraction. The method gives a framework that could fuse semantic structures and language expression features into object-entities and attributes entities. Chen et al. (2018) used a semantic similarity of synonyms for feature extraction. In this approach, feature vectors are identified with a bi-clustering algorithm and an improvement of the PrefixSpan algorithm was then used to detect frequent phrases. Using an aspect term extraction and opinion target extraction, a hybrid unsupervised method has also been experimented with to combine rules and machine learning for feature extraction. Candidate opinion targets were classified using a chunk-level linguistic rule and based on domain correlation, irrelevant candidates were filtered. The extracted texts were then used as labelled data, to train a deep gated recurrent unit network for aspect terms of the document (Wu et al., 2018).

The topic map has been shown to be capable of being used to extract feature aspects and sentiment words. For instance, Xia et al. (2016) enumerated domain topics and established a relationship among these topics, which were then incorporated into opinion mining. Other classification studies by Li and Tsai (2013) made use of fuzzy formal concepts analysis to classify documents into abstract form concepts, so as to increase the influence of textual ambiguity. Formal concepts analysis introduced in this study trained a classifier using concepts rather than documents. One assumption made in the study was that abstract entities were capable of opposing specific noise-sensitive variations and created a generalising pattern for abstract data. Another clustering approach was used by Das et al. (2016) for feature extraction of online reviews. The method generated a feature-based summary of a specific feature with a particular orientation. A probabilistic approach was employed at the word level, and each feature opinion pair was assigned to a feature-based cluster, whether it was positive, negative or neutral. Tucker and Kim (2011) explained that customer needs change over time and as such, and the ability to estimate the future needs of consumers allows design engineers to anticipate and plan for the next products generation. Thus, they proposed a predictive trend mining algorithm that captures customers' requirements over a period to determine the next generation of product features. This study involved a data mining approach but did not consider incorporating opinion mining from online reviews. To incorporate opinion mining and engineering design, Jin et al. (2016) explored the characteristics of big data mainly due to the advantages of the volume, velocity, variety and value of data from online reviews in order to understand customers' needs. The authors proposed a framework that dealt with big consumer data for customer requirements understanding. The features of products in online reviews were identified, and the corresponding sentiment polarities were obtained. After using a Kalman filter to forecast the trend of the customer requirement, a Bayesian technique was then employed to compare product features.

The competitive environment in the marketplace is one of the main drivers for manufacturers to improve their product design. For manufacturers to keep informed about the changes in the market and the changes in customers' needs, competitive intelligence has been used for a long time as one of the strategies businesses adopt to keep ahead of their competitors. The application of competitive intelligence has shifted to an online mode of espionage. A competitive intelligence approach was considered by He et al. (2015) to understand other competitive products in e-commerce websites that consumers were talking about. The numbers of times certain aspects of a product are mentioned and the sentiment polarity words towards the features competitors' products were extracted to make an informed decision. Most businesses undertake competitive intelligence on competitive products through the use of opinion mining for their product improvement. Tuarob and Tucker (2015) found customer preferences were quantified towards different product features to get insights into the features that resulted in a negative customer's experience. A data mining approach and knowledge-based system were designed to quantify customer's satisfaction during the usage of the product. The system also provided a means for extracting notable features classified as either weak or strong from social media. Jiang et al., (2017) proposed a methodology for identifying important product features as well as future important features based on online reviews. Fuzzy theory was introduced in this study to identify the importance of product attributes obtained from online reviews. Using a fuzzy inference system (FIS) method, the authors established linguistic rules for estimating the importance of the product features extracted online. The results were defuzzified to obtain the importance of each feature extracted from online reviews. Furthermore, a fuzzy time series was also utilised to estimate the future importance of the product attributes extracted.

Another study by Kim et al. (2016) also applied competitive intelligence in product development. The proposed method showed a hybrid opinion mining approach by combining lexicon-based sentiment analysis and machine learning classification. The procedure proposed comprised conducting opinion mining, categorising product features and finally classifying the purchase intention of consumers. These processes could allow manufacturers to predict consumer's behaviour towards their product as well as towards other competitive products. A Bayesian sampling method for feature extraction for product design was proposed by Lim and Tucker (2016). The proposed algorithm identified an optimal search keyword combination for making enquiries on one specific product data. This method reduced the mistakes in feature extraction and increased product-feature related knowledge when using text mining. Identifying latent product features is one of the most relevant product aspect determination means in product planning. Latent product features are usually hidden in online reviews and using the conventional feature methods described above for extracting aspects of products presents a challenge. Recently, topic modelling has been introduced in the field of sentiment analysis in an attempt to identify hidden customer requirements that are usually not explicitly stated. The latent features are a source of product identification opportunity since latent product features are usually identified from a small section of reviewers who can be classified as innovators. Van De Kauter et al.(2015) proposed a sentiment annotation scheme, which in contrast to most schemes, captures explicit as well as implicit sentiment. They used coarse-grained and fine-grained level sentiment analysis methods to accurately pinpoint the sentiment expressed about a given company. Xu et al. (2015) proposed to classify the non-explicit sentences for implicit feature identification in Chinese product reviews using a SVM. They extended the Latent Dirichlet Allocation to construct an explicit topic model. Some types of prior knowledge, such as must-links, cannot-links and relevance-based prior knowledge were identified, extracted, and incorporated into the explicit topic model automatically.

Wang et al. (2013) also proposed a hybrid association rule mining method, by mining as many association rules as possible via several complementary algorithms to extract implicit features. Zhang et al. (2017) presented a hierarchical constrained topic model and support vector regression to extract implicit features from texts. Schouten and Frasincar (2014) proposed to predict the implicit feature based on the choice of words in a sentence by leveraging the co-occurrence between a set of known implicit features and notional words. A deep convolution neural network and the sequential algorithm were proposed by Feng et al. (2018) to identify implicit features. They extracted the aspects comprised by words vectors, part of speech vectors, dependent syntax vectors to train the deep convolution neural network and then employed the sequential algorithm to obtain the sentiment annotation of the sentence. Zeng and Li (2013) formulated implicit feature identification into a text classification problem and designed a topic-feature-centroid classifier to perform the classification task. Similarly, Dosoula et al. (2016) developed a classifier that predicts the presence of multiple implicit features in sentences. The classifier makes its prediction based on a score function and is trained using a threshold. Multiple implicit features are identified when a score exceeds the threshold. Lee et al. (2016) devised a method to mine perceptual maps based on online reviews that automatically build perceptual maps and radar charts to understand implicit customers' needs and experience in using a product. Also, Zhou et al. (2015), explained that analogical case reasoning from sentiment analysis could be used to extract latent customer needs. Consequently, Zhou et al. (2011) proposed a method of extracting the latent needs of customers from online websites using case-based analogical reasons. In their study, a two-layer model was developed; the first, a sentiment analysis model projected to extract explicit features with a fuzzy SVM to build sentiment prediction. In the second layer, an analogical reasoning method was used to identify latent customer needs by establishing similarities and differences analogically between ordinary and extraordinary use cases. In the latter case, analogical reasoning, hybrid reasoning, combining a case-based and a rule-based case for an understanding in extracting latent features were considered.

2.5 Google Trends

Information seeking behaviour is common among consumers in order to make informed purchasing decisions. One of the popular mediums for consumers to seek information is the Google search engine. In 2006, Google introduced Google Trends. Google Trends is a platform that highlight and breaks down search activities from the google search engine on different topics into simple metrics.

It was estimated that about 2 trillion searches were recorded per year as of 2016 (Sullivan, 2016). Google later introduced Google Insights, a more advanced and detailed oriented platform that generated data on search trends in 2008. However, in 2012, Google merged Google Insight with Google Trends (Google Trends, 2021, Jun et al., 2018). Although other search engines exist, global statistics on search engines shows that since 2010, Google has maintained a 90% of the market share (Johnson, 2021).

2.5.1 Google Trends for Forecasting

The introduction of Google Trends has enabled some studies to project the future and integrate the Google Trends in business activities (Choi and Varian, 2012; Ettredge et al., 2005). Carriere-Swallow and Labbe, (2013) proposed to perform Automobile sales nowcasting in the emerging market was using data on Google search queries. The Google search queries data known, as Google Trends Index was observed to provide accurate assessment compared to other economic variables used in nowcasting. Nowcasting tends to be challenging to conduct due to the delay in releasing a key macroeconomic variable. Similarly, Fantazzini and Toktamysova, (2015) also used Google Trends and another economics variable to forecast sales of the car. This study considered only price as a car attribute among the variables for forecasting. Google Trends has also been adopted by product manufacturers to determine how their customers consume different competitive products (Vosen and Schmidt, 2011). Consequently, Won et al.

(2018) integrated sentiment analysis and the Google Trends index to develop a perceptual map that will enable product manufacturers to understand the competitive structure of a market. The uncertainties associated with consumer opinion in this study were not addressed. Similarly, Jun and Park (2017) proposed to enhance product brand visibility by positioning product branding using web search information. They derived relationships between competitive products as well relationships between the product of a brand by tracking the time-series information of the web search traffic among the products and verifying the changes in the status of each brand that occurred in the consumers' minds over time. Some studies also focused on forecasting tourism demand using web search data to reflect changes in tourism in demand in real-time (Bokelmann and Lessmann, 2019; Feng et al., 2019; Höpken et al., 2019).

2.6 Forecasting demand for a product

A comparative review of the demand forecasting methods is presented in this section. In demand forecasting, the future demand for a product is predicted to enable manufacturers to plan their operations effectively. In most cases, the forecasting of the demand for a product are done based on past sales data. The conventional methods for forecasting the demand for products can be classified into four categories. The first category, known as the judgment method, that requires minimal quantitative skills, short term and medium-term forecasts has limited or no past sales data. The second category of forecasting is the counting method, mainly used for medium term forecasting and requires minimal to moderate term forecasting. Past data are not necessarily needed in the counting method and the required quantitative skills could be high. The third category is the time series forecasting mainly used for short, medium, and long term forecasting. This method of forecasting requires minimal to moderate quantitative skills and past data for the forecasting is necessary. The fourth category is the association or casual method where basic to high quantitative skills and past data is necessary. This method of forecast works for medium to long term forecasting.

The non-conventional methods for product demand forecasting make use of artificial intelligence methods, machine learning methods and other complex mathematical forecasting model. For large amounts of past dataset, artificial intelligence techniques such as the artificial neural network, genetic algorithm and backpropagation neural network can generate an arbitrary nonlinear approximation function from the historical dataset for demand forecasting (Liu et al, 2013). Wavelet transform, Bayesian network and fuzzy logic have also been used for demand forecast (Kumar and Kumar, 2019). The conventional and non-conventional demand forecasting methods are summarised in Figure 2.6.

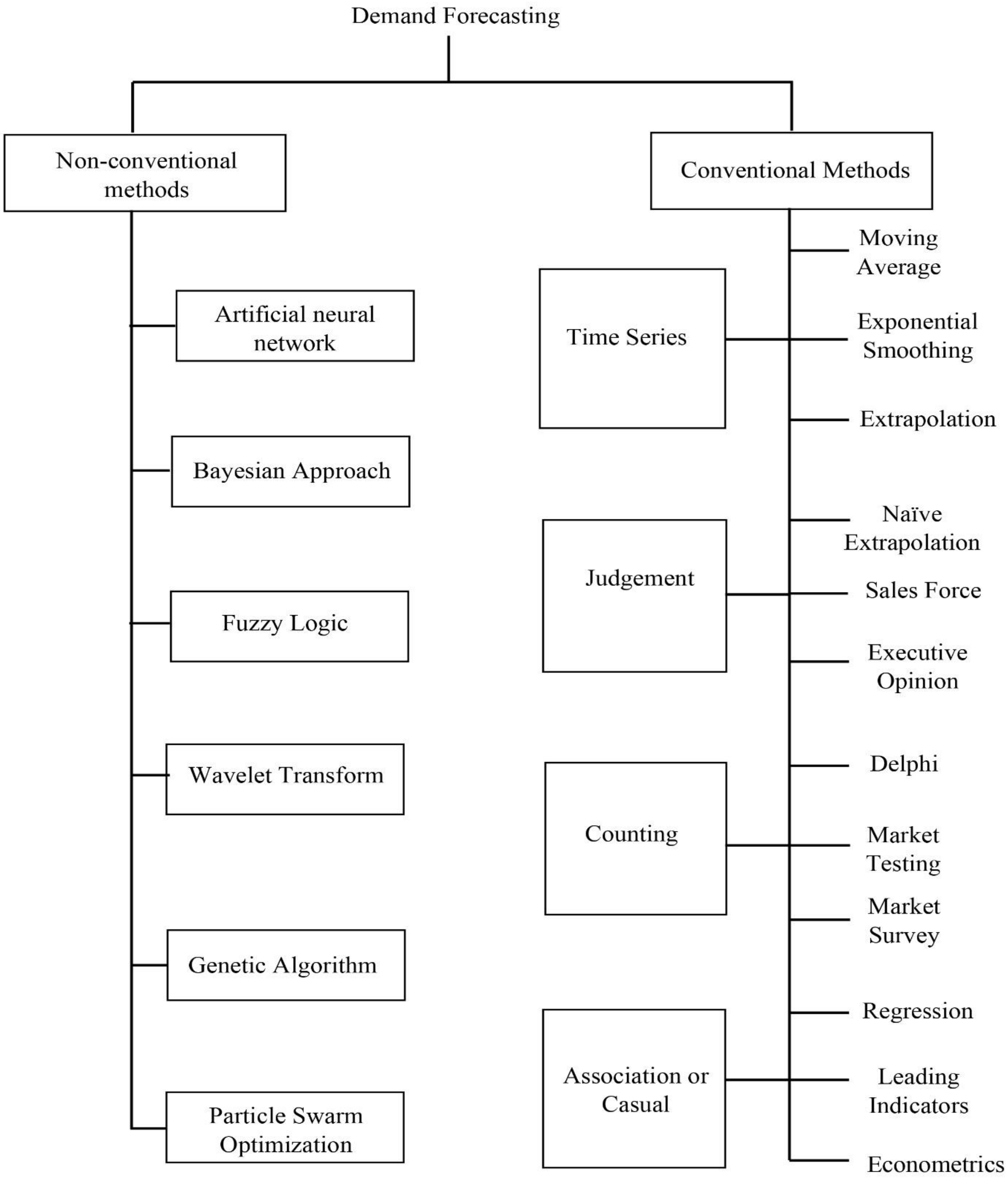

Figure 0.6 The convention and non-conventional methods of demand forecast

2.6.1 Statistical approaches to forecasting

In the last two decades many demands forecasting models have been developed, although each forecasting methods developed came with its own drawbacks and limitations. Traditionally, the autoregressive integrated moving average (ARIMA) has been widely used in forecasting the sales of a product (Irhami and Farizal, 2021; Krutz et al., 1986). Although ARIMA is simple and easy to implement, the ARIMA method does not consider the other factors that may affect the future demand. For time series forecasting, models are fitted into past data to make predictions for the future and other factors that affect the future demand for a product are not considered. The time series methods (Lee and Cho, 2009; Neto et al., 2016) however, do not address the issue of seasonality and may be affected by multiple factors which may be missed.

To compensate for lack for a historical dataset, Mostard et al., (2011) proposed a "top-flop" classification approach and found that the expert judgement methods for forecasting small products gave better results compared to methods that depends on advanced information methods. Gupta et al., (2019) also examined the Bayesian compressed vector autoregressive approach that considered other factors like housing sentiment that could affect the prediction of house sales. The proposed method was valuable to handling high dimensional datasets. Linear regression has also been used to develop forecast models that relates demand to other variables that can affect the demand forecasts. A good example can been seen in (Netisopakul and Leenawong, 2017) whereby the multiple linear regression and gradient descent method has been used to forecast the sales of cars in Thailand. The results of the multiple linear regression outperformed other forecasting methods when compared. In some of the statistical forecasting, although some optimal and objective results were obtained, “experts” knowledge was required to choose the method for demand forecasting. Identifying the most relevant explanatory variables affecting the demand forecast in some of the statistical techniques can be challenging or even unknown. This leaves many of the methods associated with statistical forecasting to be fraught with uncertainties.

2.6.2 Artificial intelligence methods for forecasting

With the drawbacks in the statistical methods, artificial intelligence techniques have been adopted in recent studies to forecast the demand for a product. Particularly, the artificial neural network (ANN) models and other models in the same categories have been reported to provide better forecasting results compared to the statistical methods. In the study by Tebaldi et al., (2019) the ANN was used to forecast the demand in automobile industry. Eighteen car components were considered in the ANN and after series of training and testing of the model, the results proved to be satisfactory. In Güven and Şimşir, (2020) the ANN was used to forecast the demand for retail apparel. The ANN was adopted due to its ability to work on large datasets and high input dimensions. The results of the ANN were compared to the support vector machine and the former have better forecasting performance. Other applications of the ANN can be seen as studies where the ANN was used to forecast the demand for a product. Other hybrid models of the ANN to forecast the demand for a product were also established to address the limitation of the ANN. Hao et al., (2018) developed a hybrid prediction model based on ANN and a grey model to forecast vehicles at the end-of-life cycle. The ANN and Bayesian rules training were used together to also forecast the demand in the furniture industry (Yucesan et al., 2017).

Similarly, the ANN and the genetic algorithm were also used for forecasting and the results achieved better forecasting performance when compared with standalone artificial intelligence methods (Cicek and Ozturk, 2021). The evolutionary neural network was also used to forecast demand for apparel using two years of sales data. The evolutionary neural network shows better forecasting performances when compared with other neural networks and traditional forecasting models (Au et al., 2008). Rahman et al. (2011) also proposed to forecast demand using the Bayesian approach. The Bayesian model can predict the probability of the demand in the future by expressing it as conditional on the observed demand before the peak sale season of US apparel dealers.

The adaptive neuro-fuzzy inference system (ANFIS) that combines fuzzy logic and the learning capacities of neural networks, has been adopted to forecast the demand of computers (Atsalakis, 2014). Another hybrid system combines the wavelet kernel support vector machine and particle swarm optimization for demand forecasting. This method was proposed to address forecasting with small samples while addressing the fuzziness and non-linearities associated with demand forecasting (Wu, 2010).

Data mining methods such as the k-nearest neighbour models have also been used to forecast using large time-series data sets (Kück and Freitag, 2021). Some studies also combined multiple machine learning methods as seen by those who suggested combining K-means, random forest and the quantile regression forest to estimate both the demand and variability of demand for new products with short lifecycles (van Steenbergen a Mes, 2020). Similarly, hierarchical self-organising maps and support vector regression was also proposed to predict the demand quantity of a product. The hierarchical self-organising map was used to classify the data, thereafter the support vector regression was used to develop the demand forecasting model (Lu and Wang, 2010). Lu et al., (2012) made a case that the identification of important forecasting variables was challenging in the ANN. Hence, they suggested to use a multivariate adaptive regression spline to construct a sales forecasting model. This model was found to have better forecasting results when compared to neural networks, multiple regression, support vector machines and ARIMA. This method however does not address the uncertainties in forecasting the demand of a product.

Most of the literature on forecasting the demand for products developed these models with the assumption and knowledge of past or historical sales data. To enable demand forecasting of new products, the Bass model has been explored in the extant literature to forecast the demand for both new and existing studies. The Bass model also considered two main relevant factors that could affect the future demand of product with limited or no past data. These two factors, considers the drivers of product adoption related to the behaviour of consumers. Moreover, the uncertainty in demand forecasting in the literature are limited. Table 2.6 shows a summary some of the literature in demand forecasting in the area of artificial intelligence.

Table 0.6 A summary of the demand forecasting models using artificial intelligence

Reference	Method	Application field	Finding	Limitation
Tebaldi et al., (2019)	Artificial neural network	Automobile	Had a better forecasting performance compared to traditional forecasting	Did not address uncertainty in demand forecasting
Güven and Şimşir, (2020)	Artificial neural network	Retail apparel	Had better forecasting performance compared to the support vector machine	Does not address understand and customers sentiment in the forecast model
Hao et al., (2018)	Artificial neural network /grey model	Vehicle	Had better forecasting performance compared to another neural network	Uncertainty in demand forecasting not addressed.
(Yucesan et al., 2017)	Artificial neural network/ Bayesian training rules	Furniture	The forecasting results had satisfactory performance	Required large amount of past dataset. Uncertainty in the forecast were not also addressed
(Cicek and Ozturk, 2021)	Artificial neural network/ genetic algorithm	Vehicle	Performed better than the support vector machine, back propagation neural network and the ARIMA/SARIMA model.	Uncertainties in the forecasts were not addressed.
(Au et al., 2008)	Evolutionary neural network	Fashion apparel	Had better forecasting performance when compared with SARIMA	Does not address the uncertainty in demand forecast
Rahman et al. (2011)	Bayesian approach	Apparel	Addresses the fluctuating demand forecasting during peak season. The model had better forecast performance compared to the adaptive Holt-Winters seasonal forecasting model.	Does not include other factors that may affect the future demand. Uncertainties not addressed in the demand forecast
(Atsalakis, 2014)	ANFIS	Computer	The ANFIS had better performance compared to autoregressive and autoregressive moving average. Does not require any prior assumption of the distribution of the dataset.	Uncertainty in the forecasting not addressed. Consumers sentiment not considered in the forecast
(Wu, 2010).	Support vector machine/ particle swam optimization	Car sales	Addressed the issue of uncertain data and finite samples.	Customer sentiment not considered in the forecast. Other factors that will affect the future demand of the product not considered
(van Steenbergen a Mes, 2020)	K-means, random forest, and quantile regression forest	Home and industrial items	Used the product attributes of new and existing product to predict the profile of new products.	Consumer sentiment and uncertainty not considered in the forecast

2.7 Bass model

The demand for a product in researched into in extant literature, has mainly been assessed using traditional methods like the time series and the judgement approaches. However, these approaches require past data in order to make predictions on future demand. Demand forecasting may also be used in estimating prices, predicting capacity requirement, or decision making on resources allocation. Product manufacturers rely on demand forecasting to ensure that the required and appropriate quantity of inventories are in place to ensure the smooth running of production processes. New product demand forecasting however presents some difficulties to product manufacturers due to increasing uncertainty in the market. The lack of or limited historical sales also present a challenge in forecasting demand for a new product. Thus, some studies have proposed different methods to enable product manufacturers to forecast new products and existing products.

To address the challenges of the traditional methods, Bass (1969) introduced the Bass model in the context of innovation of diffusion of innovation. The innovation of diffusion of innovation describes the process in which innovations are communicated through specific mediums over a period in a system (Everett, 2013). Through the bass model, the demand for new and existing products can be determined by considering the two main drivers of product adoption. According to Bass, products adoptions are driven by two factors, namely the coefficient of innovation and coefficient of imitation. The coefficient of innovation describes individuals who adopt product based on external influences like the mass media. These individuals are called innovators. The coefficient of imitation describes individuals who adopt products based on social pressure and various influences. These individuals are known as the imitators. By considering consumers' adoption behaviour in demand forecasting, the Bass model has been attractive to industry players when trying to develop demand forecasts for their products.

Since the introduction of the Bass model in 1969, several extensions from the original Bass model have been developed. The effects of variables such as price and advertisement of a new product were integrated into the original Bass model (Bass et al., 1994). Chien et al. (2010) also proposed a multi-generation diffusion model that considered growth rate, repeated purchase and technological substitutions. Other studies like Lee et al. (2006) integrated the conjoint analysis into the Bass model to forecast the demand of products. The extensions of the Bass model however fails to address uncertainty in demand forecasting. Also, the coefficients of the Bass model, although defined with specific values, does not factor the potential of current trends affecting the value of these coefficients. The Bass model coefficients can be made adjustable by defining them with an equation to factor into them current trends. In Table 2.7, the extant literature on the Bass diffusion model extension is shown.

Table 0.7 Bass diffusion models and its extensions

Reference	Remarks	Limitation
(Song et al., 2015)	1) Developed a hybrid Bass-Markov model to determine the demand of wireless broadband service. 2) The proposed model allowed continuous forecasting using past activities in the field of service forecasting	1) Does not address the fuzziness in demand forecasting 2) Innovation and imitation coefficient not explicitly defined by a equation to factor current trends
(Zhang et al., 2020)	1) Integrated online reviews, search data and macroeconomic data in the Bass model to improve the forecasting ability of the Bass model. 2) Coefficient of imitation was with a word of mouth index 3) Coefficient of innovation was a replaced with a Baidu search index to reflect search engine Baidu activities	1) Does not address the uncertainties in market demand forecasting
(Fan et al., 2017)	Integrated online reviews in Bass model	1) Does not address the uncertainties in demand forecasting
(Liang et al., 2015)	1) Developed a self-restraining Bass model 2) Address the case where adopters restrain the growth of adoption in the diffusing process of the Bass model	1) Does not address the uncertainties in demand forecasting 2) Does not factor other effects that might affect the demand forecast in the future
(Wang et al., 2017)	Developed a grey Bass model to address the subjectivity in determining the market potential for the Bass model.	1) Does not address the fuzziness in demand forecast 2) Does not consider other factors that might affect the forecast of demand

2.8 Discrete choice models

A common practice among product manufacturers is to investigate consumers purchasing behaviours by measuring trade-offs among attributes in a given set of alternatives. This practice known as choice modelling has been implemented in transportation and tourism and in the marketing research teams of product manufacturers to explore the financial feasibility of developing new products or services (Kuklys, 2002; McFadden, 1974). Discrete choice analysis (DCA) is an approach to choose modelling that determine the maximum utility of a given product among a set of competitive products. From the marketing perspectives in DCA, the utility of choice is a function of the attributes of all possible choices and the attributes of consumer making the choice. In DCA, consumer preferences are used to estimate the choice probabilities of all possible choices of a consumer. Different DCA models reported in the literature have been used in demand modelling. Such models include the MNL (Chen et al., 2013; Resende et al., 2011), nested logit (Ma and Kim, 2015) and mixed logit (Hoyle et al., 2010) have been applied the demand modelling. In the study by Higgins et al. (2012), a diffusion model for electrical vehicles by integrating multi-criteria analysis and discrete choice model was developed. For demand forecasting with limited historical information, Eggers and Eggers, (2011) proposed a choice-based conjoint adoption model that used individual-level preferences as a basis for prediction electrical vehicles adoption. A latent class choice model, a kind of discrete choice model was also integrated with a network effect model to predict technology adoption (El Zarwi, Vij, and Walker, 2017). Jun and Park, (1999) also suggested a choice-based diffusion model for multiple generations of products. The study incorporated exogenous variables such as price though the choice behaviour of consumers. Liu and Cirill (2018) proposed a generalized dynamic discrete choice model for product adoption. The model considered the dynamic market conditions by a stochastic diffusion process. A choice modelling methodology for usage in context-based design was also developed to quantify the degree of usages context consumer choice. A taxonomy user context based design was defined by modelling the usage context influence on both product performance and customer preferences He et al., (2012).

2.9 Discussions and research gap identification

Many frameworks and methodologies for modelling customer satisfaction have been proposed in extant literature. However, the majority of these studies relied on surveys to elicit the needs of consumers. Surveys depend on respondents' responses to the specifically designated questions which can limit the actual concerns of consumers. To address this problem, online reviews from the internet can be used to provide numerous opportunities for product manufacturers to elicit consumers' needs. Online reviews have not been explored beyond the extracting of consumers concerns. Thus, opportunities exist to develop customer satisfaction models from online reviews. Also, nonlinear regressions have commonly been used to develop customer satisfaction models in past studies. The developed models can contain interacting explanatory variables and higher-order explanatory variables. However, most of these studies rarely addressed the fuzziness associated with the modelling. Therefore, new approaches that will utilize online reviews and address the non-linearities and fuzziness associated with modelling customer satisfaction models are required.

MCDM approaches have been used extensively to identify trade-offs that will enable product manufacturers to identify the best product attributes for a product. Most of the studies also relied on surveys. Surveys do not keep up with the dynamic needs of consumers and do not allow product manufacturers to design products that satisfy their customers'. Thus, product design attributes relevant today maybe irrelevant tomorrow. Online reviews and Google Trends which have not been given much attention in product design can be used to predict the future importance of product design attributes. Online reviews are considered in this study because of the ease of obtaining real time data on the needs of consumers while Google Trends can provide data attributes of products that are regularly searched on the google search engine. Thus, data obtained from online reviews and Google Trends can be used in a robust framework that can aid product manufacturers to predict which product attributes will be relevant in the future.

While many studies have developed forecasting models, few have considered consumers concerns and sentiments on products and how these sentiments can drive product adoption. Moreover, fuzziness and uncertainties in forecasting models are rarely considered. Thus, the research opportunity is to integrate consumers sentiment in forecasting models while addressing the uncertainties in forecasting the adoption of a products.

Chapter 3. Methodology

In this chapter, an in-depth description of the methodology related to this research is presented. The proposed methodology comprises the use of online reviews and computational methods to assist manufacturers in the new product development and the design process. The three main areas of focus in this research are; modelling customer satisfaction from online reviews, estimating the importance of customer requirements and formulating a model using online activity metrics to estimate the demand of a new product. The proposed methodology mainly comprises the development of MGGP-FR for modelling customer satisfaction based on online reviews. Next the determination of the importance and future importance of product attributes based on Shapley value, the Choquet integral and the fuzzy rough set time series method is demonstrated. Lastly, a fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews is presented in this research. The overall framework for this research is shown in Figure 3.1.

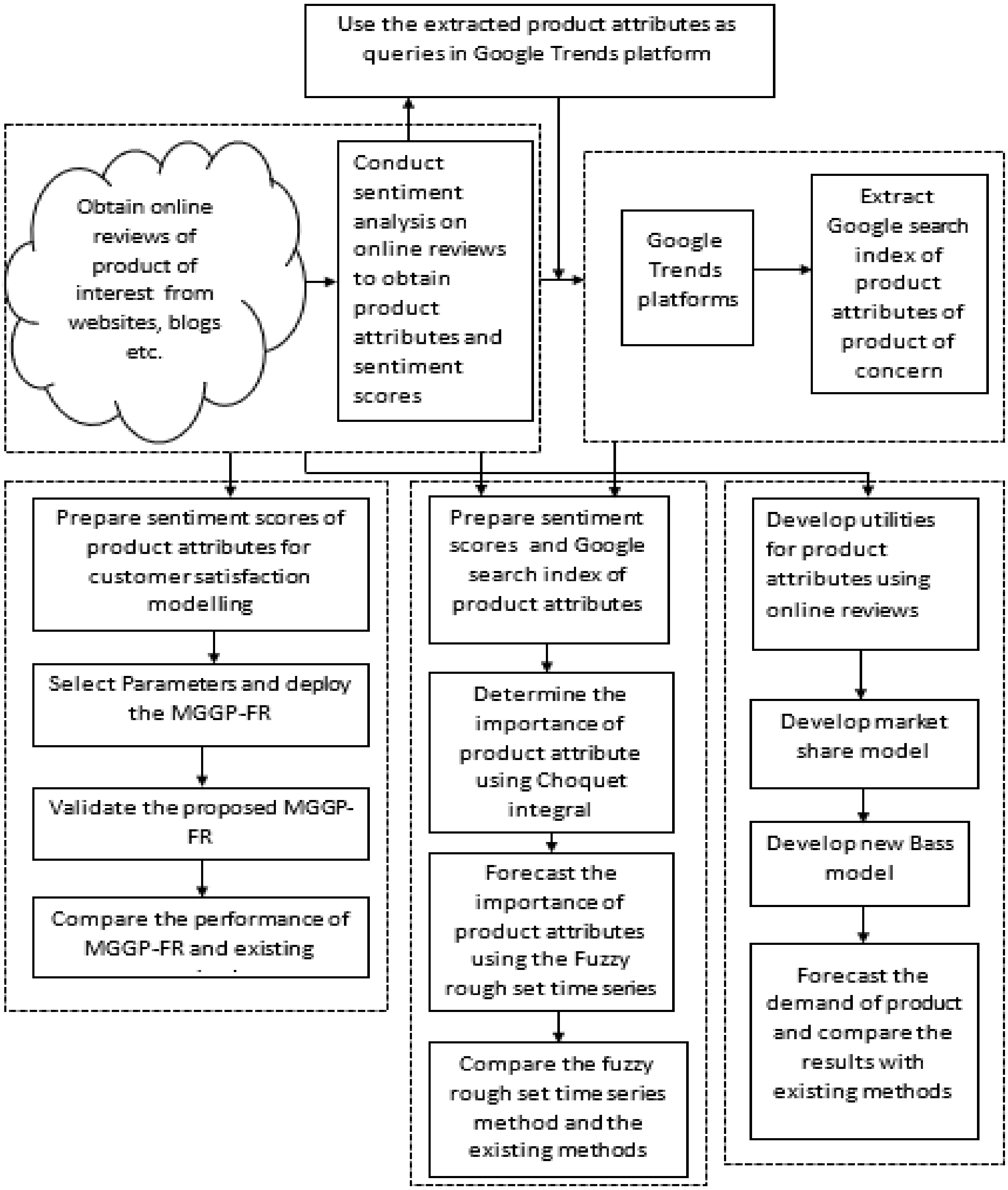

Figure 0.1 Overall framework methodology

3.1 Opinion mining for product design

Opinion mining involves a series of steps to classify unstructured texts from online reviews into three categories of sentiments; neutral sentiments, negative sentiments, and positives sentiments. Different text analysis tools exist on the market to facilitate the process of opinion mining on online reviews. With many choices, this study settled on the use of Semantria, a text analysis tool that performs sentiment analysis through an application programming interface or an excel plugin. Semantria caters for sentiment analysis in various industries. Since this study is focused on analysing texts from e-commerce websites, retail/e-commerce, the voice of customer category analysis in Semantria was selected for opinion mining. Different brands/models of products of the same type belonging to the same categories are considered in this study. As most e-commerce websites have sections for potential customers to leave their reviews on the product, the web crawler needs to be customised to extract these reviews. The information extracted includes a customer or reviewer ID, the reviews, and the ratings for the product. Besides, the development of customer satisfaction models will require information on the main aspects of the products that customers have a particular interest in, also known as customer concerns or customer requirements.

In this research, online customer reviews are used in the formulation of a customer satisfaction model. The first step in developing the customer satisfaction model is to conduct sentiment analysis on the extracted online reviews. There are three methods or approaches popularly used in conducting opinion mining, follows: The keyword based classification approach; the key-word based classification approach, for instance, classifies text according to the presence of positive or negative polarity of words. Words such glad, excused, sad, unhappy, scared and frustrated dictate the polarity of words and the text as a whole. Thus its classification mainly depends on the presence of these polarity determining words (Cambria *et al*., 2013). The second method is the lexicon-based classification method: this method develops a list of words manually labelled as having a positive and negative polarity. For each word, a polarity score is created. The lexicon formed is used to compute the overall sentiment score of a text. The lexicon-based method requires no training data as opposed to a supervised machine-learning method. This form is fit for analysing text like reviews from forums and blogs (Aggarwal and Zhai, 2012; Taboada *et al*., 2011).

The last method used in sentiment analysis is the use of the machine learning method, which is used to handle large volumes of data. It comprises both supervised and unsupervised learning algorithms. Both approaches require a large volume of data to train an algorithm to uncover hidden associations in unlabelled data. Examples of the machine learning based approach include support vector machine, naive Bayes, K-nearest neighbour, decision trees, artificial neural network etc. (Shayaa *et al*., 2018). In this research, Semantria for excel plugin, a product of Lexalytics made for analysing large volumes of texts, is employed for the opinion mining. Semantria combines the three methods commonly used in sentiment analysis to analyse opinions in texts.

3.2 Modelling customer satisfaction based on online reviews

In this research, the results of opinion mining from online reviews are used to develop customer satisfaction models. Explicit customer satisfaction models are developed to address the fuzziness and nonlinearity associated with customer satisfaction modelling. Recently, some fuzzy polynomial regression approaches, namely the GP-FR and chaos-based FR approaches, have been proposed to model customer satisfaction. The chaos based FR is based on the stochastic search algorithm integrated in an optimization strategy to accelerate the search for a global optimal solution. It is preferred for its ergodicity, intrinsic stochastic property, and its sensitivity to initial conditions. The draw back with the chaos-based optimization is that it has the worst stability compared to other algorithms, such as the genetic algorithm. This is because, chaotic sequences are developed from both poor and good solutions as a results of a large-scale solution space. The population obtained has the potential to lead to poor areas during searching resulting in a population with unstable generations (Laili et al., 2015). Since the genetic algorithm provides a better stability compared to the chaos-based optimization algorithm, Chan *et al*. (2012), proposed a GP-FR for modelling customer satisfaction. The method of genetic programming is based on the principle of the traditional genetic algorithm that provides more stable individuals after a set number of generations. However, GP-FR method exhibited some drawbacks. Although it generated stable solutions, it had a poor generalizability and gave poor predictions with high errors compared to its lower training errors.

Thus, a more robust MGGP was proposed to generate a mathematical model with a better generalising and fitting capability (Jiang *et al*., 2016). The MGGP provided more accurate results through fewer evaluations. This is made possible because, in the MGGP, more outputs per GP individual are generated. The conventional GP treats a tree function as an individual in the GP population and each of the population is a weighted linear combination of sparse trees. However, according to Akhil *et al*. (2013) and Ankit *et al*. (2015), MGGP generates a model that does not give satisfactory performance in test data. This poor performance on the test data is likely to present false information about the underlying principles in the data. Since the formulation of the MGGP is random, a gene with poor performance could be regressed with genes of higher performance, and that could degrade the performance of the model according to Garg *et al*. (2014), so the best model may not perform the best on the testing data. The MGGP has been successfully applied in (Akbari *et al*., 2014; Chen *et al*., 2016; Garg and Lam, 2015). The fuzziness associated with variables have not been considered so far in the formulation of the MGGP method. This has become relevant in this study because opinions on online reviews on certain products attributes vary in terms of sentiments scores. In opinion mining, a positive sentiment score signifies positive reviews, whiles a negative sentiment score indicates negative reviews. A zero score indicates a neutral opinion. A positive sentiment score could vary due to differences in sentiments on the same product attributes, and this vagueness and uncertainty could be addressed by employing FR analysis. The integration of FR analysis and the MGGP could increase the performance of the training and testing data by addressing the vagueness and uncertainties in data during modelling.

MGGP-FR is proposed in this research to model customer satisfaction in order to develop non-linear structures which will exhibit interactions between terms or possess higher order terms. FR analysis is then employed to generate the fuzzy coefficients of the individual terms evolved from MGGP method. The proposed method utilises the natural selection process of the genetic algorithm to select models with better prediction performance. Moreover, operations such as mutation enable variables to be omitted or included to increase the diversity of the solutions proposed. With the capability of the MGGP to represent each gene with a “tree”, a better generalising capability is obtained compared to the GP where a whole chromosome represents a tree.

3.3 Determination of the importance of product attributes

One of the MCDM methods used to determine the importance of product attributes is the fuzzy analytical hierarchy (FAHP). This method relies on pairwise comparison of product criteria to rank product attributes or criteria. The FAHP also depends on surveys which sometimes can be time consuming and maybe unreliable when some respondents do not provide all the required responses to a questionnaire required. Moreover, the pairwise comparison does consider the influence of more than two criteria on the determination of importance of product attributes. To address the limitation of studies similar to the FAHP that depends surveys for determining the importance of product attributes, online reviews have been proposed in some recent studies to identify the relevant product attributes needed for product development (Jiang et al., 2017; Kim et al., 2016)**.**

Consumers in recent times have also developed the habit of searching for products online before making any purchases as a result of different modes of advertisement which have played a key role for consumers to seek information on products in order to make informed purchasing decisions. This change behaviour is enabled by the development of internet technologies, where a new paradigm in advertisement relies on the internet to propagate information on a product. The online advertisement comes in the form of search engine marketing, search engine optimisation, social media marketing, email marketing, online referral marketing etc. (Nosrati *et al*., 2013). The google search engine is one of the most commonly used search engine used globally. Google has a platform called Google Trends that provides data and graphical descriptions on the number of searches on certain search terms or topics of interest by internet users over a specified period. It is noticeable from Google Trends that before new products of popular brands of products are released into the market, consumers search activity online can indicate their interest towards the products and the product attributes. Therefore, Google Trends data of product attributes can reflect the importance of the attributes perceived by consumers to a certain extent. However, no previous studies were found to have considered the Google Trends data in the determination of the importance of product attributes. This research aims to incorporate Google Trends data, sentiment scores and frequencies from online reviews to determine the importance and future importance of product attributes.

To determine the importance of product attributes, the approach employs the Shapley value and Choquet integral method. An electric hairdryer is used as a case study to demonstrate how to determine the importance value of a product's attributes. Thereafter, the fuzzy rough set time series method is employed to predict the future importance of the product attributes of the electric hairdryer.

Three online metrics (sentiment scores, frequencies and Google Trends data) were considered to estimate the importance and future importance of product attributes in this study compared to existing approaches, where only two online metrics (sentiment scores, frequencies) were employed in the estimation of the importance and future importance of product attribute (Jiang *et al*. 2017). The proposed approach mainly involves the development of determining the weights of the sentiment scores, frequencies and the Google Trends using the Shapley value. Next the densities of these online metrics are determined for the

Choquet integral, in order to estimate the importance value of each online metric (Cherkassky, 2011).

The Choquet integral is a complementary statistical tool that can facilitate the analyses of complex evaluations and the behaviour of consumers. It allows for estimating the importance of different online metrics such as sentiment scores, frequency of product attributes mentioned in online reviews and Google Trends data by considering the dependencies existing between them. It allows a better understanding of the relationships existing between the online metrics, especially in the presence of multicollinearity. Furthermore, comparison of the different product attributes can be easily performed easily using the Choquet integral. Moreover, the Choquet integral assists in the determination of the overall importance of each product attribute while reflecting the importance of each metric (Alfonso, 2013).

In the Choquet integral, the weight of an individual online metric, also known as the fuzzy density, is required to formulate an aggregation model. The fuzzy density of each online metric is determined by first conducting a survey among experts to determine how significant each online metric is in determining the importance of a product attribute. Based on the survey results, the Shapley value method is employed to estimate the fuzzy density of each online metric.

The fuzzy densities describe how important each online metric is in the determination of the importance of a product attribute. The fuzzy measures describe how the weight of each metric is when considered alone or combined. A typical example and application of the use of fuzzy measures is in an organisation where the efficiency of workers is measured. The efficiency of workers who smoke and chat simultaneously is less than the aggregation of the effectiveness of workers who smoke and chat independently where Effectiveness *(smoke and chat) < Effectiveness (smoke) + Effectiveness (chat).* Thus, in reality, the variables of interest are usually dependent on each other, and such dependence presents higher insight than variables that are considered individually. Fuzzy measures are also abstract and are suitable for measuring the effectiveness of drugs, the effectiveness of workers or opinions from people. The abstract nature in the measurement sets it apart from the explicit and definitive variable whose aggregation gives an exact measure in drawing insight from data. With the fuzzy measures obtained, the Choquet integral is used to estimate the importance of product attributes an electrical hair dryer.

3.4 Determination of the future importance of products attributes

A fuzzy time series approach based on rough set rule induction is proposed to estimate the future importance of product attributes. This phase of the research aims to bridge the time lag of estimating the importance of product features and launching a new product into the market. Moreover, by taking advantage of the big data available online, manufacturers can be updated with the fast-evolving customers' needs, especially concerning technology products. The evolving needs at each point in time reflect features of products that customers have much interest in and what they will most likely prefer in the future. Jiang *et al*. (2017) employed the fuzzy time series to estimate the future importance of product attributes, but according to Hosseini *et al*., (2011) the fuzzy time series ignores some fuzzy logical relations (FLR) leading to an increase in the uncertainties in predictions. In this study, all the FLR are maintained and based on the FLR, the rough set method is employed to generate rules for forecasting the importance of product attributes. Linguistic values are defined for each fuzzy variable in each period for forecasting the future importance of product attributes. In the formulation of rule induction, the Learning from Examples version 2 (LEM2) algorithm evaluates a local covering and then converts it into rule set by exploring the search space of attribute-value pairs. The LEM2 has the potential to learn from the smallest set of minimal rules to describe a concept from a decision table. The algorithm was chosen because of its ability to generate rules from inconsistent data and generate certain and possible rules from a decision table.

3.5 A fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews

The artificial neural network, exponential smoothing and autoregressive models require a large amount of data to generate a model that can fit the data.

Unfortunately, obtaining data of past demands could be challenging as it could take years to gather enough meaningful data to develop a model to predict future demands. In recent times, consumer attitudes towards demand for a product are dynamic due to fast-paced changes in customers' requirements. By relying on past demands collected over a long period of time for demand forecasting, businesses do not benefit from such forecast due to delay in meeting the demands of their customers in a timely manner. To enhance forecasting abilities in existing models, some studies have sought to leverage the availability of big data from online reviews in econometric models to forecast the demand of products. None of these studies considered the uncertainties associated with big data. Moreover, since most businesses are concerned with meeting the demands of their consumers, it is not favourable for businesses to depend on collecting data on customers past demand for a long period before developing forecast models. Thus, by being able to develop a forecasting model with certain product data and online reviews, businesses could forecast product demand. Also, to address the uncertainty in the demand modelling, FR is adopted in this study to obtain three scenarios in in the demand modelling. They are: i) Worst case scenario ii) Best case scenario iii) Normal case scenario. The demand for Apple tablet P.C in U.S.A is used as a case study in this phase of the research.

The main steps involved in this phase of research are a) to obtain online reviews and extract sentiment scores of old and new competitive products b) to obtain market shares of competitive products, c) to estimate the importance of product attributes using a fuzzy inference system (FIS) d) to estimate the utility of the product attributes, e) to develop the FR that relates the utility of the product attributes to the product ratings in order to obtain three scenarios for forecasting the demand for a product , f) to formulate an optimization model that minimizes the actual market share and the predicted market share from the MNL model and finally, g) to develop the Bass model integrated with sentiment scores from online reviews. The modified Bass model is used to forecast the adoption of tablet P.C in this study.

Chapter 4. A Multigene Genetic Programming based Fuzzy Regression for Modelling Customer Satisfaction

The gap between the perceived quality of a product and pre-purchase quality expectation can be measured by customer satisfaction (Chow, 2015; Guo et al., 2017). Customer satisfaction measurement enables product manufacturers to improve their communication with their consumers. It is also supposed to enable product manufacturers to meet their customers' expectations and design products with relevant attributes to be included or improved. Lastly, customer satisfaction, modelling, and measuring also enable service providers to determine their strengths and weakness compared to their competitors (Lucini et al., 2020). However, a non-linear relationship exists between customers’ needs and customers satisfaction, and hence, determining customer satisfaction can be complex (Basfirinci & Mitra, 2015). In this study, user-generated content, namely, online reviews, extract customers’ needs. An MGGP-FR is used to develop a non-linear model that can relate customer needs and satisfaction.

Section 4.1 to section 4.5 of this chapter describes the proposed approach for modelling customer satisfaction using the MGGP-FR. In section 4.6, the application of the proposed method for modelling customer satisfaction for an electric hair dryer is presented. The validation of the proposed methodology is presented in section 4.7, and in section 4.8 the summary of the above chapter is given. The methodology for modelling customer satisfaction is summarised in Figure 4.1

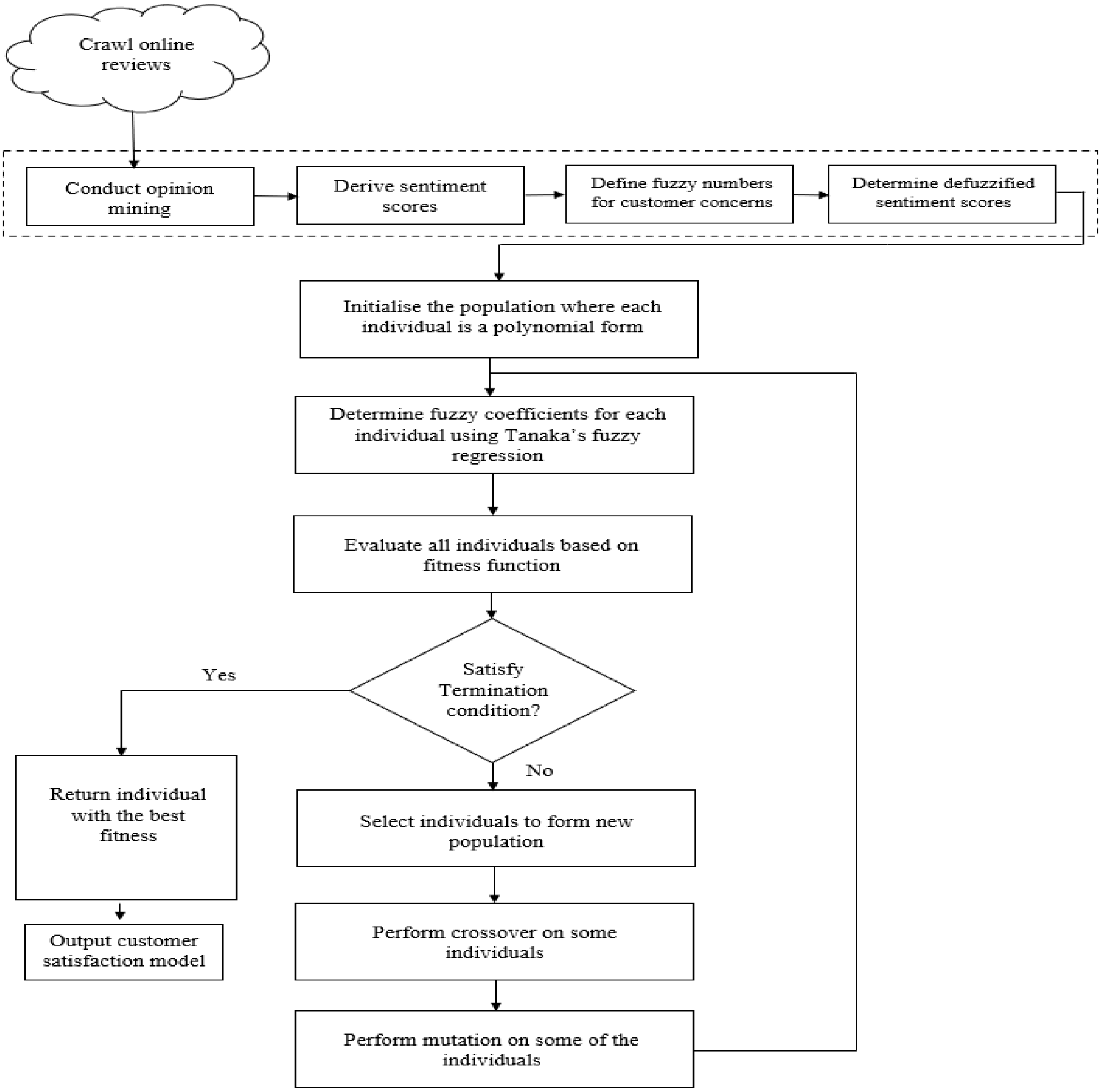

Figure 0.1 Methodology for modelling customer satisfaction

4.1 Fuzzy polynomials

In developing a customer satisfaction model in this research, a datasets obtained from online reviews are used as inputs for the model to generate polynomials of second and or higher order terms and interaction terms using the MGGP. Tanka's FR is then employed to determine the fuzzy coefficients of the nonlinear terms generated by the MGGP method by solving a linear programming problem. Polynomials for nonlinear models in the literature have been developed using the Weiner series, based on models of geometric transformation. This series is also known as the "Kolmogorov-Gabor polynomials". Fuzziness in these polynomial models is usually not addressed. To address this deficiency, the fuzzy polynomials proposed in this research is used to generate non-linear terms in which there is interaction between explanatory variables and higher order polynomials from MGGP. Thereafter, Tanaka's FR method is used to determine the fuzzy coefficients of the polynomials generated from MGGP. The fuzzy polynomial developed based on MGGP-FR is expressed as shown below:

$$\tilde{y} = \widetilde{\mathrm{A}}_0 + \widetilde{\mathrm{A}}_1 x'_1 + \widetilde{\mathrm{A}}_2 x'_2 \ldots. + \widetilde{\mathrm{A}}_{\mathrm{N}} x'_N \tag{4.1}$$

where $\tilde{y}$ is the fuzzy value of customer satisfaction; x'_n, where n = 1, … N, is a single customer concern, interaction term involving several customer concerns or a higher-order term of a customer concern; and $\tilde{A}_N$ are the fuzzy coefficients of the terms of the model.

$\tilde{A}_N$ is expressed as (a^c, a^s), where a^c and a^s are the central value and the spread of the fuzzy coefficients, respectively. Thus, the fuzzy polynomial model shown in Eq. (4.1) can be rewritten as follows in Eq. (4.2):

$$\tilde{y} = (a_0^c, a_0^s) + (a_1^c, a_1^s)x'_1 + (a_2^c, a_2^s)x'_2 + \cdots + (a_N^c, a_N^s)x'_N \quad (4.2)$$

4.2 Multigene genetic programming

GP uses evolutionary algorithm inspired by biological evolution which aims at optimising user defined fitness functions by using genetic operators such as crossover, reproduction, and mutation. The structure of GP is embedded in the underlying principles of the genetic algorithm (Garg *et al*., 2014). The GP has been successfully applied in rule discovery by Hu *et al*. (2015) and its application is seen in predictive modelling by Afzal and Torkar (2011), natural language processing by Araujo (2006), function optimization by Miller and Mohid (2013), robot optimization control by Kala (2012) and fuzzy logic by Yang and Soh (2000). The GP is well known for its application in symbolic regression, where mathematical expressions for a group of functions and variables are determined. The symbolic regression in GP is encoded in the form of a hierarchical tree in which each tree node represents a function, variable, or a constant number. The functions comprise arithmetic operators such as (+,-,/, x) nonlinear- function such as (sin, cos, tan, exp, tanh, log) or Boolean operators, a set of terminals $T = \{x, p\}$ containing a design attribute set $x = \{x_1, x_2, \ldots, x_n\}$

of the hierarchical tree is shown and the coefficient set $p = \{p_0, p_1, p_2 \dots p_{n_t}\}$ of the tree, with n_t being the number of terms of the function. Every tree represents an individual of the population of GP and every potential solution is depicted as a tree with branches made of operations (internal nodes of tree) from the function set and arguments (terminal nodes of the tree) from the terminal set T. An example of a hierarchical tree is shown in Figure 4.2, and can be expressed as a nonlinear polynomial of higher order terms and nonlinear terms, is as shown below as:

$$(x_1 * x_1) - (x_2 * x_2) + (x_1 * x_2 * x_4)$$

Which is also the same as;

$$x_1^2 - x_2^2 + x_1.x_2.x_4$$

The coefficient for the trees is represented as $\boldsymbol{p} = (p_0, p_1, p_2 and\ p_3)$ and is determined after the polynomial structure is determined. Thus, the number of coefficients in this example is four. The tree can finally be represented as:

$$p_0 + p_1.x_1^2 - \ p_2.x_1^2 + p_3.x_1.x_2.x_4$$

The coefficients of the of the nonlinear function generated from the GP can be estimated using the orthogonal least square algorithm

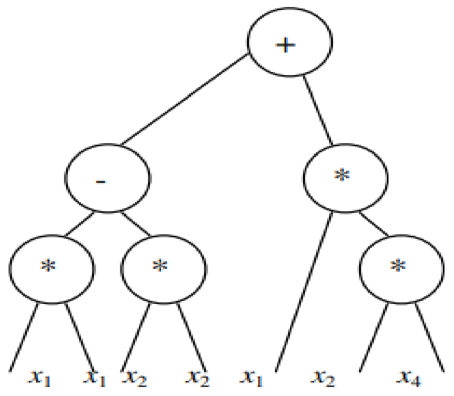

Figure 0.2 An example of a hierarchical tree of a GP

In MGGP, each individual of the GP is expressed as the weighted linear combination of the outputs from a large number of GP trees, with each gene in an individual is represented by a tree. An example of a multigene model is illustrated in Figure 4.3. In the model, the output is determined from three explanatory variables (b_0, b_1, b_3). The multigene GP model is linear in its parameters with three coefficients (c_0, c_1, c_3) . It also comprises nonlinear terms such as square root and sin (x).

The parameters settings of the MGGP are characterised by the maximum number of genes, known as T_{max}, and the maximum tree depth D_{max} of each gene in a chromosome. Moreover, as mentioned earlier, in traditional genetic programming, the evolved model is composed of a single tree while in MGGP, each regression model is a weighted linear combination of several trees. This allows compact models to be developed, unlike in GP where there is a higher chance of complex and over parameterized models being developed (Gandomi and Alavi, 2012; Nuo *et al*., 2019). A higher transparency, ease of interpretation and a possible higher accuracy of the model is achieved by allowing more outputs per GP individual. A canonical GP formulation treats a tree (function) as an individual in the GP population (Maynard *et al*.,2018; Searson, Willis *et al*.,2007; Strachan *et al*., 2014). A simple approach is to enable more functions per individual and to combine their output. In this way, multi-gene (or multi-tree) GP concepts can become important. The models are also of low order and the linear-in parameters of the coefficients are determined from the training data using the ordinary least square method. According to (Searson *et al*., 2007), the multigene symbolic regression may be computationally more efficient than the conventional GP. The MGGP ensures that a parsimonious polynomial structure with significant terms and tree depths are generated.

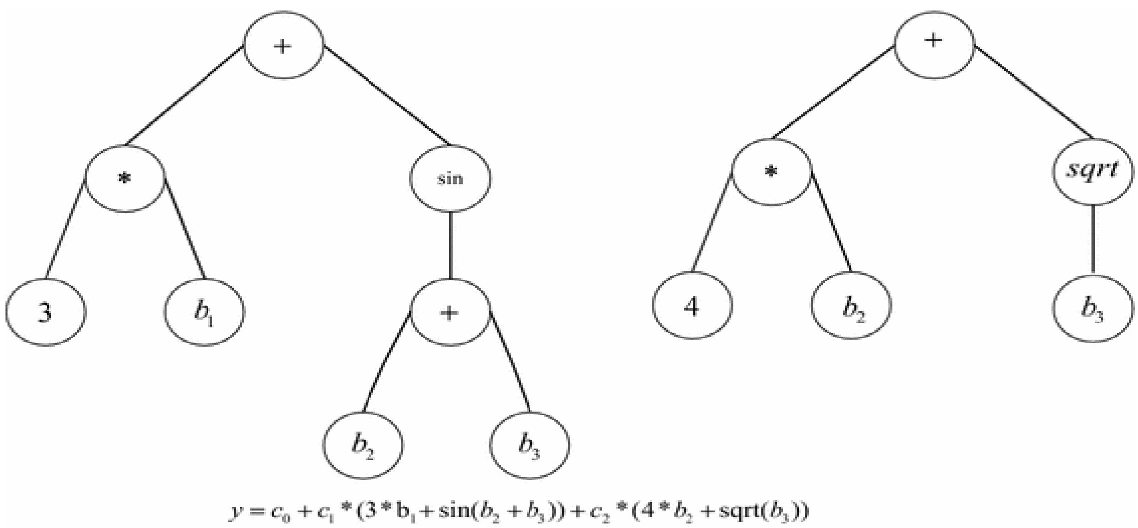

Figure 0.3 An example of genes in MGGP

4.3 Fitness function and crossover and mutation

The programs that come closest to achieving the underlying behaviour in the dataset are selected for "breeding" using the fitness function, the root mean square error (RMSE). MGGP is used to optimise the fitness function defined in Equation (4.3) below:

$$RMSE = \sqrt{\frac{1}{N_{train}} \sum_{k=1}^{N_{train}} \left(\tilde{y}(k) - y(k)\right)^2} \tag{4.3}$$

where $\tilde{y}(k)$ is the predicted customer ratings in the k^{th} data. $y(k)$ is an actual customer rating? N_{train} is the number of samples. The crossover and mutation operators are employed to enhance the diversity and likelihood of achieving a higher performance model. In the crossover operators, random nodes from two parent trees representing a pair of solutions are selected. The nodes selected could be a subtree (subexpressions), which are exchanged between the two main trees. Thus, the new pair of offspring (new trees) inherit characteristics from each parent. In mutation, a node is randomly selected from a tree, and it is replaced by a random subtree expression to generate a new solution and to introduce diversity into the genetic population. The multigene symbolic regression are used to generate the genes of the MGGP. Based on the model complexity analysis, the symbolic regression has the ability to remove redundant models. Low level crossover operations are also carried out in the MGGP to swap relevant section of genes between the two parents. The crossover operations are usually performed at the same depth for both parents to avoid the generation of infeasible mathematical tree.

The actual customer satisfaction and the predicted customer satisfaction value could be used to interpret the performance of the model. The performances of the evolved models from MGGP-FR, and other existing methods could be compared using the performance indices; the mean relative error (MRE) and the variances of errors (VOE) described by the equations below:

$$MRE = \frac{1}{N_{train}} \sum_{k=1}^{n_{train}} \frac{|\tilde{y}(k) - y(k))|}{y(k)} \tag{4.4}$$

$$VoE = \frac{1}{N_{train}-1} \sum_{k=1}^{N_{train}} \left(\frac{|\tilde{y}(k)-)\, y(k)|}{y(k)} - MRE \right)^2 \tag{4.5}$$

where $y(k)$ is the actual customer ratings in the k^{th} dataset and $\tilde{y}(k)$ is the predicted customer ratings? N_{train} being the number of samples.

4.4 Determination of fuzzy coefficients using fuzzy regression analysis.

Most systems influenced by human estimations are accompanied by fuzziness. The fuzziness in these systems is addressed in many studies using Tanaka's FR analysis, in which the parameters are represented by fuzzy sets. Zadeh's extension principle defines the fuzzy linear function. The statistical assumptions in Tanaka's FR are also relaxed. In Tanka's FR, inputs, outputs, and coefficients can be represented by fuzzy numbers. To evaluate the output, two criteria are considered, the least absolute deviation and the minimum spread. The deviations are regarded as the fuzziness of the parameters of the systems and are reflected as linear functions with fuzzy parameters. Thus, it represents a good model that shows the relationship between dependent and independent variables (Peters, 1994; Tanaka, Uejima, and Asai, 1982). The linear programming problem of the fuzzy polynomial is formulated as:

$$Min\,J \sum_{j=0}^{N_{NR}} \left(a_j^{s'} \sum_{i=1}^{M} \left| x_j'(i) \right| \right) \tag{4.6}$$

The objective function J, describes the fuzziness of the system. There are $1 + N_{NR}$ terms in the fuzzy model; M is the number of data points; and $\left| x_j'(i) \right|$ is the jth absolute value transformed variable of the ith dataset in the fuzzy model. The constraint of the objective function is defined as follows:

$$\sum_{j=0}^{N_{NR}} \left(a_j^{c'} x_j'(i) + (1-h) \right) \sum_{j=0}^{N_{NR}} a_j^{c'} \,|\, x_j'(i) | \ \geq y_i \tag{4.7}$$

$$\sum_{j=0}^{N_{NR}} \left(a_j^{c'} x_j'(i) - (1-h) \right) \sum_{j=0}^{N_{NR}} a_j^{c'} \,|\, x_j'(i) | \ \leq y_i \tag{4.8}$$

$$a_j^{s'} \geq 0, a_j^{c'} \in R, j = 0,1,2, \dots, N_{NR}$$

$x_0'(i)$=1 for all I and $0 \leq h \leq 1$

The h-factor, which take values between 0 and 1, measures the degree of fitness of the fuzzy polynomial model and y_i is the value of the ith dependent variable. The constraint in (4.7) and (4.8) makes certain the dependent variable has at least h degrees of belonging to $\tilde{y}_i$ with $\mu_{\tilde{y}_i} \geq h, i = 1,2, \ldots, M$. The last constraint ensures that $a_j^{s'}$ and $a_j^{c'}$ are non-negative.

4.5 Algorithm of multigene genetic programming based fuzzy regression.

The algorithm for MGGP-FR is specified below.

Step 1. Initialise the parameters, function nodes and fitness function.

Step 2. Generate individuals by crossover, mutation, and reproduction.

Step 3. Obtain the fitness value and when the termination condition is met, move on to step 3. When the termination criterion is not met, repeat step 2.

Step 4: After evolving the best individuals, select the individual with the best fit and simple polynomial structure.

Step 5: Determine the fuzzy coefficients for the polynomial model using the method described in section 4.1.4. Solve the LP problem of Equation (4.6), (4.7) and (4.8), the fuzzy coefficients. $\tilde{A} = \left(a_j^{c'}, a_j^{s'}\right)$ of each term of the fuzzy polynomial is determined.

The predicted variables $\tilde{y}$ are then determined from the fuzzy polynomial model. The relative error between the predicted value $\tilde{y}$ and the actual value y is determined. The MRE of the training and testing dataset is also determined. The proposed model is validated by comparing with three well-known algorithms, FR, the GP and the GP-FR method.

4.6 Implementation

A case study on modelling customer satisfaction for an electric hair dryer is illustrated in this chapter to demonstrate the proposed methodology for modelling customer satisfaction from online reviews, using MGGP-FR. The Amazon e-commerce website (www.amazon.com) is one the most prominent online retailers on the internet currently, alongside with other e-commerce websites such as AliExpress and e-bay. It was selected to identify twenty-two popular brands of electric hair dryers and their online reviews available on the review section on the Amazon website. Amazon was chosen in this study because there were a lot of online reviews found on most of the products listed there. Also, the reviewers who post on Amazon are verified to ascertain whether they had already made purchases on the product they intend to write reviews on.

Moreover, reviewers are not restricted to the number of words they can type. There is also a rating section where customers can give their overall ratings on the product. The products were selected based on the number of reviews posted about them. To make identification of the products easier, each product was denoted alphabetically from A-V. With the twenty-two hairdryers identified, the next step was to design a web crawler to extract the online reviews of each of the hair dryers into an excel sheet for sentiment analysis. A summary of the number of reviews extracted with the overall ratings for each hair dryer is shown in Table 4.1. The ratings are made on a scale from 0 to 5 on the review section on the Amazon website. The reviews collected were reviews posted from January 2017-January 2018.

Table 0.1 Summary of reviews extracted from Amazon

Product	Number of reviews	Overall product rating
A	898	4.2
B	910	4.4
C	272	4.3
D	300	3.9
E	1028	4.1
F	1018	4.1
G	119	4
H	315	4.4
I	364	4.4
J	302	4.2
K	310	3.2
L	691	3.8
M	1229	3.4
N	364	4.5
O	1772	3.3
P	1779	4.2
Q	426	3.1
R	227	3.7
S	459	4
T	638	4.6
U	139	3.7
V	360	4.3

On the Amazon product review section, consumers post their opinions on the hair dryers based on their experience with the product. Their opinions usually contain aspects of the products which could either bear a positive, negative or neutral sentiment. Some reviews contain no opinions on the aspects of the product under review. Those kinds of reviews are usually a generic statement on whether the reviewers were pleased or not pleased with the product. In Figure 4.4, the first review bears the concerns on the hair dryer regarding the weight and the location of the control buttons, whiles the second review gives a generic positive statement about the hairdryer. As a result, a large number of reviews were extracted to capture reviews that contains relevant customer concerns. Moreover, locating individual concerns from a large number of reviews is assisted with text processing and sentiment analysis. Semantria for excel was employed to process the reviews and conduct opinion mining. Semantria generates themes, concepts, topics, intentions, summarization, parts of speech tagging and entities from texts with accompanying polarity. The polarity denotes the positive and negative sentiments on the categories of the text processing results. In Semantria, each unstructured sentence of a review is broken down into tokens by a process called tokenisation. Each token represents a phrase or words contained in a sentence. The tokens in the English language are identified by the presence of white spaces and punctuation. Next, is the part of speech tagging. In this phase, each token is identified by a token tagger. Consequently, the token tagger figures out whether each token is a noun, pronoun, adverb, adjective, verb or some other part of speech.

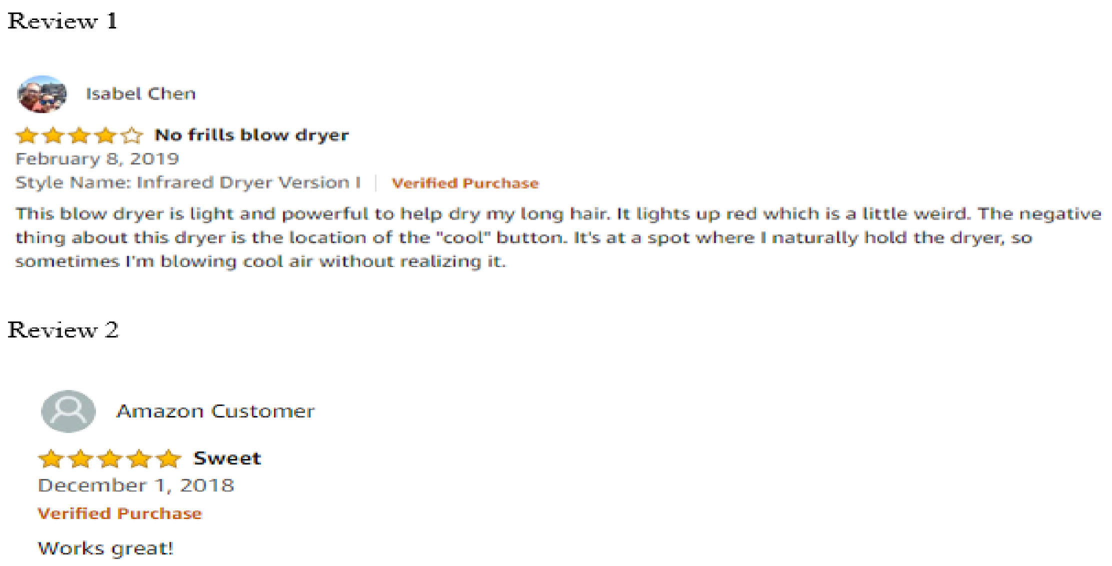

Figure 0.4 An example of customer reviews on the amazon e-commerce

The part of speech tagging is followed by syntax parsing, where Semantria identifies similar sentences presenting different meaning. For example, in the sentences below;

1. *The shop performed poorly because Sam was appointed new manager,*
2. *The shop was doing poorly until Sam was appointed as the new manager.*

The first sentence connotes a negative sentence whiles the second sentence implies a positive sentence. Semantria uses syntax parsing to differentiate between these two sentences. The last stage involves the use of lexical chaining to join individual sentences based on their connection to a larger topic. After processing of the unstructured online reviews, opinion mining was conducted, whereby sentiment scores were assigned to sentiment-bearing tokens. Adverbs and adjectives are deemed to carry sentiments on tokens. Hence, Semantria generates the cumulative sentiment score for each token under various categories such as topics, themes, entities, concepts etc. Under the concept categories, concepts could be created with examples to enhance the lexicon search in a text. Thus, concepts related to hair dryer customers' requirements and their examples were developed to enable Semantria to search for customer concerns in the pool of reviews. Each concept was assigned a weight during the concept generation. Based on the product design concepts generated in (Childs, 2014), a summary of the concepts with corresponding examples were identified and used in Semantria, Table 4.2 as shown below.

Table 0.2 Synonyms of the concepts to enhance lexicons in Semantria

Concepts	**Examples**
Safety	"Fire protection" OR "burning protection" OR" safety."
Quality	"quality" OR "gentle and even heating" OR "Positive switch position."
Appearance	"colour" OR "aesthetics" OR "form."
Price	"low-cost" OR, "expensive" OR "cheap" OR "great value."
Usability	"easy" to use" OR "low storage" OR "low noise
Comfortable to hold	"Hand-grip" OR "weight" OR "balance."
Size	"appropriate "size" OR "lightweight."
Temperature setting	"hot" OR "cold" OR "heat."
Speed setting	" fast" OR " speed" OR
Weight	"mass" OR" load" OR "heaviness" OR "weight."
Noise	"sound" OR "noise" OR "loud."
Robustness	"strong" OR "strength" OR "hard" OR "rough treatment."
Reliability	"accuracy" OR "constancy"
Efficiency	"efficiency"

Semantria was then run, and sentiment scores were generated for each of the customer concerns (concepts) from each review. This was done for all the twenty-two products. Since each hair dryer brands has a different total number of reviews, the number of concerns extracted and analysed for sentiment analysis differ. Moreover, each review from a particular product presents different sentiment scores on the product concepts. As each reviewer has a different level of sentiment regarding the hair dryer of a particular brand, all these sentiments must be aggregated to obtain a value that represents the entire market sentiment on the hairdryer concepts. In generating sentiment scores that generalise the overall perception of customer concerns (concepts), a fuzzy asymmetrical triangular number is defined for each customer concern. The fuzzy number for safety is defined by Equation (4.9) as follows.

$$\tilde{C}_s = (l_s, a_s, r_s) \tag{4.9}$$

where the concerns denote the concepts generated in Table 4.2 $\tilde{C}_s$ is the fuzzy number of the customer concerns on "safety", is $l_s$the left spread for $\tilde{C}_s$, a_s is the centre for $\tilde{C}_s$ and r_s is the right spread for $\tilde{C}_s$. The median is proposed to denote the central values because it is not affected by sentiments that are seen as outliers. The left spread and right spread are represented by the minimum and maximum sentiment score respectively.

According to (McAllister, 1996), a fuzzy number can be defuzzified by the centroid defuzzification method. The triangular fuzzy number in Figure 4.5 is defuzzified according to the Equation (4.10) below:

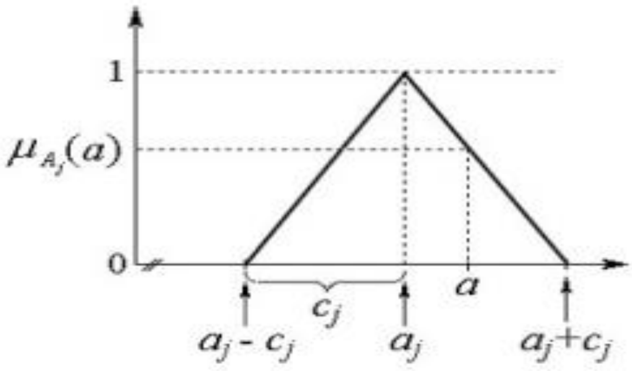

Figure 0.5 Asymmetric triangular fuzzy number

$$x^* = \frac{\int x\, \mu_{\tilde{C}}\,(x)dx}{\int \mu_{\tilde{C}}\ (x)dx} \tag{4.10}$$

where, x^*denotes the crisp value and $\mu_{\tilde{C}}$ represents the fuzzy membership function. In Table 4.3 an example of the review and the associated customer concerns, sentiment score and rating is shown. The sentiment scores for all customer concerns (concepts) of the twenty-two hairdryer brands and their overall ratings are summarized in Table 4.4.

Table 0.3 Examples of customer online reviews and their extracted data and information

Reviewer ID	Review	Extracted concept topic	Sentiment score	Rating
1	“Not worth price paid as does not have power of other 1875W hairdryers. Unfortunately, due to being out of state for unexpected family emergency, I missed time window to return and there is no way to contact the seller that I can see.”	Price	-0.55	1
2	“Good design, love the colour.”	Appearance	0.55	4
3	“After a couple of months, it does not work on high - only blows at the lower speed. The heat options still work.”	Speed setting	-0.57	2

Table 0.4 Data set of sentiment scores of customers concerns

Safety	Quality	Efficiency	Price	Temp_ Settings	Appearance	Usability	Reliability	Comfortable to hold	Robustness	Speed	Weight	Size	Easy control	Rating
-0.245	1.078	0.509	2.303	0.723	0.211	-0.123	0.451	0.200	2.236	0.370	0.229	0.475	0.716	4.2
-0.174	-0.080	0.464	-0.110	0.214	0.394	-0.483	-0.631	-0.489	0.349	-0.087	-0.265	-0.293	0.330	4.4
1.123	1.502	0.584	0.801	1.235	2.182	1.140	0.329	1.267	1.513	0.304	0.882	0.754	1.082	4.3
1.123	1.503	1.543	0.801	1.235	2.182	1.140	0.329	1.267	1.513	0.304	1.199	0.759	1.082	3.9
-0.312	0.452	0.196	1.163	0.973	0.521	0.756	-0.993	0.99	1.564	0.186	-0.094	0.524	0.268	4.1
0.2	2.152	0.732	1.843	1.705	0.996	0.647	0.044	0.320	1.176	-0.049	0.241	0.014	0.547	4.1
0.011	0.106	0.45	1.409	0.907	1.375	0.819	0.019	0.414	0.723	0.527	0.598	0.525	0.819	4
0.392	1.674	0.65	1.572	1.781	0.99	1.092	0.204	1.006	1.541	0.927	0.35	0.636	1.054	4.4
0	1.522	1.63	3.085	2.483	1.663	1.181	0.222	1.258	3.396	1.028	0.529	0.540	1.191	4.4
0.16	1.706	1.113	0.489	0.74	0.746	0.001	0.234	0.207	2.222	0.228	0.187	0.457	-0.219	4.2
0.241	1.24	0.385	2.145	2.334	0.829	0.564	1.099	1.359	1.561	0.113	0.185	0.040	0.377	3.2
0.796	0.506	0.271	0.759	0.659	0.809	0.478	-0.304	0.863	1.253	0.527	0.038	1.624	0.473	3.8
-0.293	0.44	1.907	1.595	1.771	1.141	0.933	0.493	1.231	0.744	0.576	0.523	0.974	0.958	3.4
-0.098	1.158	0.987	2.95	2.099	1.179	1.003	0.067	1.009	3.225	0.285	0.302	0.4	0.940	4.5
0.318	1.365	0.836	0.868	2.013	0.115	0.875	0.207	1.141	1.55	1.393	0.178	0.718	0.816	3.3
0.297	1.639	0.572	1.025	1.015	0.649	0.537	-0.579	0.289	1.513	0.160	0.245	-0.203	0.634	4.2
0.025	0.331	1.678	0.256	1.489	0.566	0.247	0.681	-0.162	1.673	0.055	0.144	-0.050	0.152	3.1
0.342	0.470	0.302	0.539	0.301	0.65	0.657	-0.744	-0.15	2.056	0.350	0.767	-0.512	0.604	3.7
-0.064	1.009	0.754	1.659	0.427	0.792	0.365	1.279	0.586	1.749	0.618	0.729	0.921	0.722	4
-0.245	1.312	1.158	1.921	1.266	1.251	0.321	-0.578	0.603	1.199	0.0518	0.475	0.085	0.598	4.6
0.019	0.055	1.263	1.601	1.157	1.712	0.291	0.397	0.145	0.123	0.022	0.199	0.777	0.291	3.7
0.047	0.694	2.015	0.384	1.989	0.792	1.023	0.639	-0.033	0.32144	0.327	0.395	0.853	1.009	4.3

Fourteen customer concerns were involved in the modelling, $x_k (k = 1,2,3 \ldots,14)$. The variables $x_1 = safety,\ x_2 = quality,\ x_3 = efficiency, x_4 = price,\ x_5 = temp_{setting},\ x_6 = appearance,\ x_7 = usability,\ x_8 = reliability, x_9 = comfortable\ to\ hold\ , x_{10} = Robustness, x_{11} = speed, x_{12} = weight, x_{13=}\ size,\ x_{14} = easy\ controls$. The ratings are the dependent variables in this process. Eighteen cases were used in training the model while the remaining four cases of the datasets were used to test the model. The modelling process was implemented using the MATLAB programming software. The MGGP-FR algorithm was run initially to generate the multigene polynomials. Eighteen cases of the data set were used to generate polynomial structures. An example of the polynomial and interacting terms structures generated is shown in Figure 4.6.

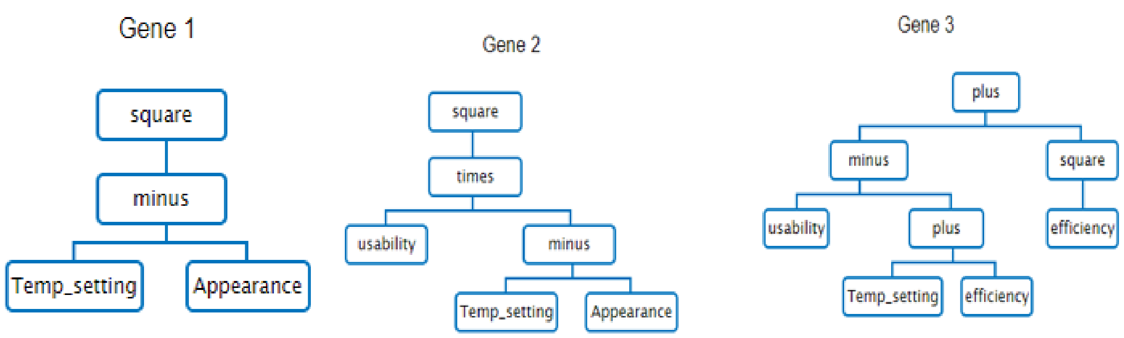

Figure 0.6 An example of a tree structure of individual genes of the MGGP

In the MGGP-FR, the maximum population was set to 250 with a maximum generation of 200. A tournament size of 25 was selected with an elite fraction of 0.7. The probability of a pareto tournament was set at 0.7. The maximum number of genes was set at 3 with a maximum tree depth set at 4. The crossover probability was 0.84, and mutation probability at 0.14. The function sets chosen were, “Times”, “Minus”, Plus” and “Square”. Moreover FR was applied to the non-linear structure generated from the MGGP algorithm to determine the fuzzy coefficients of MGGP-FR model. In this study, the h value, also known as the certain factor in FR analysis, is responsible for controlling the size of a feasible data interval and extending the support of the membership. The FR used different values of h ranging between [0,1], and it was set to 0.5 since it generates the smallest error. To evaluate the effectiveness of the proposed methodology, four validation tests were conducted, and the performance of the proposed approach was compared to three other approaches, namely the FR, GP and the GP-FR method. GP-FR is a method that is used to evolve polynomial and interacting variables using the GP method; the coefficients of the evolved polynomials are then determined using the FR. In the FR method, a fuzzy variation of classical regression analysis is used to develop a customer satisfaction model, when subjective judgements inhibit the crisp measure of a dependent variable. Figure 4.7 shows the variation of the best (log values) and mean fitness with the number of generations. It can be seen from Figure 4.7 that the fitness value decreases with an increasing number of generations. The best fitness was found at the 160th generation (fitness = 0.018146). Figure 4.8 shows the population of the evolved models. In Table 4.5, the customer satisfaction model based on the MGGP-FR was compared to customer satisfaction models based on the FR, GP, and the GP-FR. The MRE and VoE is of the MGGP-FR, FR, GP and the GP-FR is shown Table 4.6.

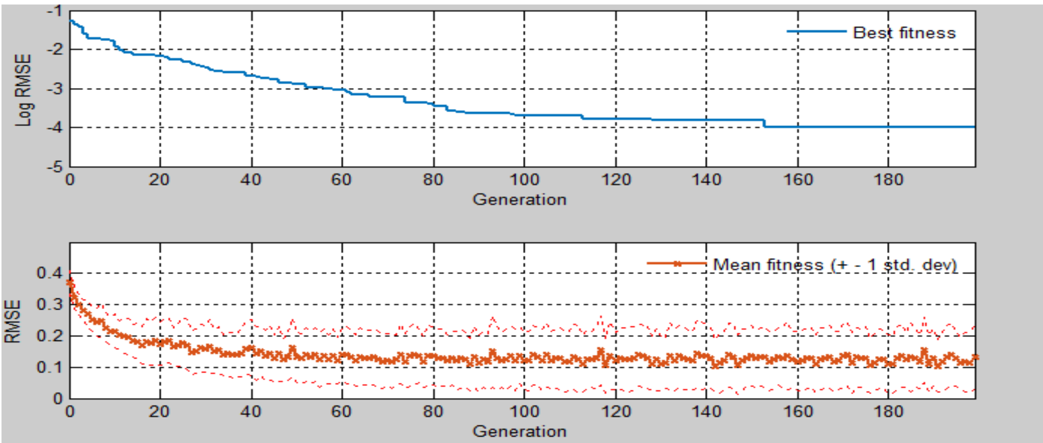

Figure 0.7 Fitness values after 200 generation

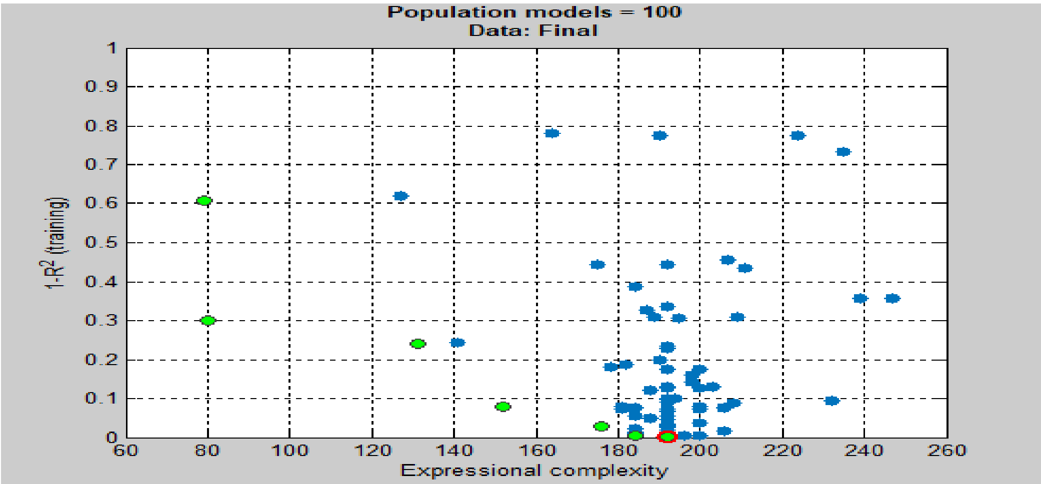

Figure 0.8 Population of the evolved models in terms of their complexity and fitness

Table 0.5 Customer satisfaction models generated from sentiment scores

Algorithms	Generated models
FR	$\tilde{y} = (3.22407, 1.734965 * 10^{-10}) + (-0.81497, 3.76252 * 10^{-10})x_1$ $+(0.521739,0.627565)x_2 + (-0.30186 , 3.23186 * 10^{-9})x_3$ $+(-0.219524,0.147184)x_4 + (-0.3362128,3.91405 * 10^{-10})x_5$ $+(0.883279,4.693 * 10^{-9})x_6 + (-0.1834253,7.9833 * 10^{-10})x_7$ $+(-0.1283631,2.18419 * 10^{-10})x_8 + (-0.2051243, ,4.461 * 10^{-10})x_9$ $+(0.245492,1.719 * 10^{-9})x_{10} + (0.0361108, 5.3748 * 10^{-10})x_{11}$ $+ (-1.018843, 4.668 * 10^{-10})x_{12} + (0.138820,2.28 * 10^{-10})x_{13}$ $+ (0.9950397,8.150 * 10^{-10})x_{14}$
GP	$\tilde{y} = 3.788476 + 0.028388\,(\big(\big((x_2 + x_5) * x_5\big) * (x_6 - x_8) * (x_7 + x_2)\big)$ $- 0.3788476x_8$
GP-FR	$\tilde{y} = (3.9368 , 4.6943) + (-0.5285 , 0)\,(x_9x_8 + x_3x_8)$ $+(0.2029, 0.0000)\, {x_{12}}^2x_4$
MGGP-FR	$\tilde{y} = (3.9174, 0.3570) + (0.5987, 8.7842 * 10^{-23}){x_{14}}^4$ $+(-0.5474, 6.6743 * 10^{-23})x_5x_9 + (0.5507, 9.092 * 10^{-23})x_4x_9$ $+(0.2937, 7.125 * 10^{-23})x_2 + (-0.3553, 0.0656)x_8$ $+(-0.3288, 6.210 * 10^{-23})x_5 + (-0.2136, 8.095 * 10^{-23})x_{10}x_{12}$

Table 0.6 Performance of models generated from sentiment scores data

Errors		FR	GP	GP-FR	MGGP-FR
Training errors	Product A	0.0242	0.1407	0.0458	0.0336
	Product B	0.0016	0.0847	0.1067	0.0543
	Product C	0.0247	0.0388	0.1090	0.0528
	Product D	0.0272	0.0598	0.0405	0.0180
	Product E	0.0111	0.0333	0.1035	0.0303
	Product F	0.0396	0.0413	0.0199	0.0262
	Product G	0.0069	0.0466	0.0128	0.0158
	Product H	0.0291	0.0702	0.1067	0.0307
	Product I	0.0320	0.0301	0.0091	0.0528
	Product J	0.0272	0.1083	0.0917	0.0532
	Product K	0.0342	0.0179	0.0468	0.0805
	Product L	0.0113	0.0336	0.0859	0.0595
	Product M	0.0150	0.0885	0.0454	0.0583
	Product N	0.0258	0.0601	0.0115	0.0482
	Product O	0.0298	0.1122	0.1423	0.0674
	Product P	0.0281	0.0028	0.0184	0.0216
	Product Q	0.0079	0.1372	0.0979	0.0778
	Product R	0.0081	0.1029	0.1269	0.0658
	MRE of training error	0.0213	0.0672	0.0678	0.0470
	VOE of training errors	$1.2336* 10^{-4}$	0.0017	0.0020	$4.0555* 10^{-4}$
Prediction errors	Product S	0.0076	0.1770	0.1350	0.0231
	Product T	0.0592	0.0686	0.0211	0.0723
	Product U	0.0047	0.0118	0.0136	0.0682
	Product V	0.2014	0.1661	0.2379	0.1014
	MRE of prediction error	0.0682	0.1059	0.1019	0.0663
	VOE prediction errors	0.0085	0.0063	0.0113	0.0010

4.7 Validation

The performance of the proposed method for modelling customer satisfaction is validated by performing validation tests to determine the training and prediction error of the proposed MGGP-FR, the FR, GP, and the GP-FR. The k-fold validation method with k=4 was employed to validate the performance of the proposed. In each validation, eighteen datasets of the hair dryer product were randomly selected to train the FR, GP, GP-FR, and the MGGP-FR model. The remaining four datasets were used to evaluate the prediction accuracy. Table 4.7 describes the experimental plan for conducting the validation tests.

Table 0.7 Experimental plan used for validation of the four approaches

Test set	Validation test no.	Defuzzified sentiment scores of the products which are used as training data	Rating of CS of the product to be predicted based on the generated model
I	1	A, B, C, …, R	S
	2	A, B, C, …, R	T
	3	A, B, C, …, R	U
	4	A, B, C, …, R	V
II	5	E, F. G, …, V	A
	6	E, F. G, …, V	B
	7	E, F. G, …, V	C
	8	E, F. G, …, V	D
III	9	A, B, C, …, V	K
	10	A, B, C, …, V	L
	11	A, B, C, …, V	M
	12	A, B, C, …, V	N
III	13	A, B, C, …, V	P
	12	A, B, C, …, V	Q
	15	A, B, C, …, V	R
	16	A, B, C, …, V	S

The CS models generated from final validation test for each validation set is shown in Tables 4.8 to 4.11. It can be seen that all the CS models based on the FR tend to be linear, but the CS models based on GP, GP-FR and MGGP-FR involves higher-order and interaction terms. Only the GP has no fuzzy coefficients compared to the FR, GP-FR, and the MGGP-FR. The performance results namely the prediction errors, MRE and the VoE. of the first, second, third and fourth validation tests are shown in Tables 4.12, 4.13, 4.14 and 4.15. From the validation performance results, the MRE of the MGGP-FR was the least compared to the MRE of FR, GP, and GP-FR in all the four validation tests. For the VoE, the MGGP-FR had less VoE compared to the other three approaches in validation test 1, 2 and 4. Further for the VoE for validation test 3, the VoE of the MGGP-FR was less than the GP and GP-FR but slightly higher than FR.

The MRE and VoE of the four approaches under the four validation tests are summarized in Figure 4.9 and Figure 4.10, respectively. From Figure 4.9 the proposed MGGP-FR had the least MRE errors from first and second validation tests conducted. In the third validation test, the MGGP-FR had the second least MRE error after the FR. Similarly, in the fourth validation test, the MGGP-FR had the second least MRE error after the GP. In Figure 4.10, the MGGP-FR had the least VOE when compared with FR, GP, and the GP-FR.

In order to determine the significance differences between prediction performance between the proposed MGGP-FR and the other three approaches, a two-sample t-test was conducted. The results of the t-test showed that there were significant differences between the MGGP-FR and the other three methods in terms of the MRE errors and VOE.

Table 0.8 Models generated from validation dataset 1

Algorithms	Generated models
FR	$\tilde{y} = (2.69418, 5.148 * 10^{-16}) + (-1.5342, 7.657 * 10^{-16})x_1$ $+(0.600107, 1.63706 * 10^{-16})x_2 + (-0.92925\ , 0.35248)x_3$ $+(-0.9638543, 1.5779 * 10^{-16})x_4 + (0.577187, 2.548 * 10^{-16})x_5$ $+(1.23319, 0.877348)x_6 + (-0.88612, 4.14537 * 10^{-16}\ )x_7$ $+(-0.242427, 4.16184 * 10^{-16})x_8 + (-0.310344, 1.735 * 10^{-16})x_9$ $+(0.4496734, 8.2799 * 10^{-17})x_{10}$ $+ (-0.56292, 2.32851 * 10^{-16})x_{11}$ $+(0.100540, 8.369 * 10^{-16})x_{12} + (0.8021364, 4.133642 * 10^{-16})x_{13}$ $+(1.59059, 3.223 * 10^{-16})x_{14}$
GP	$y = 3.857866 - 0.867910(x_{12} - (\ x_2 + x_{14})x_{12}) + 0.457820(x_7 - x_5)x_8$
GP-FR	$\tilde{y} = (3.611575\ , 5.186783) + (\ 0.178491\ , 0\ )x_1$ $+(\ 0.363831\ , 0.000000\ )\ x_6 + (\ -0.378673\ , 0)x_8$
MGGP-FR	$\tilde{y} = (4.4939, 1.887 * 10^{-14}) + (-0.4846, 0.522)x_1$ $+(-0.4490, 1.002 * 10^{-14})x_2 + (1.149, 7.046 * 10^{-15})x_5$ $+(0.2772, 2.4075 * 10^{-13})x_6 - (-0.7998, 1.7949)x_5x_{14}$ $+(-0.6287, 1.552)x_{11} + (-0.5994, 1.4458)x_{12}$

Table 0.9 Models generated from validation dataset 2

Algorithm	Generated models
FR	$\tilde{y} = (2.78084, 5.7271 * 10^{-16}) + (-1.5045, 1.6509 * 10^{-15})x_1$ $+(0.726184, 9.81609 * 10^{-16})x_2 + (-0.15260, 5.21733 * 10^{-16})x_3$ $+(-0.524827, 8.70464 * 10^{-16})x_4 + (-0.192999, 1.227 * 10^{-15})x_5$ $+(0.971985, 6.659 * 10^{-16})x_6 + (-0.248319, 1.5218 * 10^{-15})x_7$ $+(-0.431949, 6.396 * 10^{-16})x_8 + (-0.54475, 1.4104 * 10^{-15})x_9$ $+(0.24893, 3.9634 * 10^{-16})x_{10} + (0.12509, 1.3793)x_{11}$ $+ (-0.54696, 1.1542 * 10^{-15})x_{12} + (0.679975, 8.054 * 10^{-16})x_{13}$ $+(0.679975, 1.327 * 10^{-15})x_{14}$
GP	$\mathrm{y} = 3.759547 + 0.140811x_2 - 0.278392x_8 + 0.340826x_{12}x_4$
GP-FR	$\tilde{y} = (3.746218, 6.529828) + (0.079660, 0.000000)\, x_2 * x_4$
MGGP-FR	$\tilde{y} = (4.0798, 0.2240)(0.6396, 3.898 * 10^{-17})x_9x_{10}$ $+(-0.7057, 2.734 * 10^{-16})x_5x_9$ $+(0.3200, 4.445 * 10^{-17})x_{14} + (0.4270, 3.792 * 10^{-17})x_{10}$ $+(-0.4193, 7.3467 * 10^{-17})x_{10}x_3 + (-0.3081, 1.592 * 10^{-16})x_8$ $+(0.4417, 0.0941)x_2$

Table 0.10 Models generated from validation dataset 3

Algorithm	Generated models
Fuzzy regression (FR)	$\tilde{y} = (2.6530, 7.1124 * 10^{-12}) + (-1.2489, 4.4232 * 10^{-11})x_1$ $+(0.4854, 1.1387)x_2 + (-0.738539, 8.3367 * 10^{-12})x_3$ $+(-0.57961, 4.9960 * 10^{-12})x_4 + (0.6639, 1.07168 * 10^{-11})x_5$ $+(0.926059, 8.257704 * 10^{-12})x_6 + (-0.7834, 2.270 * 10^{-11})x_7$ $+(-0.384146, 1.05765 * 10^{-11})x_8 + (-0.67408, 1.49732 * 10^{-11})x_9$ $+(0.4062914, 3.655 * 10^{-12})x_{10} + (-0.36150, 1.5462 * 10^{-10})x_{11}$ $+(-0.1205, 2.9058 * 10^{-11})x_{12} + (0.8086, 9.41658 * 10^{-12})x_{13}$ $+(1.3039, 1.9879 * 10^{-11})x_{14}$
GP	$\tilde{y} = 4.156304 - 0.574141x_9 + 0.132889(x_4 + x_6)x_{14}x_2$
GP-FR	$\tilde{y} = (4.040545, 2.171688) + (-0.165660, 1.703708)\, x_5x_9$ $+(0.212090, 0.000000)x_2 + (-0.198910, 0)\, x_8$
MGGP-FR	$\tilde{y} = (-4.0591, 0.30140) + (0.4619, 1.6820 * 10^{-11})x_3{}^2x_8$ $+(0.0867, 0.0046)x_5{}^3x_6x_{14} + (0.0724, 1.290 * 10^{-12})x_5{}^3x_6x_{14}$ $+(0.4351, 1.4338 * 10^{-11})x_5{}^2x_6x_8{}^2$ $+(0.3711, 1.263 * 10^{-11})x_5{}^2x_8{}^2x_9$

Table 0.11 Models generated from validation dataset 4

Algorithm	Generated models
FR	$\tilde{y} = (2.9451232, 5.741 * 10^{-14}) + (-1.9218,\ 2.78183)x_1$ $+(0.78412, 3.98260 * 10^{-14})x_2 + (-1.04858\ , 6.7150 * 10^{-14})x_3$ $+(-0.701930, 2.291 * 10^{-14})x_4 + (0.53469, 3.70929 * 10^{-14})x_5$ $+(1.1083071, 4.7049 * 10^{-14})x_6 + (-0.3716177, 7.495299 * 10^{-14})x_7$ $+(-0.24027, 1.37739 * 10^{-13})x_8 + (-0.72931, 9.071 * 10^{-14})x_9$ $+(0.297091, 2.0546 * 10^{-14})x_{10} + (-0.38032, 9.46787 * 10^{-14})x_{11}$ $+(-0.073570, 1.6676 * 10^{-13})x_{12} + (1.1818747, 1.10920 * 10^{-13})x_{13}$ $+(1.1818747, 7.28022 * 10^{-14})x_{14}$
GP	$\tilde{y} = 3.838973 + 0.274737(x_{14}^2 x_5 - x_8 x_5)$
GP-FR	$\tilde{y} = (3.898583\ , 5.758779) + (\ -0.342934\ , 0)x_8$ $+(\ 0.295162\ , 0\ )x_{12}x_4$
MGGP-FR	$\tilde{y} = (3.8609, 9.2542 * 10^{-18}) + (0.2209, 4.238 * 10^{-19})x_3x_2$ $+(-0.4013, 4.598 * 10^{-19})x_4 + (-0.3808, 4.070911 * 10^{-18})x_5$ $+(-0.5350, 7.4508 * 10^{-19})x_3x_9 + (-0.1422, 0.7703)x_8$ $+(-0.49113, 8.80047 * 10^{-19}){x_7}^4$

Table 0.12 Performance of models generated from validation dataset 1

Errors		FR	GP	GP-FR	MGGPFR
Training errors	Product E	0.0128	0.0294	0.0385	0.0187
	Product F	0.0276	0.0224	0.0589	0.0519
	Product G	0.0341	0.0454	0.0308	0.0551
	Product H	0.0249	0.0185	0.0469	0.0927
	Product I	0.0034	0.0067	0.0195	0.0485
	Product J	0.0249	0.0814	0.0240	0.0111
	Product K	0.0270	0.0416	0.1621	0.0617
	Product L	0.0212	0.0217	0.0820	0.0271
	Product M	0.0492	0.1322	0.1526	0.1163
	Product N	0.0307	0.0863	0.0618	0.0680
	Product O	0.0120	0.1918	0.1572	0.0903
	Product P	0.0183	0.0131	0.0379	0.0396
	Product Q	0.0351	0.0988	0.1673	0.0119
	Product R	0.0184	0.0234	0.1393	0.0377
	Product S	0.0240	0.0712	0.1012	0.0319
	Product T	0.0327	0.0253	0.0174	0.0473
	Product U	0.0526	0.0305	0.1065	0.0244
	Product V	0.0327	0.1125	0.1206	0.0977
	MRE	0.0268	0.0585	0.0847	0.0518
	VOE	$1.486 * 10^{-4}$	0.0025	0.0029	$9.5709 * 10^{-4}$
Prediction errors	Product A	0.0523	0.0853	0.1166	0.0402
	Product B	0.0005	0.0379	0.0958	0.0620
	Product C	0.2566	0.1761	0.0578	0.0329
	Product D	0.1661	0.4087	0.1663	0.0175
	MRE	0.1189	0.1770	0.1091	0.0381
	VOE	0.0132	0.0271	0.0020	$3.4185 * 10^{-4}$

Table 0.13 Performance of models generated from validation dataset 2

		FR	GP	GP-FR	MGGP-FR
Training Errors	Product A	0.0122	0.0557	0.0609	0.0457
	Product B	0.0027	0.1060	0.1484	0.0617
	Product C	0.0098	0.0418	0.1065	0.0034
	Product D	0.0108	0.0787	0.0149	0.0191
	Product E	0.0052	0.0092	0.0761	0.0393
	Product F	0.0017	0.0248	0.0092	0.0296
	Product G	0.0182	0.0141	0.0605	0.0280
	Product H	0.0290	0.0622	0.1009	0.0672
	Product I	0.0322	0.0005	0.0636	0.0904
	Product J	0.0075	0.0558	0.0922	0.0647
	Product O	0.0582	0.1961	0.1638	0.0900
	Product P	0.0053	0.0088	0.0762	0.0715
	Product Q	0.0025	0.1707	0.2106	0.0254
	Product R	0.0131	0.1281	0.0180	0.0476
	Product S	0.0012	0.0104	0.0301	0.0437
	Product T	0.0016	0.0400	0.1420	0.0148
	Product U	0.0008	0.0179	0.0144	0.0742
	Product V	0.0105	0.1323	0.1238	0.0692
	MRE of error	0.0123	0.0641	0.0840	0.0492
	VOE	$2.1360 * 10^{-4}$	0.0035	0.0033	$6.6703 * 10^{-4}$
Prediction errors	Product K	0.3934	0.1762	0.2369	0.0072
	Product L	0.0064	0.0330	0.0061	0.1351
	Product M	0.1513	0.1673	0.1183	0.1105
	Product N	0.0841	0.0651	0.1070	0.0594
	MRE	0.1588	0.1104	0.1171	0.0781
	VOE	0.0280	0.0052	0.0089	0.0032

Table 0.14 Performance of models generated from validation dataset 3

Errors		FR	GP	GP-FR	MGGP-FR
Training Errors	Product A	0.0240	0.0237	0.0108	0.0647
	Product B	0.0023	0.0082	0.0820	0.0475
	Product C	0.0398	0.0527	0.0465	0.0258
	Product D	0.0439	0.0445	0.0513	0.0417
	Product E	0.0126	0.1183	0.0300	0.0585
	Product F	0.0598	0.0773	0.0748	0.0937
	Product G	0.0030	0.0123	0.0002	0.0003
	Product H	0.0433	0.0502	0.0684	0.0781
	Product I	0.0394	0.0403	0.1260	0.0751
	Product J	0.0462	0.0533	0.0421	0.0824
	Product K	0.0441	0.1127	0.1806	0.0962
	Product L	0.0152	0.0234	0.0667	0.0811
	Product M	0.0147	0.0597	0.1096	0.0741
	Product N	0.0250	0.0723	0.1255	0.020
	Product O	0.0471	0.1050	0.1968	0.0208
	Product T	0.0325	0.0997	0.0886	0.0087
	Product U	0.0017	0.1028	0.0877	0.0558
	Product V	0.0184	0.0035	0.0236	0.0177
	MRE	0.0285	0.0589	0.0784	0.0528
	VOE	$3.1840 * 10^{-4}$	0.0014	0.0029	$9.0689 * 10^{-4}$
Prediction errors	Product P	0.0162	0.0051	0.0332	0.1183
	Product Q	0.0843	0.3725	0.3389	0.1086
	Product R	0.0305	0.1588	0.1210	0.0022
	Product S	0.0916	0.0144	0.0533	0.0842
	MRE	0.0557	0.1377	0.1366	0.0783
	VOE	0.0014	0.0295	0.0196	0.0028

Table 0.15 Performance of models generated from validation dataset 4

Errors		FR	GP	GP-FR	MGGP-FR
Training errors	Product A	0.0163	0.0830	0.0171	0.0745
	Product B	0.0111	0.1176	0.1887	0.0995
	Product C	0.0726	0.0409	0.0816	0.0019
	Product D	0.0801	0.0575	0.0521	0.0154
	Product I	0.0000	0.0212	0.2029	0.0210
	Product J	0.0106	0.0949	0.0050	0.0386
	Product K	0.0210	0.0080	0.0620	0.1100
	Product L	0.0582	0.0354	0.0721	0.0233
	Product M	0.0240	0.1899	0.0018	0.1005
	Product N	0.0018	0.0421	0.2512	0.0103
	Product O	0.0268	0.2402	0.0331	0.0434
	Product P	0.0197	0.0209	0.0082	0.0173
	Product Q	0.0022	0.1516	0.0155	0.1516
	Product R	0.0257	0.0624	0.1071	0.1394
	Product S	0.0044	0.0625	0.0099	0.0307
	Product T	0.0008	0.0946	0.0902	0.0872
	Product U	0.0014	0.0108	0.0659	0.062
	Product V	0.0030	0.0589	0.0049	0.0581
	MRE of error	0.0211	0.0773	0.0705	0.0603
	Variance of errors	$6.1139 * 10^{-4}$	0.0040	0.0056	0.0021
Prediction errors	Product E	0.0641	0.0058	0.1054	0.0180
	Product F	0.1526	0.0344	0.3809	0.0374
	Product G	0.0787	0.0003	0.0224	0.0525
	Product H	0.0067	0.0267	0.1290	0.0058
	MRE	0.0755	0.0168	0.1594	0.0285
	VOE	0.0036	$2.6759 * 10^{-4}$	0.0239	$4.2699 * 10^{-4}$

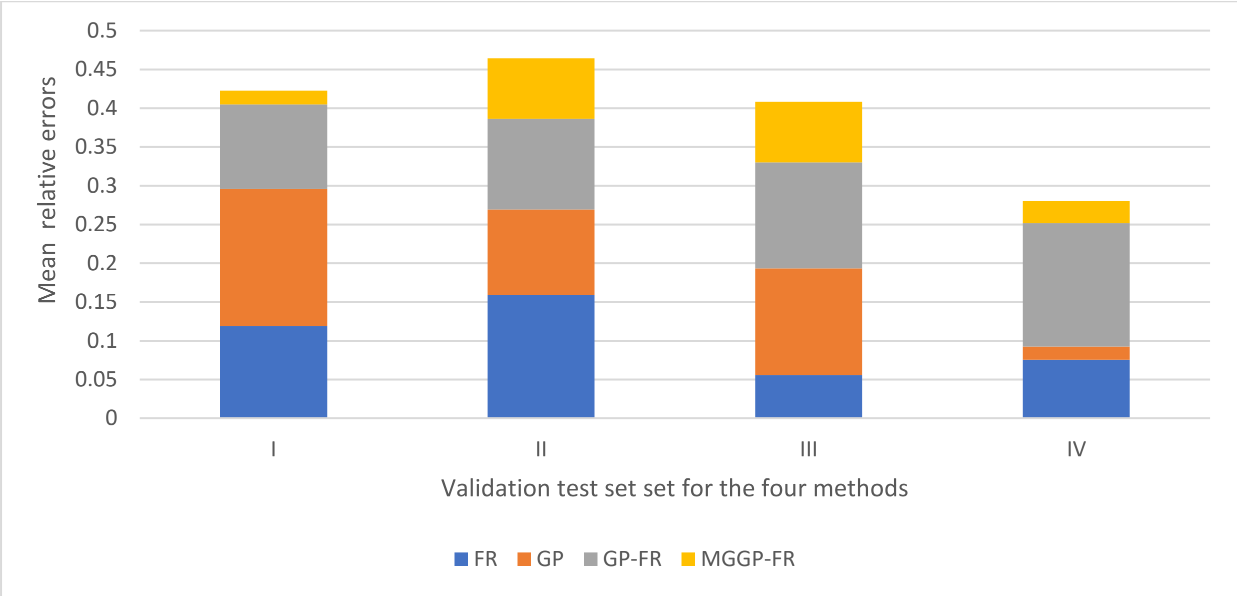

Figure 0.9 Mean relative error of the four test sets

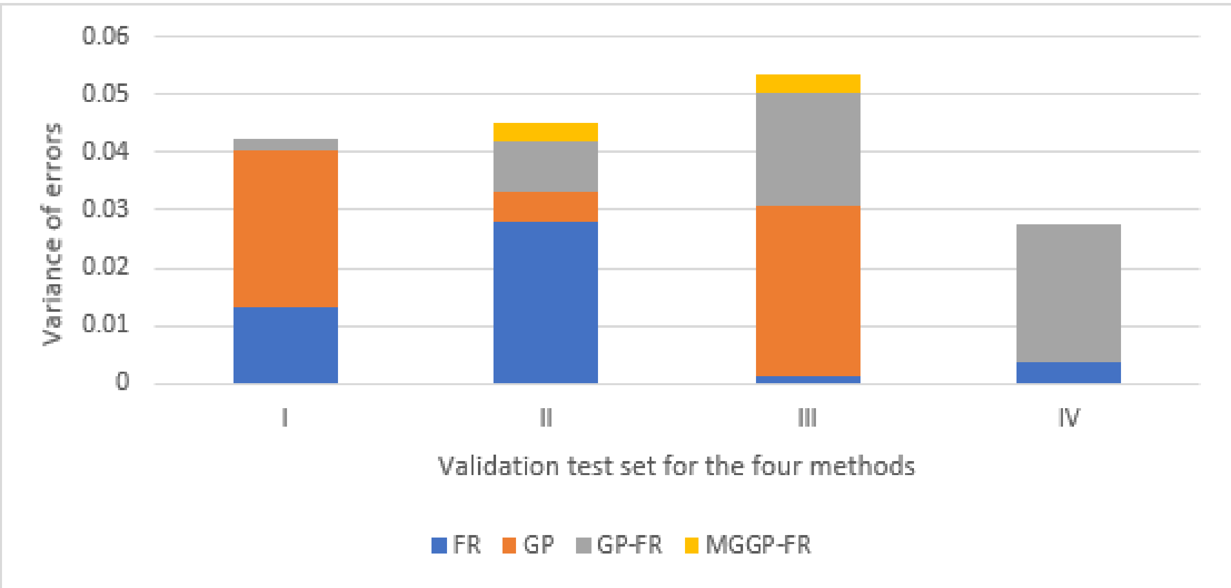

Figure 0.10 Variance of error of the four test sets

A two sample t-test was conducted to find the significance differences in the prediction performance using Equation (4.11). The null hypothesis established was that there were no significance differences between the proposed MGGP-FR and the other three approaches. At the α level set at 0.05, and critical value at 2.776, the t-values were determined for prediction errors and variance of error between the proposed method and the other three approaches. Table 4.16 shows the results of the t-test. From Table 4.16, the t-values for both the prediction errors and the variance errors were all greater than 2.776 indicating significant differences between the MGGP-FR and the other three methods.

$$t - value = \frac{\mu_1 - \mu_2}{\frac{\sqrt{var_1 + var_2}}{N}} \tag{4.11}$$

Table 0.16 t-values of prediction errors

	t-value for prediction errors	t-value for variance of prediction errors
t-Test between FR and MGGP-FR	3.604	4.588
t-Test between GP and MGGP-FR	3.0507	8.289
t-Test between GP-FR and MGGP-FR	3.8712	5.153

4.8 Summary

In this chapter, a new MGGP-FR is proposed to model customer satisfaction is presented. The symbolic regression of the MGGP was employed to generate polynomial structures of higher-order and interaction terms. The FR was then used to determine the fuzzy coefficients of the generated fuzzy polynomials by solving linear programming problems.

A case study on developing customer satisfaction model for electric hairdryers was conducted. In order to evaluate the effectiveness of the proposed method, four validation tests were conducted, and the performance of the validation test were compared to the FR, GP and GP-FR. The validations results showed that the proposed MGGP-FR had better prediction performance in terms of the MRE and the VoE.

Chapter 5. Forecasting the importance of product attributes using online customer reviews and Google Trends

With dynamic consumers needs and rapid changes in technologies, keeping up with consumers' needs and opinions has never more challenging than before (Ying et al., 2018). User-generated content in the form of online reviews presents an opportunity for product manufacturers to understand consumers' dynamic needs through consumers' perspectives. Moreover, this study considers consumer online search behaviours that have not been extensively applied in product development research to identify the relevant product attributes for the present and future consumers’ needs.

In this chapter, a proposed methodology for estimating the importance and forecasting the future importance of product attributes from online reviews and Google Trends is demonstrated. The approaches for estimating the importance of product attributes is described in section 5.1 whiles the methodology for predicting the future importance of product attributes is presented in section 5.2. Section 5.3 describes the implementation of the proposed method using an electric hairdryer as a case study. The results and validation of the proposed method are presented in section 5.4. Finally, a summary of this chapter is presented in section 5.5. The methodology for determining the importance and future importance is shown in Figure 5.1 below.

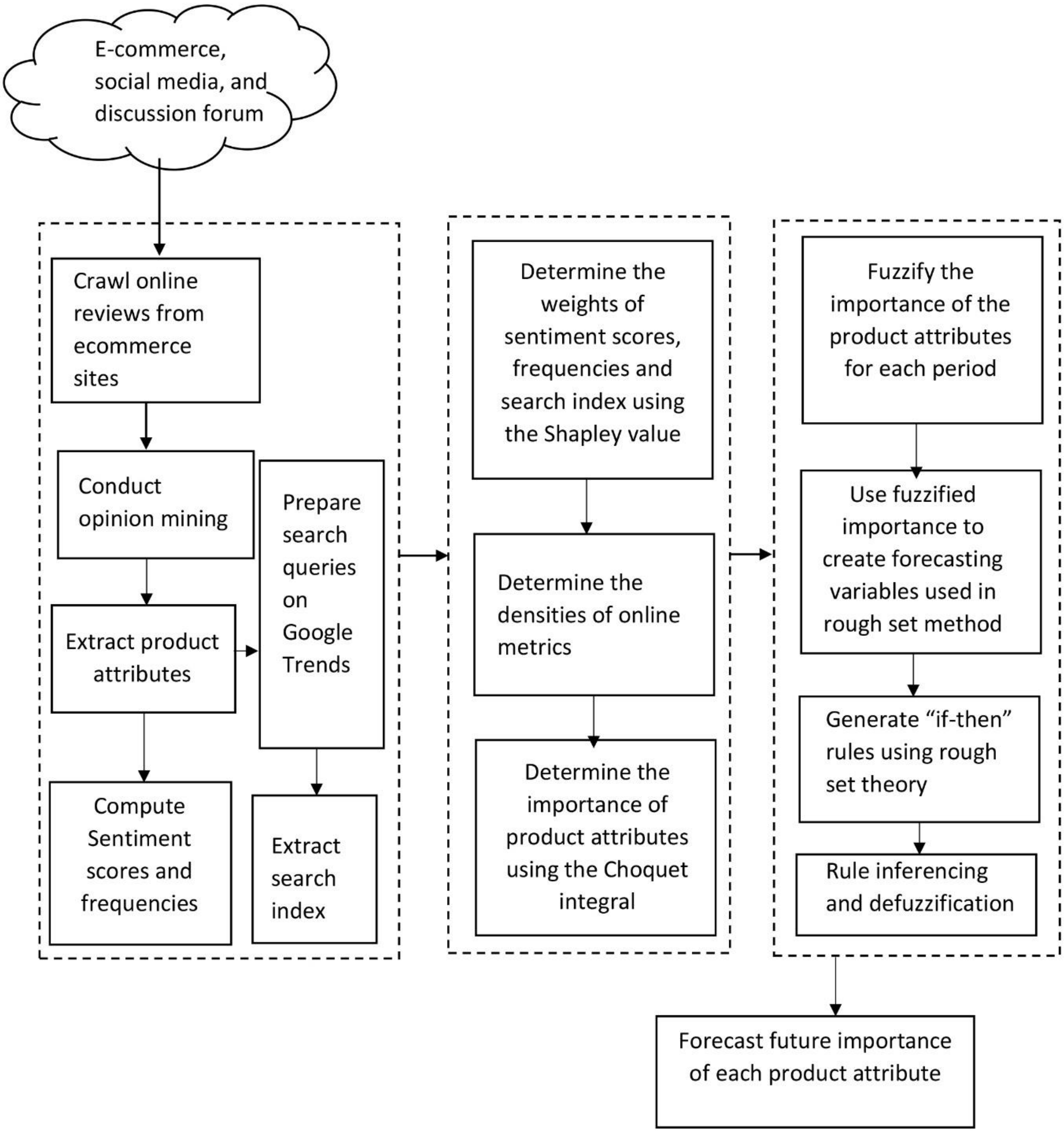

Figure 0.1 The framework for determining the importance and future importance of product attributes

5.1 Determining the importance and future importance of product attributes

As product manufacturers are moving towards data integrated product design, the conventional methods of determining the most significant product attributes that will appeal to the dynamic needs of consumers is becoming more prominent than ever in the face of globalisation. Thus, to address the limitations of past studies, as mentioned in section 2.6, this study employs the Shapley value and Choquet integral for determining the importance of product attributes using online reviews and the Google Trends index. In predicting the future importance of product attributes, the fuzzy rough set time series is used to predict the future importance of product attributes of an electric hair dryer.

5.1.1 Shapley value

This study employs three main metrics to estimate the importance of product attributes. The metrics are namely sentiment scores from online reviews, frequencies from online reviews, and Google search indexes. The concept of the Shapley value works to ensure the distribution of benefits and cost across multiple actors working in coalition (von Neumann and Morgenstern, 2007). The Shapley value is implemented in scenarios with unequal contribution of each player in any situation, but all players cooperate to achieve the desired outcome. The three online metrics in this studies are used as players in the Shapley value estimation. The weight is each online metric is initially determined by conducting a survey among particular experts. The survey allows the experts to rate the significance of online metric in the determination of the importance of product attributes. The weight obtained for each online metric as well as the weight of interaction among the three online metrics, obtained from the survey, is determined by finding the average of the ratings given by the experts. The weight obtained is then used to determine the Shapley value of the three online metrics. The Shapley value (S_k) (Besner, 2019) is described by the Equations (5.1) and (5.2) below:

$$S_k = \sum_{M \subset N\{k\}} \gamma_n \, M[v\,(M \cup \{k\}) - v(M)] \qquad (5.1)$$

where the weight $\gamma_n(M)$ of entering a coalition is defined as

$$\gamma_n(M) = \frac{m!\,(n-1-m)!}{n!} \qquad (5.2)$$

In both Equation (5.1) and Equation (5.2), n refers to the total number of all players in a game and m is the subset of n. M refers to the players in the coalition and $v(.)$ represents the value function determining the utility of each coalition. It is the marginal contribution of player k to the coalition M. $M \cup \{k\}$ represents the kth players involved in the coalition. The Shapley value and the average Shapley values are determined using Equation (5.1). The Shapley values obtained from the online metric estimated is next used in the Choquet integral to determine the importance of the product attributes. In this study, the Shapley values are used as fuzzy densities in the Choquet integral.

5.1.2 Choquet integral

The Choquet integral is a function that aggregates data whiles considering the fuzzy measure. The fuzzy measure refers to the function set on all possible combinations of a set of criteria (Beliakov et al., 2020). Fuzzy measures have been useful in areas where information on the value of a variable is uncertain, and it has been applied in decision making where there are different sources of uncertain information (Yager, 2016). It is the generalisation of the classical measure through the use of non-additivity, where additivity is replaced by the monotonicity for fuzzy measures, allowing the fuzzy measure to determine the importance of each attribute. In classical measures, the attribute (A_1, A_2) is demonstrated as:

$$\mu(A_1, A_2) = \mu(A_1) + \mu(A_2) \tag{5.3}$$

where μ represents the weight of the criteria. Equation (5.3) does not represent the fuzzy measure. In fuzzy measures, two results could be obtained as shown in equations 5.4 and 5.5 below:

$$(a) \quad \mu(A_1, A_2) < \mu(A_1) + \mu(A_2) \tag{5.4}$$

$$(b) \quad \mu(A_1, A_2) > \mu(A_1) + \mu(A_2) \tag{5.5}$$

(a) is also known the substitutive effect while (b) is known as the complementary effect. Thus, the importance of considering the sum of all possible combinations of attributes is more significant than considering the sum of individual attributes.

Definition 1(Wang and Klir, 1992)***:***

In fuzzy measures, let $X = \{x_1, x_2, \dots, x_n\}$ be a set of attributes and let $P(X)$ be the subsets of X, which can also be represented as the power set of X, thus 2^x. A fuzzy measure g is a set function defined on $P(X)$ is denoted by $P(X) \rightarrow [0,1]$ with the following properties represented by Equations (5.6) and (5.7) and (Shiau and Lee, 2017) :

(1) $g(\emptyset) = 0, \;\; g(X) = 1;$ (5.6)

(2) $A \subset B \subset X \; implies \; g(A) \leq g(B)$ (5.7)

Thus, to determine the regular fuzzy measure in the space of n elements, $(2^n - 2)$ coefficients are needed. The complexity of the fuzzy measures also increases with an increase in the number of attributes (Wang *et al*., 2011). The concept and framework applied in a probabilistic environment is used in the context of measures in the presence of uncertain information.

Definition 2 (Wang and Klir, 1992)***:***

Let X be a finite set and 2^xbe the power set of X. If a fuzzy measure $g: 2^x \rightarrow [0,1]$ satisfies the following conditions represented by Equations (5.8) and (5.9):

(1) $g(\emptyset) = 0, g(X) = 1;$ (5.8)

(2) $g(A \cup B) = g(A) + g(B) + \lambda . g(A) . g(B)$ (5.9)

$$\forall A, \;\; B \subset X, \;\; A \cap B = \emptyset, \lambda \in (-1, \infty)$$

g is called the regular λ-fuzzy measure defined on 2^x. Condition (2) of definition 2 is called the λ-rule. When $\lambda = 0$, the λ-rule generates an additivity for the classical measure. For $X = \{x_1, x_1, \ldots, x_n\}$ the value of $g_i = g((\{x\})$ is known as the measure density.

The parameter λ of a regular λ-fuzzy measure is determined from the Equation (5.10) (Wang and Klir, 2013).

$$\prod_{i=1}^{n}(1+\lambda g_i) = 1+\lambda \tag{5.10}$$

Let set $X = \{x_1, x_2, \ldots, x_n\}, \lambda \neq 0$ then:

$$g(E) = \frac{1}{\lambda}\left[\prod_{x_i \epsilon E}^{n}(1+\lambda g_i) - 1\right], \forall E \subset X \tag{5.11}$$

The values of the λ-fuzzy measures of the individual attributes can be obtained in theorem 1 to obtain the values of λ. Using the λ obtained from theorem 2, the fuzzy measures of other sets can be determined. The λ-fuzzy measure is also determined using the fuzzy measure densities. Moreover, the λ-fuzzy measure must also satisfy the boundary conditions below:

1) If there exists some g_i=1, then $g_j = 0$ for any j≠i.
2) If $g_i < 1$ for all g_i, then there exist at least two of them positive.

If the measured density g_i on finite set X = $\{x_1, x_2, \ldots, x_n\}$, then there is only one solution λ obtained from (5.1).

5.1.3. Fuzzy integrals

Mathematically, the fuzzy integral generalises the classical Lebesgue integral and is employed for aggregation purposes in MCDM. Many fuzzy integrals have been proposed, and as seen in (Vicen *et al*., 2007; Wang and Klir, 2013), two of the fuzzy integrals popularly used in the literature is the Sugeno integral and the Choquet integral. The concept of the fuzzy measure and fuzzy integral was introduced by Sugeno who replaced the additive characteristics of classical measures with a non-additive monotonic requirement. Grabisch, (1996) compared the fuzzy integrals, the Sugeno and Choquet integrals, according to their characteristics. The Sugeno integral is based on the Min and Max operators. On the contrary, the Choquet integral is based on linear operators and is more commonly applied in real problem applications. One feature of the Choquet integral is that it is similar to the expected value in a measure-based scenario. Thus the Choquet integral can be used to derive the expected value for any function of a measure type with an uncertain variable (Yager, 2016).

Definition 3 (Wang and Klir, 1992)***:***

Let X be a finite set, 2^X be the set of X, $f: X \rightarrow [0, \infty]$, with g a regular fuzzy measure defined on 2^X. Then the Sugeno integral (expressed in Equation (5.12) and the Choquet integral (expressed in Equation (5.13) of function f with respect to g are defined as shown in the following :

$$\text{(Sugeno integral)} = \int f d = \bigvee_{i=1}^{n} (f(x_i) \wedge g(A_i)) \qquad (5.12)$$

$$\text{(Choquet integral)}$$

$$= \int f d\, g = (\ f(x_i) - f(x_{i-1}). g(A_i) \qquad (5.13)$$

where $0 \leq f(X_1) \leq f(X_2) \leq \cdots f(X_n) \leq 1,\ A_i = \{X_i, X_{i+1}, \ldots, X_n\}, f(X_o = 0, g(A_{n+1}) = 0.$

From Definition 3, the Sugeno integral does not have any relation with the Lebesgue integral, but the Choquet integral has a relation with the Lebesgue integral. This phase of the research employs the Choquet integral in the aggregation process in estimating the importance of the product attributes

The fuzzy densities of the three online metrics are obtained for each product attribute identified from the online reviews. Based on these fuzzy densities, the fuzzy measures for the three online metrics are calculated using Equation (5.1). The Choquet integral (fuzzy integral) for each product attribute is estimated in order to determine the importance of the product attribute of interest. In Table 5.1, the functions for determining the fuzzy measures are shown. In Figure 5.2 is the fuzzy integral graph used to estimate the importance of product attributes.

Table 0.1 Functions for determining the fuzzy measure

Function	Fuzzy measure
v({∅})	0
v({Google trend index})	0.3
v({Sentiment score})	0.4
v({Frequency of attribute})	0.75
v({Google trend index, Sentiment score})	0.2
v({Google trend index, Frequency of attribute})	0.9
v({Frequency of attribute, Sentiment score})	0.6
v({Google trend index, Sentiment score, Frequency of attribute})	1

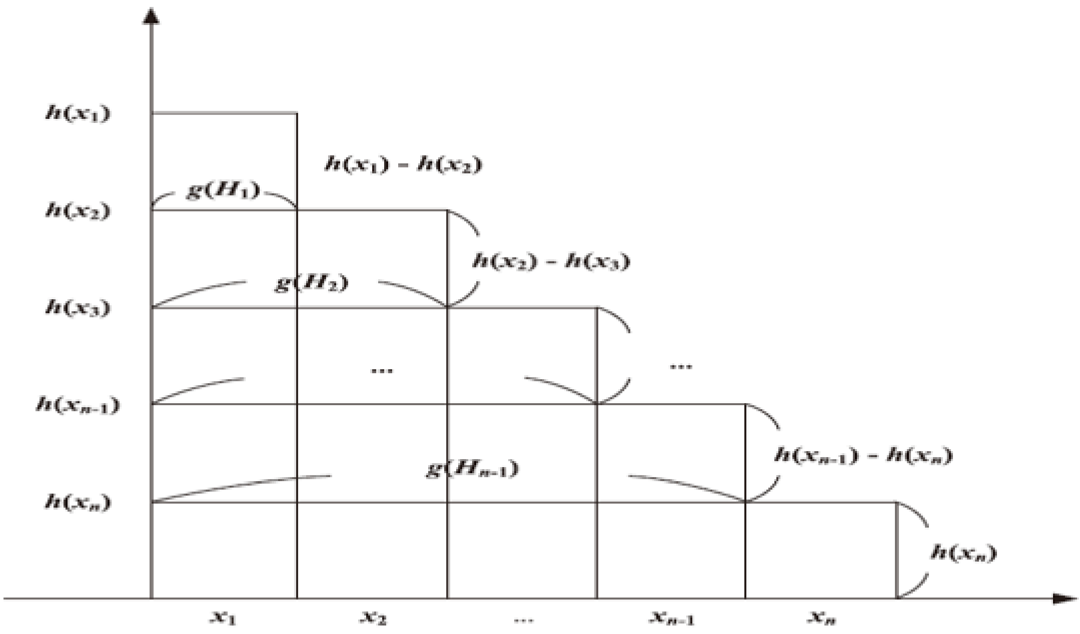

Figure 0.2 A graph of fuzzy integral construction

The fuzzy integral is determined from Equation (5.14) below.

$$\int hdg = h(x_n) * \ g(H_{n)} + [h(x_{n-1} - h(x_n) * g(H_{n-1})] \dots + [h(x_1 - h(x_2) * g(H_1)] \tag{5.14}$$

where $H_1 = \{x_1\}$, $H_2 = \{x_1, x_2\}$, …, $H_n = \{x_1, x_2, \dots, x_n\} = X$

5.2 Determining the future importance of product attributes

5.2.1 Fuzzy time series

Most firms make projections on what consumers may like by observing trends they infer from surveys the have conducted. These methods are usually not sufficiently explicit, are also cumbersome in nature and may not be able to predict consumers' interests and needs in an era of the fast pace of change occurring due to the dynamic nature of the technology industries worldwide. In order to address these problems faced by manufacturers, there is also the issue of dealing with the time lag that occurs between predicting customers concerns and the launching of a new product. Launching products late to the market may lead to the failure of the products since consumers needs and requirements may have changed at the time of the launch of the product.

There are large numbers of current customer reviews available on e-commerce websites for businesses to learn from. This presents an opportunity for manufacturers to be updated on the current trends in consumers' opinions. To transform such opportunities into reality, a model is required to determine how future trends for certain product attributes develop. This section focuses on how fuzzy time series, based on the rough set, can be used to estimate the future importance of product attributes from online reviews. The fuzzy time series approach was proposed by (Song and Chissom, 1994) to forecast the enrolment of students, with a fuzzy linguistic variable as data. In their work, the fuzzy time series was defined as follows:

Let $\mathrm{Y(t)(t} = \cdots, 1,2, \dots)$, a subset of a real number be a universe of discourse on which fuzzy sets $f_i(t)(i = 1,2,\dots)$ are defined and, $F(t)$ is a collection of $f_1(t), f_2(t)$ ... then $F(t)$is defined as a fuzzy time series defied on $\mathrm{Y(t)(t} = \cdots, 1,2, \dots)$. $F(t)$ represents linguistic variables while $f_i(t)(i = 1,2,\dots)$ represents the possible linguistic values of $F(t)$. The linguistic variables are also known as fuzzy sets whose function is time dependent, occurring at different times. The fuzzy time series proposed by (Song and Chissom, 1993) used a first order model. In the first order model, it is supposed that $F(t)$ is the result of $F(t-1)$, i.e., $F(t) \rightarrow F(t-1)$. The relation between the fuzzy variables can hence be described as $F(t)$= $F(t-1)°R(t,t-1)$ where $R(t,t-1)$ is the relationship between $F(t-1)$ and $F(t)$ and $F(t)$= $F(t-1)°R(t,t-1)$ is the first order model of $F(t)$= $F(t)$. $R(t,t-1)$ is independent of time that is $R(t,t-1) = R(t-1,t-2)$ for any time t. For a higher order series or $the\ nth$, fuzzy series, $F(t)$ is denoted by $F(t-n),\dots F(t-2), F(t-1) \rightarrow F(t)$ .The fuzzy relationships is described as $R^n(t,t-1) = F^t(t-2) \times F(t-1) \cup F^t(t-3) \times F(t-2) \cup, \dots, F^t(t-n) \times F^t(t-n+1)$ where n is the parameter affecting the prediction of $F(t)$ and is the number of years before t. For example, when n=3, the third order fuzzy time series model is $F(t) = F(t-1)°R^3(t,t-1)$

and $R^3(t,t-1) = F^T(t-2) \times F(t-1) \cup F^T(t-3) \times F(t-2)$. Prediction at $t = 5$ for a third order fuzzy time series model is represented as $F(5) = F(4)^{\circ}R^3(5,4)$ and the $R^3(5,4) = F^T(3) \times F(4) \cup F^T(2) \times F(3)$.The predicted $F(t)$ is defuzzied to obtain a crisp predicted output.

In conventional fuzzy time series prediction, a first order, second or third order lag period of a fuzzy variable is applied in the fuzzy logical relationship to predict the next data. In this research, the fuzzy logical relationships are replaced by the rough set in order to generate rules that forecasts the importance of product attributes. The fuzzy variable is a universe of discourse U that can be partitioned into several fuzzy intervals (also called linguistic intervals) of equal length $u_1, u_2 \dots u_n$, where n is the number of intervals. For example, fuzzy variable (A), can be partitioned into five fuzzy intervals expressed as shown below:

$$A_1 = \{1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + \cdots + 0/u_n\}$$

$$A_2 = \{0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + \cdots + 0/u_n\}$$

$$A_3 = \{0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + 0/u_5 + 0/u_6 + \cdots + 0/u_n\}$$

$$A_4 = \{0/u_1 + 0/u_2 + 0.5/u_3 + 1/u_4 + 0.5/u_5 + 0\ /u_6 + \cdots + 0/u_n\}$$

$$A_5 = \{0/u_1 + 0/u_2 + 0/u_3 + 0.5/u_4 + 1/u_5 + 0.5\ /u_6 + \cdots + 0/u_n\}$$

where "+" is the union operator, n is the number of intervals, and the length of each interval is defined by $L = [D_{min} - D_1, D_{max} + D_2]/n$. The fuzzy variable obtained from this study is used to generate forecasting rules using the rough set method. Apart from the lagged variables, the Moment, which is the difference between a fuzzy variable and a lagged fuzzy variable, the lagged first order

Moment and the Slope will be used as variables for the rough set-in order to generate forecasting rules. Table 5.2 shows all the variables used in the rough set generate forecasting rules.

Table 0.2 Independent variables for fuzzy rough set time series

Variable	Notation	Description	Definition
I	P	Importance	I
A	Lag1	First lag order of importance at time $t-1$	(I_{t-1})
B	Lag2	Second lag order of importance $t-2$	(I_{t-2}),
C	Lag3	Third lag order of importance $t-3$	(I_{t-3}),
M	M	Moment	$(I_t - I_{t-1})$
		Moment at time t	M_t
		Moment at time $t-1$	M_{t-1}
S	S	Slope	$\frac{(M_t - M_{t-1})}{t-(t-1)}$

5.2.2 Rough set method

Rough set theory was proposed by (Pawlak, 1982) presenting a new approach to imprecision, vagueness and uncertainty. The rough set method was also structured to extract rules associated with complex data and was developed on the assumption that in an information system, every object in the universe is associated with some form of information. Object classifications are characterised by equivalence relations or indiscernibility Any set of indiscernible objects is known as the elementary set, and a union of any elementary sets is known as a crisp set, otherwise referred to as a rough set. Vague concepts cannot be characterized based on the information on their elements. Rough sets deal with the vagueness of concepts by using two kinds of approximations known as the lower and upper approximations. The lower approximation describes the domain objects which are known to belong to a subset with certainty, while the upper approximation describes objects which possibly belong to the subset. Thus, when the lower and upper approximation of the set are not equal, then the set is described as being rough. The rough set is conducted solely with the data presented for feature selection. A four-tuple design table is described as $S = (U, Q, V, \rho)$ where the universe U is a finite non-empty sets of objects; Q is a finite set of attributes; $V = \cup_{q \in Q} V_q$, where V_q is the domain of the attribute q. The information function can be described as $\rho: U \times Q \rightarrow V$, such that $\rho(s, q) \in V_q$ for every $q \in Q$, s $\in U$ and $\exists(q, v)$ where $q \in Q$ and $v \in V_q$ are the descriptions of S. A subset of a set of attributes, $\in Q$, of two objects $x, y \in U$ and are both indiscernible with respect to the attributes R, if and only if, $\rho(x, r) = \rho(y, r)$ for $\exists r \in R$. The indiscernible relation, which is the equivalence relation defined for the set U, is written as $ind(R)$, where $ind\ (R)$ partitions the universe into disjoint subsets and $U/ind(R)$ is used denote these partitions of U. The lower and upper approximation of the set $Y \subseteq U$ can be defined Equations (5.15 and (5.16) as:

$$\underline{R}Y = \cup \{X: X \in U/ind(R), x \subseteq Y\} \tag{5.15}$$

$$\overline{R}Y = \cup \{X: X \in U/ind(R), x \cap Y \neq \emptyset\} \tag{5.16}$$

where $\underline{R}Y$ is made up of all the object in U that certainly belong to Y and $\overline{R}Y$ is made up of all the objects in U that possibly belong to Y under the equivalence relation R. The element existing solely in the upper approximation make up the boundary region(BN). It depicts the area which cannot be certainly classified into Y or its complement. Thus, the boundary region is expressed by Equation (5.17) as:

$$BN(Y) = \overline{R}Y - \underline{R}Y \tag{5.17}$$

The positive region $Pos_R(Y)$ and the negative region $Neg_R(Y)$of Y on R are defined by equations 5.17 and 5.18 respectively.

$$Pos_R(Y) = \overline{R}Y \tag{5.18}$$

$$Neg_R(Y) = U - Pos_R(Y) \tag{5.19}$$

In the process of feature selection, redundant attributes are removed, and the most relevant attributes are retained as the indispensable information for forecasting the future importance of products attributes. Thus, the attribute reductions are also defined:

If R is a set of equivalent relation, $r \in R$, and $Pos_R(Y) \neq Pos_{R-\{r\}}(Y)$, namely $ind(R) \neq ind(R - \{r\})$, R is an independent attribute and r is indispensable attribute in R otherwise r is dispensable.

If R is independent $R \subseteq P$ and $ind(R) - ind(P)$, R is reduction of $P,$ also known as reducts. $R \in RED(P),$ where $RED(P)$ represents the set of all attributes reductions of P.The intersection of $RED\ \ (P)$is the core P, which is expressed as the $Core(P)$. In this study the LEM2 algorithm is adopted to generate a rough set to predict the importance of product attributes. The LEM2 algorithm has the potential to generate a minimal set of rules to describe any size of data including datasets with missing data, hence, it is adopted in this study. Moreover, it can achieve a higher accuracy when a lower and upper approximation of data is used as input. The rules generated from LEM2 algorithm are easy to understand since no complicated models are developed.

5.2.3 Learning from examples algorithm (LEM2)

The LEM2 (Orlowska, 1998) algorithm learns minimal rules from examples. In the LEM2 algorithm, unnecessary conditions are excluded from generating minimal rules. Thus, any conditions that are excluded from a minimal rule will no longer be consistent with the concept description. The algorithm defines the dependency for the attribute-value pairs, for which the concept C depends on the set $T$of attributes -value pairs,

$$T \rightarrow C \ iff \ [T] \subseteq C \text{ and } [T] \neq \emptyset$$

A minimal complex of C is then described as set T, such that C depends on T and T is minimal. If T is a set of such minimal complexes of T, it is said to be a local covering of C if

$$\bigcup_{T \in T} [T] = C$$

and T is minimal.

This kind of local coverage is made up of a single set of rules for a minimal discriminant description of the concept, i.e., a set of rules sufficient enough to differentiate examples belonging to the concept from those examples belonging to the concept complements. The rules generated satisfy the completeness and consistency requirements. The strategy of the LEM2 algorithm involves determining the minimal complexes to cover the examples in the concept until all the examples are covered. For the algorithm to be sufficient, each new iteration procedure seeks a new minimal complex to cover the members of the concept that have not been included by the previous complexes. This continues until the whole concept is covered. The algorithm for the LEM2 method is described below:

```
Procedure:    Find _Single_Local_covering
Input:        Concept C
Output:   Covering T of C
Start
          G=C {goal=concept}
          Ṭ= ∅
          do
               T = Find_Minimal_Complex (C, G)
               Ṭ=Ṭ∪ T
                G = C − ∪_{T∈Ṭ}[T]        {update G}
          Until G = ∅
          {Minimize Ṭ }
          Do ∀ T ∈ Ṭ
               If ( ∪_{S∈Ṭ−{T}} [S] = C    then Ṭ=Ṭ −{T}
          end do
stop
```

First, all the attribute-value pairs relevant to examples in G are collected in a set and denoted as T(G) shown below:

$$T(G) = \{t \mid [t] \cap G \neq \emptyset\}$$

The next best pairs are selected from T (G) and integrated into a minimal complex T until $[T] \subseteq C$ is obtained. Set T is then minimized by attempting to drop redundant attributes -value pairs before returning the result as a minimal complex. The pairs are then removed in order of their selection. The algorithm for finding the minimal complex is described below:

```
Procedure:        Find_Minimal_Complex
Input:            Concept (C) and goal (G), G ⊆ C
Output:           A minimal complex of C
start
                  found = false
                  T = ∅
                  T(G) = { t | [t] ∩ G ≠ ∅
                  do
                  select a pair of t ∈ T(G) such that] |[t] ∩ G is the maximum; break ties
                  by selecting t with smallest [t]; if further ties select first.
                  T1 = { t′ ∈ T(G) | [t′] ∩ G = [t] ∩ G }
                  T = T ∪ T1
                  If [T] ⊆ C then found = true
                          else
                                  G=[t] ∩ G
                                  T(G) = { t | [t] ∩ G ≠ ∅}
                                  T(G) = T(G) − T1
                          end if
until found
do ∀ t ∈ T  {minimize T}
        if [T-{T}]] ⊆ then T =T-{t}
end do
return T                { T is a minimal complex}
```

The forecasting variables given in Table 5.2 are used to generate the rough set If-Then rules. The roughset rules with support greater than 2 are used in forecasting the importance of the product attributes. An example of the rough set rules is shown below:

If Lag1(A1) AND Lag3 (C1) AND Slope (S5) Then Importance (I2)

A1 is the fuzzy set for the first interval and first order importance variable. C1 is the fuzzy set for the first interval and third order importance variable. S5 is the fuzzy set for the fith interval defined for variable Slope. I2 is the fuzzy set for the second interval defined for importance and I2 is the future importance forecast. I2 could then be defuzzified by finding the average of the interval in order to obtain the value of the importance forecast. I2 is defuzzified as shown in Equation (5.20).

$$I2 = \frac{[I_{low} + \quad I_{high}]}{2} \tag{5.20}$$

where I_{low} is the lowest value in the second interval and I_{high} is the highest value in the second interval.

5.3 Implementation

This section describes the case study of an electric hairdryer in determining the importance of product attributes and predicting the future importance of product attributes. A well-known electric hairdryer brand was selected from Amazon.com. The reviews associated with the electric were extracted. A total of 8, 319 reviews from fifteen time periods each month from January 2018 to March

2019 were scraped from the reviews section of the website. The reviews were pre-processed, and opinion mining was conducted on these reviews using the Semantria, an opinion mining software.

Lexicons of terms related to electrical hairdryers were developed to train the Semantria software. Based on the lexicons, Semantria was able to determine and extract customer concerns related to product attributes. The product attributes extracted were categorised under twelve main types. The sentiment scores and the frequencies of these product attributes for fifteen periods were obtained. The search index data were also extracted from the Google Trends. The data obtained from the Google Trends and online reviews were of different scales, hence the data was normalised. Table 5.3 shows the normalised data for the three online metrics.

Table 0.3 Normalized data of the three online metrics for the first two periods

Product attributes	Sentiment scores		Frequencies		Google Trend index	
	Period 1	Period 2	Period 1	Period 2	Period 1	Period 2
Control settings	0.848	1	0.192	0.363	0.00	0.074
Airflow	0.9	0.9	0.948	0.8284	0.328	0
Weight	0.957	0.978	0.641	0.794	0.093	0.139
Usability	0.989	1.00	0.591	0.798	0.148	0.398
Noise	0.596	0.603	0.952	0.508	0	0.052
Price	0.947	0.943	0.545	0.675	0.023	1.00
Easy to use	0.885	0.971	0.326	0.157	0	0.147
Power	0.940	1	0.078	0.092	1	0.695
Comfortable to hold	0.909	0.522	0.855	0.878	0.244	0.236
Durability	1.00	0.766	0.48	0.5	0.173	0.253

Portability	0.98	1.00	0.349	0.616	0.031	0.115
Efficiency	0.75	0.761	0.761	0.891	0.423	1.00

The next step was to determine the weight of each online metric and the interaction of the weights of the online metric. Some experts filled out a questionnaire shown in Table 5.4 ‘1’ denotes highly insignificant while '5’ denotes highly significant. The averages of the rating were determined. In Table 5.5 the average ratings are shown.

Table 0.4 Questionnaire for experts to determine the weight of the online metrics

	Rating
1. How significant is Sentiment scores of product attributes in assessing the importance of the attributes	1 2 3 4 5
2. How significant are Frequencies of product attributes in assessing the importance of the attributes	1 2 3 4 5
3. How significant are Google Trends Indices of product attributes in assessing the importance of the attributes	1 2 3 4 5
4. How significant are Sentiment Scores and Frequencies of product attributes in assessing the importance of the attributes	1 2 3 4 5
5. How significant are Sentiment Scores and Google Trend Indices of product attributes in assessing the importance of the attributes	1 2 3 4 5
6. How significant are Frequencies and Google Trend Indices of product attributes in assessing the importance of the attributes?	1 2 3 4 5
7. How significant are Sentiment Scores, Frequencies and Google Trend Indices of product attributes in assessing the importance of the attributes	1 2 3 4 5

Table 0.5 The average rating for the coalition of online metrics by experts

Coalitions	Average ratings
{Sentiment score }	4
{Frequencies}	3.8
{Google trend index }	4. 2
{Sentiment scores & Frequencies}	4.5
{Sentiment scores & Google trends }	4.8
{Frequencies & Google search}	4.3
{Sentiment & Frequencies & Google trend index}	5

Using the average ratings shown in Table 5.5, the marginal contribution of each online metric sentiment score(S), frequencies (F) and google search index (G) was determined using the Shapley value. Equation (5.21) below indicates how the Shapley value for Sentiment scores (S) can be found. The marginal contribution of the online metrics is shown in Table 5.6.

$$S = \frac{1}{3!}[v(\{S\}) + v(\{S\}) + (v(\{S,F\}) - v(\{F\})) + (v(\{S,F,G\}) - v(\{F,G\})) + (v(\{S,G\}) - v(\{G\})) + (v\{S,F,G\}) - v(\{F,G\})] \quad (5.21)$$

$$S = \frac{1}{3!} \times (4 + 4 + 0.7 + 0.7 + 0.6 + 0.7)$$

$$S = 1.78$$

The relative weight of each online score is found and used as the density in Choquet integral.

Table 0.6 Shows the Shapley value calculation and the relative weight of the individual online metric

Coalitions of online metrics	Sentiment scores	Frequencies of product attribute	Google Search Index
S, F, G	4.0	0.5	0.5
S, G, F	4.0	0.2	0.8
F, S, G	0.7	3.8	0.5
F, G, S	0.7	3.6	0.7
G, S, F	0.6	0.2	4.2
G, F, S	0.7	0.1	4.2
Mean	1.78	1.4	1.86
Relative weight (Density)	0.35	0.28	0.37

The Choquet integral is employed to determine the importance of the twelve product attributes of the electric hairdryer. The variables x_1, x_2 *and* x_3 are used to represent the sentiment scores, frequencies and google search index, respectively. The density of the n_{th} online metric is represented by $g_\lambda(\{x_n\}$. Hence for the three online metrics, the associated density is shown below:

$$g_\lambda(\ \{Sentiment\ scores)\}) = g_\lambda(\{x_1\}) = 0.35$$

$$g_\lambda(\ \{\text{Frequencies of product attribute}\}) = g_\lambda(\{x_2\}) = 0.28$$

$$g_\lambda(\ \{\text{Google Search Index}\}) = g_\lambda(\{x_3\}) = 0.37.$$

The significance of the aggregated online metric is determined by considering the interaction and interdependence of the online metrics using the density of the aggregate online metrics and polynomial equation below: We first determined the

λ value using Equation (5.10). In the case study, the λ value was calculated by solving the following polynomial equation below:

$$\lambda + 1 = (0.35\lambda + 1)(0.28\lambda + 1)(0.37\lambda + 1) \tag{5.22}$$

$$0.03626\lambda^3 + 0.311\lambda^2 = 0$$
$$\lambda = -9.13127 \; and \; 0$$

Since $\lambda \in (-1, \infty)$, then $\lambda = 0$

With the λ value, the density of the occurrence of multiple online metrics can be determined as shown below

$$g_\lambda(\{x_1, x_2\}) = g_\lambda(\{x_1\}) + g_\lambda(\{x_2\}) + \lambda\, g_\lambda(\{x_1\})g_\lambda(\{x_2\}) = 0.63$$
$$g_\lambda(\{x_1, x_3\}) = g_\lambda(\{x_1\}) + g_\lambda(\{x_3\}) + \lambda\, g_\lambda(\{x_1\})g_\lambda(\{x_3\}) = 0.72$$
$$g_\lambda(\{x_2, x_3\}) = g_\lambda(\{x_2\}) + g_\lambda(\{x_3\}) + \lambda\, g_\lambda(\{x_2\})g_\lambda(\{x_3\}) = 0.65$$
$$g_\lambda(\{x_1, x_2, x_3\}) = g_\lambda(X) = g_\lambda(\{x_1\}) + g_\lambda(\{x_2\}) + g_\lambda(\{x_3\}) + \lambda\, g_\lambda(\{x_1\})g_\lambda(\{x_2\})g_\lambda(\{x_3\}) = 1$$

$$If\ \{x_1\} \subseteq \{x_1, x_2\} \text{ then } \mathrm{g}\,\{x_1\} \leq g\,\{x_1, x_2\}$$
$$If\ \{x_1, x_2\} \subseteq \{X\} \text{ then } \mathrm{g}\,\{x_1, x_2\} \leq g\,\{X\}.$$

The importance of the attributes “Control settings”, “Airflow” and “Weight” of the first period are illustrate in Figure 5.3, Figure 5.4, and Figure 5.5 respectively.

$f(x_1)$ is sentiment score, $f(x_2)$ is the frequency and $f(x_3)$ is the Google search index.

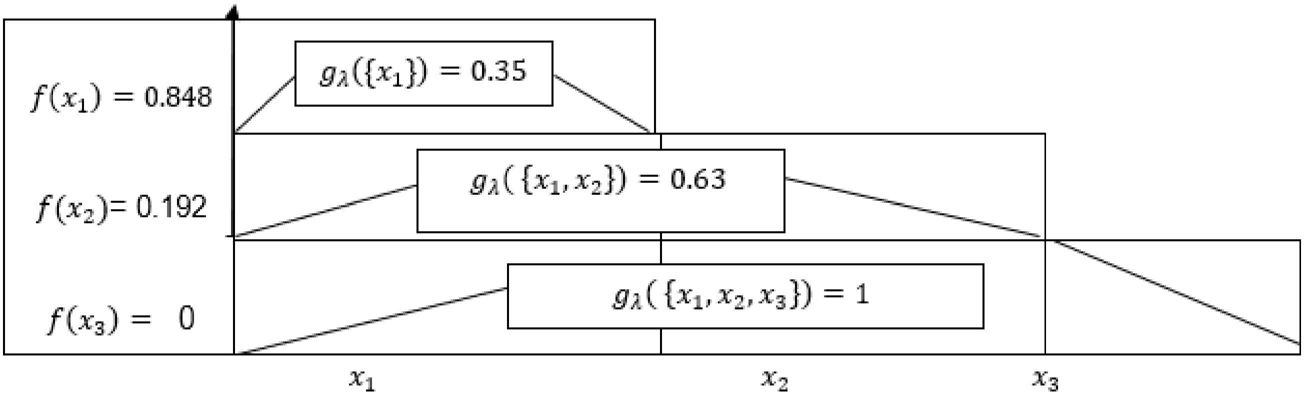

Figure 0.3 Illustration of the Importance of "Control settings" with Choquet integral

$$\int f dg = f(x_3) * g_\lambda(\{x_1, x_2 x_3\}) + \big(f(x_2) - f(x_3)\big) * g_\lambda(\{x_1, x_2\}) + \ (f(x_1) - f(x_2)) * g_\lambda(\{x_1\}) \qquad (5.23)$$

Importance of "Control settings" $= 0 * 1 \ + \ (0.192 - 0) * 0.63 \ + (0.848 - 0.192) * 0.35) = \mathbf{0.35}$

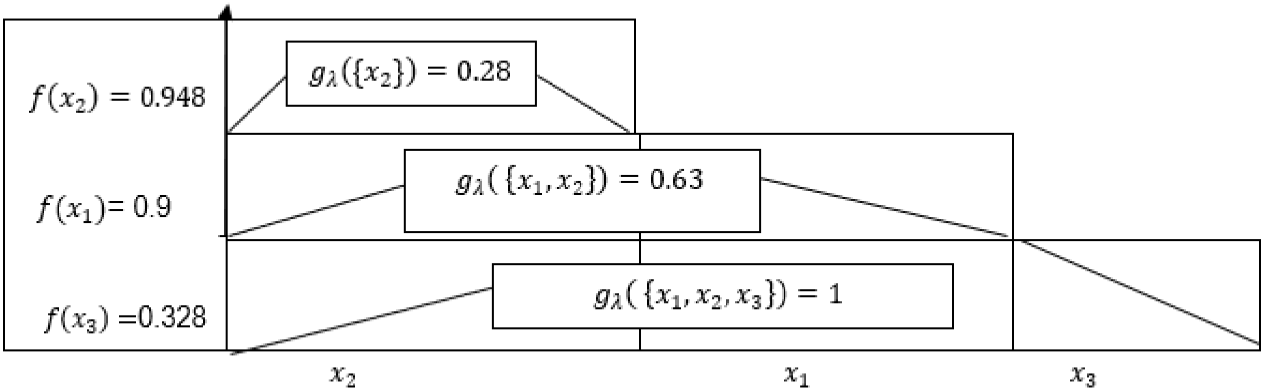

Figure 0.4 Illustration of the Importance of "Airflow" with Choquet integral

$$\int f dg = f(x_3))* g_\lambda(\{x_1, x_2 x_3\}) + \ f(x_1) - f(x_3) * g_\lambda(\{x_1, x_2\}) + \big(f(x_2) - f(x_1)\big) * g_\lambda(\{x_1\}) \qquad (5.24)$$

Importance of "Airflow" $= 0.328 * 1 \ + ((0.9 - 0.328) * 0.63\ ) + ((0.948 - 0.9) * 0.28) = \mathbf{0.702}$

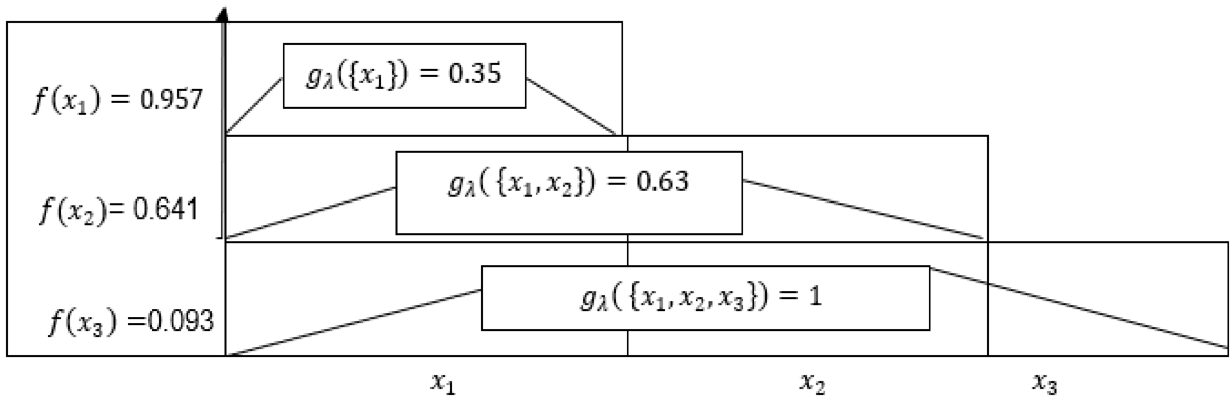

Figure 0.5 Illustration of the Importance of "Weight" with Choquet integral

$$\int f dg = f(x_3) * g_\lambda(\{x_1, x_2 x_3\}) + \big(f(x_2) - f(x_3)\big) * g_\lambda(\{x_1, x_2\}) + ((f(x_1) - f(x_2)) * g_\lambda(\{x_1\}) \quad (5.25)$$

Importance of "Weight" $= 0.093 * 1 + (0.641 - 0.093) * 0.63 + (0.957 - 0.641) * 0.35) = \mathbf{0.549}$

The importance of each product attributes in for each period is computed and scaled by a factor of factor of 10 and the results are shown Table 5.7.

Table 0.7 Importance of product attributes

Product attributes	Period 1	Period 2	Period 3	Period 4	Period 5	Perio d 6	Period 7	Perio d 8	Period 9	Period 10	Period 11	Period 12	Period 13	Period 14	Period 15
Control-settings	3.504	4.788	4.739	7.487	5.525	4.457	3.944	1.678	1.820	1.736	2.275	4.466	2.668	3.862	1.558
Airflow	7.020	5.470	6.670	7.760	5.447	5.147	3.293	2.864	3.785	6.420	4.808	3.900	4.010	4.267	4.020
Weight	5.492	6.165	4.947	7.121	4.094	5.504	4.701	2.639	4.135	5.609	2.234	2.524	4.302	4.554	4.221
Usability	5.667	7.208	6.362	7.514	5.692	4.499	2.451	2.716	4.063	3.116	2.298	3.632	2.866	1.005	2.837
Noise	4.754	4.009	6.479	7.230	5.672	4.245	1.905	1.143	4.237	1.553	1.844	2.569	1.768	3.225	1.721
Price	4.929	8.893	4.959	4.514	3.189	2.347	1.623	1.544	1.086	1.312	1.891	3.020	1.765	1.008	0.082
Easy to use	4.015	4.387	3.213	6.493	7.879	4.204	4.254	2.144	4.833	2.276	2.485	1.746	2.166	1.995	3.181
Power	7.210	6.334	4.219	6.483	7.167	4.640	3.579	3.095	3.804	2.161	3.755	5.290	5.062	5.053	2.954
Comfortable-to-hold	6.479	5.163	3.323	6.021	7.384	6.093	5.099	3.249	5.411	5.508	3.293	2.862	3.983	4.886	2.429
Durability	5.494	5.021	3.965	7.067	6.755	2.478	2.611	2.267	2.150	5.200	5.076	4.743	2.951	2.519	4.591
Portability	4.524	5.650	4.454	5.637	7.616	6.157	4.277	3.661	4.033	6.251	2.619	0.704	1.753	3.034	1.186
Efficiency	6.325	8.860	8.722	7.218	6.987	5.982	6.168	4.051	5.767	2.979	2.691	3.218	4.029	4.782	3.795

Next, the fuzzy rough set time series method is employed to predict the future importance of the electric hairdryer product attributes. Five independent variables are used in the fuzzy rough set time series are namely the first-order lagged importance (Lag1), second-order lagged importance (Lag2), third-order lagged importance (Lag3), moment (M), and slope (S). The dependent variable is the variable is Importance (I) of a product attributes. The universe of discourse, U is defined for each variable Lag1, Lag2, Lag3, M, S, and I using the importance values of each product attribute. U is then divided into several linguistic variables ranging from four to ten as decided in section 5.1.3. For example, in order to determine the interval for Lag1 of product attribute “Control settings”, the, U of the attribute “Control settings” is divided into to five intervals. In Table 5.8, the linguisic variables for all the independent variables for the product attributes “Control settings” are shown.The variables for the fuzzified linguistic intervals for the “Control settings ” for the fifteen periods are shown in Table 5.9. The universe of discourse for all the product attributes are shown from Table A.1 to Table A.12 of Appendix A. The data for determining the rough set variables are shown from Table B.1 to to Table B.12 of Appendix B.

Table 0.8 Fuzzified linguistic intervals of the product attribute "Control settings"

Linguistic interval	Lag1(A)	Lag2 (B)	Lag3 (C)	Moment (M)	Slope (S)	Importance (P)
u_1	[0 2]	[0 2]	[0 2]	[-4 -2]	[-5 -3]	[0 2]
u_2	[2 4]	[2 4]	[2 4]	[-2 0]	[-3 -1]	[2 4]
u_3	[4 6]	[4 6]	[4 6]	[0 2]	[-1 1]	[4 6]
u_4	[6 8]	[6 8]	[6 8]	[2 4]	[1 3]	[6 8]
u_5	[8 10]	[8 10]	[8 10]	[4 6]	[3 5]	[8 10]

Table 0.9 Rough set variables of the product attribute "Control settings"

Period	Lag1	Lag2	Lag3	Moment	Slope	Importance
1	A1	B1	C1	M4	S5	I2
2	A2	B1	C1	M3	S2	I3
3	A3	B2	C1	M2	S2	I3
4	A3	B3	C2	M4	S4	I4
5	A4	B3	C3	M2	S4	I3
6	A3	B4	C3	M2	S3	I3
7	A3	B3	C4	M2	S3	I2
8	A2	B3	C3	M1	S2	I1
9	A1	B2	C3	M3	S4	I1
10	A1	B1	C2	M2	S3	I1
11	A1	B1	C1	M3	S3	I3
12	A3	B1	C1	M4	S5	I3
13	A3	B3	C1	M2	S1	I2
14	A2	B3	C3	M3	S4	I2
15	A2	B2	C3	M1	S1	I1

The fuzzy linguistic variables for each of the product attributes for fifteen periods are determined. Using the Rosetta software, the rough set method was implemented to generate decision "if-then" decision rules from the fuzzy linguistic variables. The "support" of the rough set is used screen decision rules by indicating the number of times a particles decision rules is generated. This allows for a better forecasting accuracy. In this study, the decision rules with support less than 2 were removed from the generated decision rules. The linguistic variables for each product attributes for each time period were used in generating rules from rough sets where, A1, A2, $\ldots, A_n$ represents the fuzzy sets for the first, second to the nth interval of the fuzzy sets of the Lag1 variable. B1, B2, $\ldots, B_n$ are the fuzzy sets for the first, second to the nth interval of the fuzzy sets of the Lag2 variable while C1, C2, $\ldots, C_n$ are the fuzzy sets for the first, second to the nth interval of the fuzzy sets of the Lag3 variable. M1, M2, $\ldots, M_n$ are the fuzzy sets for the first, second to the nth interval of the fuzzy sets of the Moment variable, and S1, S2…,S_n are the fuzzy set for the first, second to the nth interval of the fuzzy sets of the Slope variable. I1, I2…, I_n are the fuzzy sets for the first, second to the nth interval of the fuzzy sets of the importance variable. The rules generated for forecasting all the product attributes are shown in Table C.1 to C.12 of Appendix C. An example of the first rule generated for Control settings forecasting obtained from Table C.1 is shown in Table 5.10 below.

Table 0.10 “Control settings” first rough set rule

Rough set rule
LAG1(A)(A5) AND LAG2(B)(B4) => I(P6)

From Table 5.10, to forecast the importance of “control settings” for period 16, the rule is interpreted as : **If** the first lagged importance at fifth fuzzy interval

(LAG1(A)(A5)) and the second lagged importance at the fourth fuzzy interval (LAG2(B)(B4)) is establish for control settings at period 15, **Then** the importance is the sixth fuzzy interval (P6). The fuzzy interval I(P6) is then defuzzified to obtain the forecast of importance of "control settings" for period 16.

With the decison rules generated, the future importance of the product attributes at the 16th is determined using the three online metrics obatind from the previous preriods. Table 5.10 shows the predicted future importance of the product attributes in the 16th period. The product attribute

"Weight" had the most important value. The product attribute "Price" had the least importance value indicatingt the least concern to customers among all the eletcric hairdryer attributes in the future. Similarly, "Control settings" also had a low importance value in predicted period. Thus fewer resources should be invested in improving the "Price" and "Control settings".

Table 0.11 The future importance of product attributes of an "electric hairdryer"

Product attribute	Importance at Period 16
Control Settings	1
Airflow	4.5
weight	5
Usability	2.5
Noise	1.5
Price	1
Easy to use	1.5
Power	4.5
Comfortable to hold	3.5
Durability	2.5
Portability	3.5
Efficiency	2.5

5.4 Validation

The effectiveness of the proposed fuzzy rough set time method is validated by using the proposed method to forecast the importance of product attributes in the periods 13, 14 and 15. The results of the forecasts are compared with the actual importance values of the products attributes, the fuzzy k medoid clustering times series and the adaptive neuro-fuzzy inference system (ANFIS) for periods 13, 14 and 15. The results of the forecast are summarised in Table 5.11, Table 5.12 and Table 5.13 for periods 13, 14 and 15 respectively. The root-relative square error (RRSE), relative absolute error (RAE) and the root mean square error (RMSE) were used as performance metrics to measure the accuracy of the of the forecast. The results of the performance metrics are shown in Table 5.14, Table 5.15, and Table 5.16 respectively. Based on the performance indices, the fuzzy rough set time series methods had better forecasting accuracy compared with the other three methods.

Table 0.12 Comparison of the predicted importance values of the product attributes and actual values for Period 13

Period 13					
Actual value		Predicted value			
Product attribute		Proposed method	Fuzzy time-series	Fuzzy k medoid clustering time series	ANFIS
Control Settings	2.668	3.000	5.000	4.961	5.590
Airflow	4.010	4.500	3.500	6.604	1.030
weight	4.302	5.000	3.500	1.991	1.430
Usability	2.866	2.500	2.500	2.158	3.900
Noise	1.768	1.500	1.500	1.513	2.140
Price	1.765	1.000	2.000	2.137	6.750
Easy to use	2.166	1.500	1.500	6.720	8.130
Power	5.062	4.500	3.500	7.239	0.521
Comfortable to hold	3.983	3.500	2.500	4.662	3.300
Durability	2.519	2.500	4.160	4.008	5.330
Portability	1.753	1.500	1.500	6.029	0.003
Efficiency	4.029	4.500	5.500	5.895	3.950

Table 0.13 Comparison of the predicted importance values of the product attributes and actual values for period 14

Period 14					
Actual value		Predicted value			
Product attribute		Proposed Method (Fuzzy rough set time series)	Fuzzy time series	Fuzzy k medoid clustering time series	ANFIS
Control Settings	2.668	1.000	3.000	5.710	0.990
Airflow	4.010	4.500	3.000	6.604	4.130
weight	4.302	5.500	3.000	1.991	5.950
Usability	2.866	1.500	2.000	3.399	3.420
Noise	1.768	3.500	2.830	1.513	2.540
Price	1.765	1.000	2.000	2.958	1.560
Easy to use	2.166	2.500	2.500	7.493	6.120
Power	5.062	4.500	3.500	3.901	3.600
Comfortable to hold	3.983	4.500	3.500	3.093	3.160
Durability	2.519	2.500	2.300	4.008	3.470
Portability	1.753	3.500	3.000	6.627	0.082
Efficiency	4.029	4.500	5.500	7.314	5.390

Table 0.14 Comparison of the predicted importance values of the product attributes and actual values for period 15

Product attribute	Actual value	Predicted value			
		Proposed method (Fuzzy rough set time series)	Fuzzy time series	Fuzzy k medoid clustering time series	ANFIS
Control Settings	2.668	1.000	3.000	5.710	9.900
Airflow	4.010	4.500	3.000	6.604	2.360
weight	4.302	5.000	3.000	1.991	6.850
Usability	2.866	2.500	1.500	2.158	2.050
Noise	1.768	1.500	1.500	1.513	1.090
Price	1.765	1.000	2.000	2.137	1.340
Easy to use	2.166	3.500	1.500	4.401	9.690
Power	5.062	2.500	4.500	5.925	2.730
Comfortable to hold	3.983	2.500	4.500	5.103	3.360
Durability	2.519	4.500	2.500	4.008	2.730
Portability	1.753	1.500	3.000	3.424	0.311
Efficiency	4.029	4.500	5.500	5.895	5.250

Table 0.15 Comparison of the root relative squared errors of the forecasted importance value of product attributes

Product attributes	Proposed method (Fuzzy rough set time series)	Fuzzy time series	Fuzzy k medoid clustering time series	ANFIS
Control Settings	1.105	1.082	1.917	3.500
Airflow	17.091	40.217	102.458	80.543
weight	23.381	35.360	68.108	68.575
Usability	0.309	0.751	1.138	1.206
Noise	0.303	0.360	1.192	0.686
Price	0.841	1.529	2.010	3.639
Easy to use	1.088	2.284	8.810	11.854
Power	0.308	0.912	1.308	1.618
Comfortable to hold	0.201	0.939	1.065	0.672
Durability	0.995	0.963	0.757	1.325
Portability	0.344	1.022	3.356	1.976
Efficiency	1.680	4.443	7.108	2.968
Average RRSE	**3.9705**	7.4885	16.602	14.8801

Table 0.16 Comparison of the relative absolute error of the forecasted importance value of product attributes

Product attributes	Proposed method (Fuzzy rough set time series)	Fuzzy time series	Fuzzy k medoid clustering time series	ANFIS
Control Settings	1.609	2.210	3.556	6.062
Airflow	3.580	9.392	22.362	14.217
weight	6.213	12.071	18.212	17.685
Usability	0.487	1.277	1.535	1.721
Noise	0.387	0.240	1.101	0.855
Price	0.972	2.816	2.516	3.907
Easy to use	1.015	2.553	7.682	11.312
Power	0.559	2.269	2.246	2.217
Comfortable to hold	0.352	2.877	1.924	1.249
Durability	1.481	2.467	1.291	2.098
Portability	0.495	1.699	4.845	2.673
Efficiency	1.257	4.219	5.601	1.847
Average	**1.5339**	3.6741	6.072	5.4869

Table 0.17 Comparison of the root mean square errors of the proposed method and the fuzzy time series method

Product attributes	Proposed method (Fuzzy rough set)	Fuzzy time series	Fuzzy k medoid clustering time series	ANFIS
Control Settings	1.694	1.659	2.939	5.366
Airflow	0.418	0.984	2.507	1.971
weight	0.814	1.231	2.372	2.388
Usability	0.405	0.985	1.493	1.583
Noise	0.256	0.304	1.006	0.579
Price	0.690	1.254	1.649	2.985
Easy to use	0.517	1.084	4.181	5.626
Power	0.525	1.554	2.228	2.756
Comfortable to hold	0.359	1.674	1.900	1.199
Durability	1.681	1.627	1.280	1.828
Portability	0.356	1.058	3.474	2.045
Efficiency	0.516	1.365	2.183	0.912
Average RMSE	**0.686**	1.232	2.268	2.436

5.5 Summary

In this chapter, data from the online reviews and Google Trends were used to determine the importance of product attributes using the Shapley value and the Choquet integral. The online reviews provide customer concerns and needs in a manner that conventional surveys are limited. Using opinion mining, the customer needs and the associated sentiment scores can be obtained. The Google Trends on the other hand provides metrics that shows consumer interest in a product and its attributes over a time period. A case study on an electric hairdryer was conducted to determine the importance of product attributes of an electric hairdryer. The Choquet integral was used to determine the importance of product attributes at different periods. To address the lack of studies addressing the dynamic needs of consumers, the fuzzy rough set time series method was used to predict the future importance of the product attributes. The effectiveness of the proposed method was determined by comparing its performance with the fuzzy time series, fuzzy k medoids clustering time series and the ANFIS. The RRSE, RAE and RMSE were determined for each forecasted importance value of a product attribute using the fuzzy rough set time series, the fuzzy time series, fuzzy k medoids clustering time series and the ANFIS method. The performance metrics showed that the fuzzy rough set time series had a better forecasting accuracy compared to the fuzzy time series, fuzzy k medoids clustering time series and the ANFIS.

Chapter 6 A fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews

Product lifecycles become shorter and shorter due to rapid changes in technologies. New products are regularly introduced to the market by product manufacturers. Thus, the demand for existing products and new products may not be forecasted accurately. Hence, there is a need for product adoption forecasting methods to address the uncertainties about the demand of the products. In this research, user-generated content in online reviews is used to develop fuzzy market-share models. The conventional Bass model is redefined to include word-of-mouth from the internet.

This chapter addresses the limitations of past studies that rely on large amounts of data on past product sales or adoptions in order to make accurate product adoption forecasts. Moreover, in the current dispensation where big data is integrated in businesses, there is a need to integrate data such as online reviews in demand forecast models. This enables demand forecasts to reflect the needs of consumers. The fuzziness in product demand forecasting is also not really considered in extant studies and in this chapter, the product adoption forecasting is categorized into three scenarios to address the fuzziness in demand forecasting. The methodology for forecasting the adoptions of a product using online reviews is presented in section 6.1. The implementation of the proposed method is described in section 6.2. Validation of the proposed method is presented in section 6.3 and finally, a summary of chapter 6 is presented in section 6.4. The framework for developing a market share model under uncertainty and a modified Bass model is summarised in Figure 6.1.

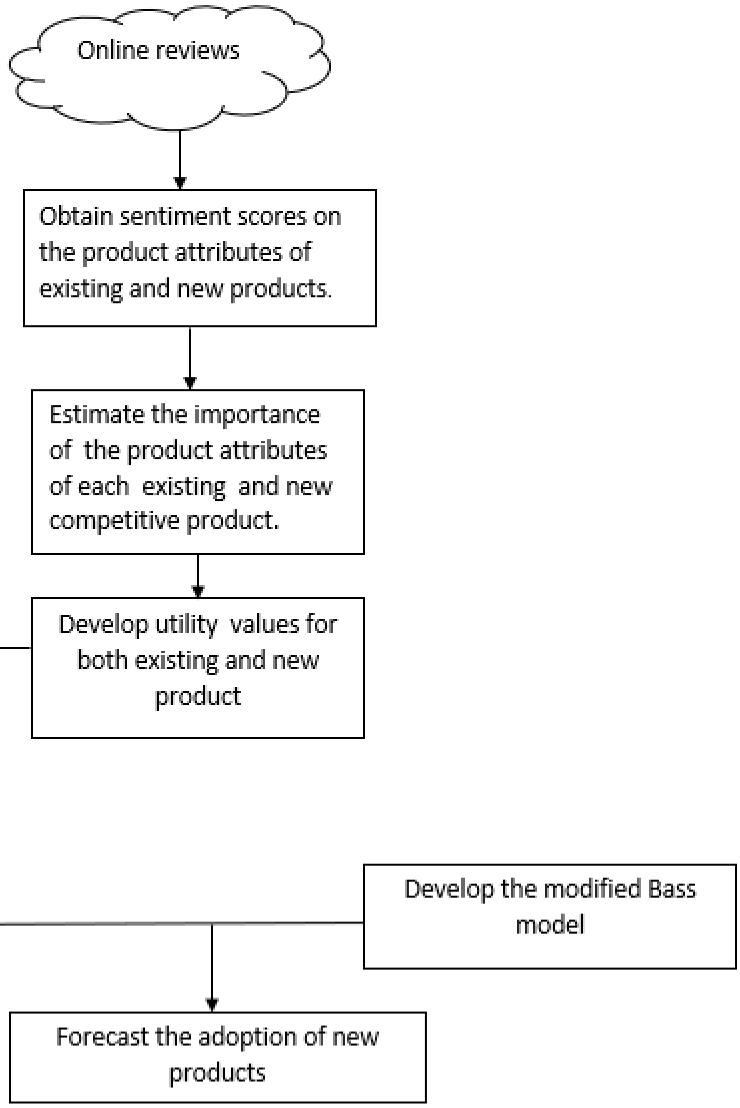

Figure 0.1The framework for developing a new market share model and a new modified Bass model.

6.1 Market demand modelling using online reviews

First online reviews of competitive products in the same category were extracted from e-commerce websites using the opinion mining technique. The most common product attributes were extracted, and the associated sentiment scores and frequencies were obtained. The sentiments scores were obtained for past products and new products released in the USA market. Using the FIS (fuzzy inference system), the importance of each product attribute was estimated. Based on the sentiment scores, and products ratings, the fuzzy utilities were obtained for each product using a utility function. Next, an optimization model was developed to obtain parameters for market share model. A modified Bass model integrated with sentiment scores from online reviews was developed. With the modified Bass, the market model, the demand/adoptions for the new products could be estimated.

6.2. Fuzzy discrete choice modelling

6.2.1 Fuzzy inference system

The fuzzy inference system (FIS) provides a systematic framework for reasoning and decision making when addressing uncertain and imprecise information. For FIS, the input-output relationships are defined by experts using a set of IF (antecedent) - Then (Consequent) rules. In this study, sentiment scores and frequencies are fuzzified and converted to fuzzy sets. Membership functions (MFs) are defined to correspond to linguistic variables that are assigned to the fuzzy sets. Different MFs such as the triangular MF, trapezoidal MF, gaussian MF etc. are available to be used in the fuzzy inference system. This study employs the triangular MF due to its simplicity. The relationship between sentiment scores, frequencies and importance of product attributes is determined by establishing a set of fuzzy inference rules in an "if-then" format. The fuzzy rules are described by Equation (6.1) below:

$$R_i = If \ (X_{i1} is \ x_{i1}, and \ \ X_{i2} \ is \ x_{i2}), then \ \ Y_i \ is \ y_i \tag{6.1}$$

where $x_{i1} and \ x_{i2}$ are fuzzy sets that correspond to the input linguistic variables "sentiment scores" (X_{i1}) and "frequencies" (X_{i2}), respectively; y_i is fuzzy set of the output linguistic variable "importance" Y_i.

A membership degree is estimated for each antecedent and consequent corresponding to the fuzzy inference rules. For an antecedent with more than one-part, fuzzy operators are applied to generate a single value. The min (minimum) and prod (product) operators are used in the "and "connective rule whiles max (maximum) and probor (probabilistic OR) are used in the "or" connective rule. The implication operator is next applied to the consequent using the *min* or *prod.* Lastly, the output is aggregated using the max, probor and sum method and defuzzified to obtain the crisp output value of importance. Various defuzzification methods exist but the centroid method of defuzzification is employed in this study. The importance value obtained from the defuzzification process is used to represent the importance weights of the product attributes.

6.2.2 Fuzzy utility function

To determine the market share model of a product, fuzzy utilities are determined in this study for each product of interest. The utility defined for each product is expressed as shown in Equation (6.2) below:

$$U_j = \left(\sum_{k=}^{m} R_{kj}\right) * \left(\sum_{k=}^{m} SS_{kj}\right) \tag{6.2}$$

where R_{kj} is the rating of the jth product profile, SS_{kj} is the sentiment score of the jth product profile with $k = 1,2,3 \dots, m$ representing the number of product attributes. The rating R of the jth product profile is determined by adopting the fuzzy linear regression. The explanatory variables of FR are the importance of product attributes, while the dependent variables are the ratings of the products of interest obtained from the online reviews. The FR is expressed by Equation (6.3).

$$\tilde{Y} = \tilde{A}_0 + \tilde{A}_1 x_1 + \cdots + \tilde{A}_j x_j + \cdots + \tilde{A}_n x_n \qquad \text{j=1,2…, n} \qquad (6.3)$$

where $\tilde{Y}$ is the estimated fuzzy output, $\tilde{A}_j$ is the fuzzy coefficient of the jth independent variable and x_j is the non-fuzzy vector of the jth independent variable. The fuzzy coefficient $\tilde{A}$ is made up of a centre and spread value $\tilde{A} = (\alpha, c) = (\alpha - c, \alpha, \alpha + c) where\ \alpha$ is the centre value and $\alpha - c$ is the left value of the fuzzy coefficient and $\alpha + c$ is the right value of the fuzzy coefficient. Hence the Equation (6.4) can be rewritten as:

$$\tilde{Y} = (\alpha_{0,} c_0) + (\alpha_{1,} c_1) x_1 + \cdots + (\alpha_{j,} c_j) x_j + \cdots + (\alpha_{n,} c_n)_n x_n \qquad (6.4)$$

To determine the fuzzy coefficients Tanaka’s FR is employed by solving the linear programming problem below:

$$\mathrm{Min} J \ \textstyle\sum_{j=0}^{N} (c_j \ \sum_{j=0}^{M} | x_{ij} |) \qquad (6.5)$$

$$\sum_{j=0}^{N} (\alpha_{j,} x_{ij} + (1-h) \sum_{j=0}^{N} c_j | x_{ij} | \ \geq y_i \qquad (6.6)$$

$$\sum_{j=0}^{N} (\alpha_{j,} x_{ij} - (1-h) \sum_{j=0}^{N} c_j | x_{ij} | \ \leq y_i \qquad (6.7)$$

$$c_j \geq 0, \qquad (6.8)$$

$\alpha \in R$, x_{i0}=1, $i = 0,1,2, \dots, M$ $j = 0,1,2, \dots, N$ for all i and $0 \leq h \leq 1$

$\alpha_j and\ c_j$ represent the centre and spread values of the fuzzy coefficient of the jth independent variable respectively; x_{ij} is the variable for the jth independent variable of the ith data set. M indicates the number of data points and N is the number of independent variables. Equations (6.6), (6.7) and (6.8) are the constraints of the estimated data. The h-factor, which takes values between 0 and 1, measures the degree of fitness of the fuzzy linear model.

6.2.3 Developing market share model

The MNL is adopted in this study to develop the market model of products. The market share model developed are expressed as probabilities using the MNL model. The MNL model is defined as the probability of choosing a pth company existing product over a competitive product and is described as follows:

$$\Pr(p) = \frac{e^{u_P}}{\sum_{j=1}^{J} e^{u_j} + \sum_{k=1}^{K} e^{u_k} + e^{u_P}} \tag{6.9}$$

where u_P is the utility of the pth product, u_k is the utility of the kth company's existing product, $and\ u_j$ is the utility of the jth competitive product. To determine the market share of a product of interest, Equation (6.9) is fuzzified by introducing the fuzzy utilities to obtain the market share equation below:

$$\widetilde{\mathrm{MS}}(p) = \frac{e^{\alpha_p \tilde{u}_P}}{\sum_{j=1}^{J} e^{\alpha_j \tilde{u}_j} + \sum_{k=1}^{K} e^{\alpha_k \tilde{u}_k} + e^{\alpha_p \tilde{u}_P}} \tag{6.10}$$

where $\widetilde{MS}(p)$ is the market share (under uncertainty that also addresses fuzziness) of the pth product among the product alternatives. The market share of all products are expressed in three different scenarios, normal case scenario, worst case scenario and best-case scenario to corresponding to the centre, left and right value of a fuzzy set, as shown below:

$$\text{MS }(p)_{center\ (normal\ scenario)} = \frac{e^{\alpha_p {u^n}_p}}{\sum_{j=1}^{J} e^{\alpha_j {u^n}_j} + e^{\alpha_p {u^n}_p}} \tag{6.11}$$

$$\text{MS }(p)_{left\ (worse\ \ scenario)} = \frac{e^{\alpha_p {u^w}_p}}{\sum_{j=1}^{J} e^{\alpha_j {u^w}_j} + e^{\alpha_p {u^w}_p}} \tag{6.12}$$

$$\text{MS }(p)_{right(best\ \ scenario)} = \frac{e^{\alpha_p {u^b}_p}}{\sum_{j=1}^{J} e^{\alpha_j {u^b}_j} + e^{\alpha_p {u^n}_p}} \tag{6.13}$$

$$Min \left(\frac{e^{\alpha_p \tilde{u}_P}}{\sum_{j=1}^{J} e^{\alpha_j \tilde{u}_j} + \sum_{k=1}^{K} e^{\alpha_k \tilde{u}_k} + e^{\alpha_p \tilde{u}_P}} - \widetilde{\text{MS}}\ (p) \right) \tag{6.14}$$

where $\text{MS }(p)_{center\ (normal\ scenario)}$, $MS\ (p)_{left\ (worse\ \ scenario)}$ $and\ \ MS\ (p)_{Right(Best\ \ scenario)}$ refers to the normal scenario of market share, worse scenario of the market share and the best scenario of the market share of the pth product respectively. ${u^n}_p,\ {u^w}_p,\ and\ \ {u^b}_p$ refers to the utility of the pth product for the normal, worse, and best scenario, respectively. ${u^n}_j,\ {u^w}_j, {u^b}_j$ refers to the utility of the jth competitive product the normal, worse, and best scenario, respectively. α_p, is the alpha value of product P $and\ \alpha_j$ is the alpha value of the jth product brand.

6.2.4 Integrating online review in Bass model.

The Bass model was established under the assumption that new product adoption is triggered by innovation and imitation. Innovation in the Bass model is represented by the coefficient p while imitation is represented by the coefficient variable q. In this study, the coefficient of imitation is explained by sentiments from online reviews of consumers. For a large number of positive reviews, a higher value of sentiment scores can be obtained. Such large number of positive has reviews have the potential to drive the adoption of a new product since consumers tend to be influenced by positive reviews on a product. However, further reviews after certain volume of positive reviews have been reached does not further increase the adoption of a product hence q is assumed to form a logistics - S shaped curve. The parameter, q can be defined as function of sentiment scores (SS) at time t and is defined as $q = f(SSt)$, and represented as shown below:

$$q=\frac{q^m q^0}{q^0+(q^m-q^0)e^{-\gamma SSt}} \tag{6.15}$$

where $q, q^m, q^0 and\ \gamma$ are the effect of WOM (coefficient of innovation), maximum effect of WOM, minimum effect of WOM and a constant respectively.

$$n(t) = \frac{dN(t)}{dt} = [p + \frac{\frac{q^m q^0}{q^0+(q^m-q^0)e^{-\gamma SSt}}}{m} N(t)][m - N(t)] \quad (6.16)$$

where m is the market demand and $N(t)$, is the number of customers who have already bought the product.

6.3 Implementation

This chapter explains the implementation of the proposed methodology for forecasting the adoption of tablet personal computers (P.C). Online reviews and the market share of five competitive brands of tablet were initially obtained. The brands were Apple, Amazon, Samsung, Asus, and Acer. A total of 13,655 online reviews from January 2019 to December 2019 were crawled from ecommerce websites. Product ratings associated with the reviews of the tablets P.C were also obtained. For this case study, tablet P.C sold from January 2018 and December 2019 were classified as old tablet P.Cs. From January 2020 to June 2020, a total of 3,213 online reviews for these five brands of tablets P.Cs were also obtained. From January 2020 to June 2020, tablet P.Cs sold were classified as new tablet P.Cs. Opinion mining was conducted on the reviews and the main customer concerns mentioned were extracted. The customer concerns mentioned were the product attributes on which consumers sentiments were expressed. The results of the opinion mining are summarised in Table 6.1, Table 6.2, and Table 6.3.

Table 0.1 Product attributes extracted

Product attributes	Symbols
Tablet memory	TM
Size	SZ
Screen	SC
Processing power	PP
Design	DS
Camera	CM
Battery life	BL
Price	PR

Table 0.2 Sentiment Scores of old Tablet P.C

	TM	SZ	SC	PP	DS	CM	BL	PR
Apple	0.6	0.38	0.35	2.874	0.65	0.77	-1.73	-0.2
Amazon	0.437	0.81	0.63	0.46	0.79	0.66	0.64	-0.4
Samsung	0.248	0.77	1.52	2.56	0.88	0.56	0.54	-0.03
Asus	0.352	0.82	0.79	0.91	0.73	0.88	0.86	0.15
Acer	0.153	0.278	0.185	0.36	1.65	2.5	-0.27	-0.5

Table 0.3 Frequencies of old Tablets P.C (%)

	TM	SZ	SC	PP	DS	CM	BL	PP
Apple	23.8	19.5	29.5	17.6	20.1	33.6	28.6	31.4
Amazon	15.6	12.4	22.6	19.7	20.4	14.8	18.9	10.3
Samsung	32	29.3	13.4	20.2	25.3	15.9	27.6	25.5
Asus	13.2	21.4	19.3	22.8	22.4	23.4	15.3	20.5
Acer	15.3	17.4	15.1	19.7	11.7	12.2	9.5	12.3

The sentiment scores and frequencies shown in Table 6.2 and Table 6.3 respectively were converted into triangular fuzzy numbers (TFNs) with a neighbour functions intersecting a membership value of 0.5. The membership value of each variable was 1. Figure 6.2, Figure 6.2, and Figure 6.4 show the membership function of the linguistic variables used in determining the importance of product attributes. Table 6.4 shows the results of the importance of the product attributes from the fuzzy inference system.

Table 0.4 Importance of product attributes

	TM	SZ	SC	PP	DS	CM	BL	PP
Apple	2.69	2.6	3.49	2	2.46	3.4	4.08	3.69
Amazon	2.53	2.25	2.58	2.56	2.38	2.39	2.46	1.78
Samsung	3.68	3.42	1.86	2.26	2.95	2.48	3.34	2.99
Asus	2.37	2.49	2.38	2.63	2.59	2.69	2.3	2.26
Acer	2.64	2.66	2.61	2.61	1.71	1.77	1.32	2.28

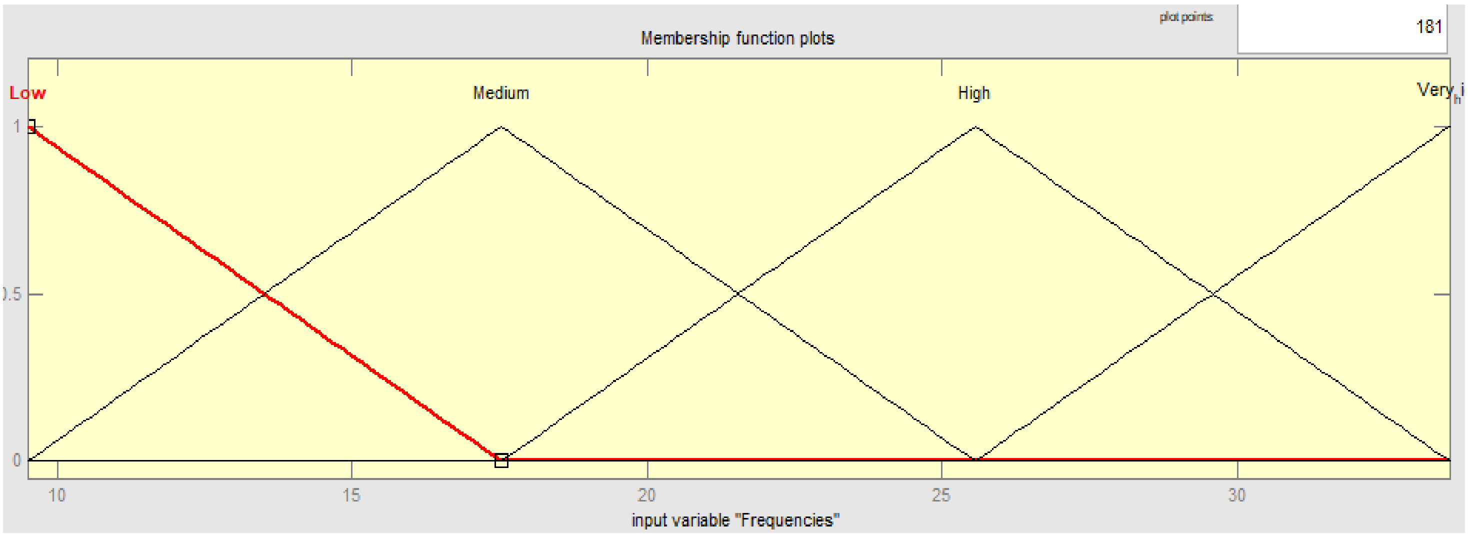

Figure 0.2 Membership function of the linguistic variable for the input "frequencies"

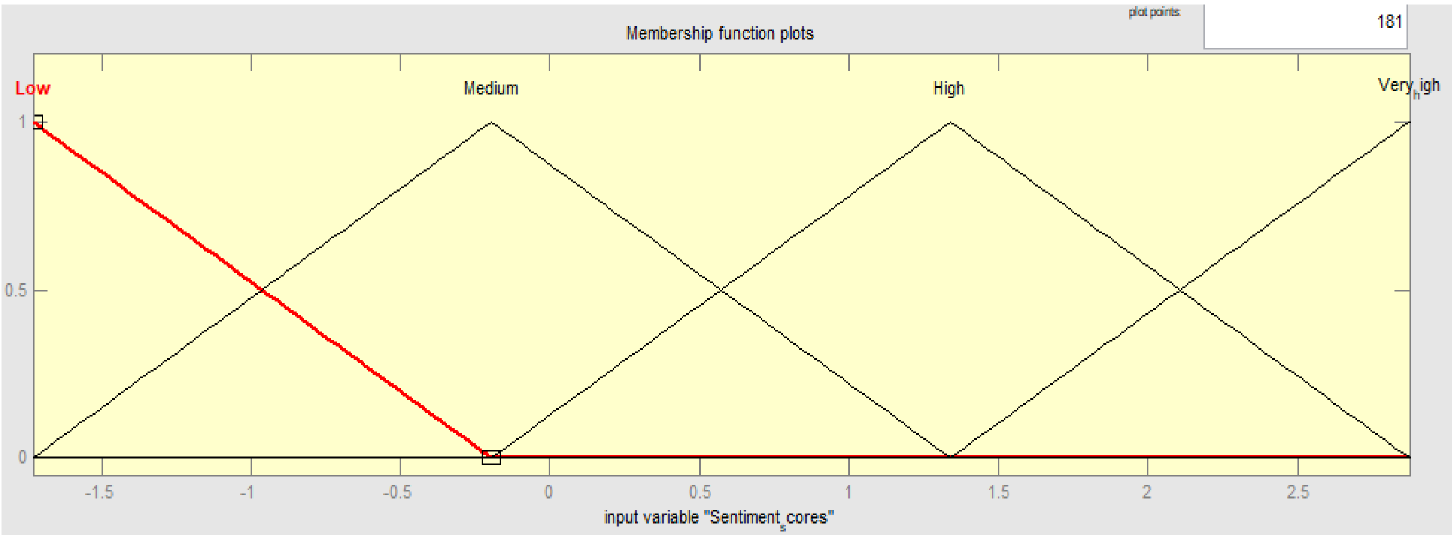

Figure 0.3 Membership function of the linguistic variable for the input "Sentiment score"

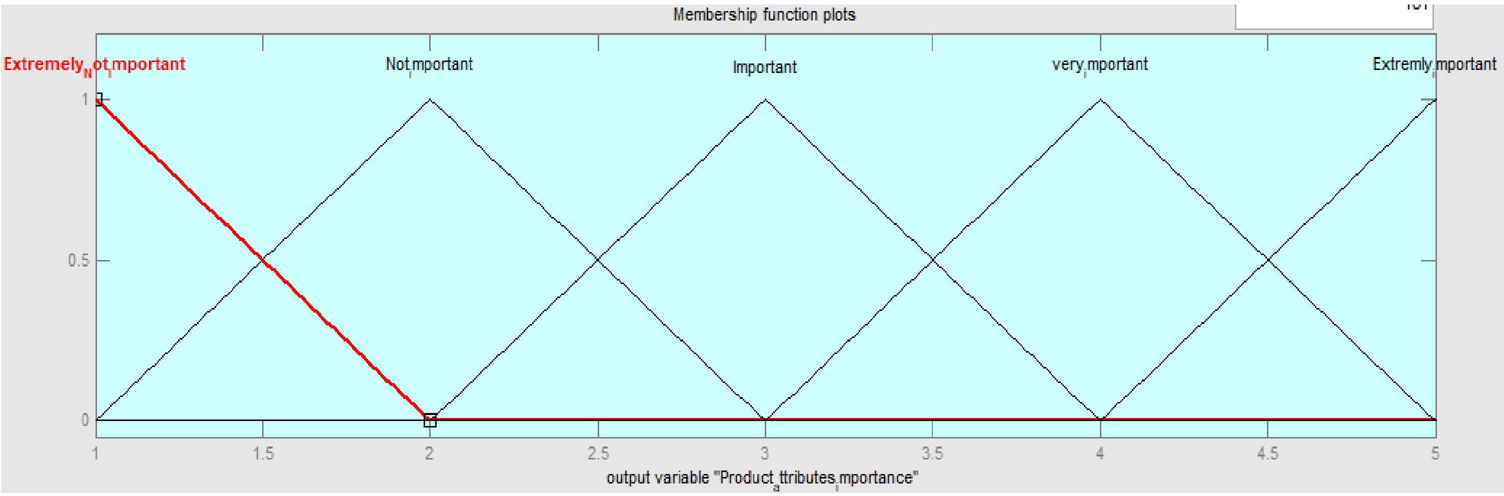

Figure 0.4 Membership function of the linguistic variable for the output "Importance

To determine the utilities of the old tablet P.Cs, the FR model was developed to relate the importance of the product attributes to the overall ratings of the product. Table 6.5, the coefficient of importance of the product attributes of the FR model is shown. The fuzzy utilities were determined using Equation (6.2) and the utilities obtained for each brand of tablet P.C are shown in Table 6.6. U_a, U_b, U_c, U_d, U_e indicates the utilities of the five brands of the tablet P.C, Apple, Amazon, Samsung, Asus, and Acer respectively.

Table 0.5 Example of fuzzy importance weights of the FR model for Apple Tablet PC

Product attribute	Centre	Spread
	0.1766	2.0238
Tablet Memory	0.6070	0.1195
Size	0.2241	0.3455
Screen	0.6159	0.4362
Processor	0.4463	0.4372

Design	-0.0223	0.9464
Camera	-0.1594	0.9213
Battery life	0.0989	0.0242
Price	-0.2008	0.4207

Table 0.6 Fuzzy utility of the old Tablet P.C

Product (Centre)	Utility value
Ua	47.8083
Ub	20.9168
Uc	28.6961
Ud	19.4425
Ue	14.4908
Product (Left)	Utility value
Ua	32.0393
Ub	20.1569
Uc	27.1950
Ud	18.6254
Ue	14.3432
Product (Right)	Utility value
Ua	63.5772
Ub	38.3258
Uc	53.5798
Ud	34.3627
Ue	26.5467

Using Equations (6.10), (6.11) and (6.12), the alpha values were determined by solving Equation (6.14). The alpha values generated were used to determine the market share models of the Tablet P.Cs Table 6.7 shows the alpha values determined.

Table 0.7 Alpha coefficient for market share model

Product	Centre alpha	Left alpha	Right alpha
alpha 1(Apple)	0.15	0.32	0.37
alpha 2 (Amazon)	0.25781	0.42	0.48
alpha 3 (Samsung)	0.25781	0.31	0.32
alpha 4 (Asus)	0.0952	0.01	0.63
alpha 5 (Acer)	0.168	0.525	0.556

To determine the market share of the new tablet P.Cs released from 2020, the utilities of the tablet P.Cs were first obtained by determining the importance of product attributes by using sentiment scores and frequencies of the new tablet P.Cs Tables 6.8, 6.9 and 6.10 show the values of the sentiment scores, frequencies, and importance of the attributes of the new tablet P.C, respectively.

Tanaka's FR was adopted to develop fuzzy utilities for the market share models. Besides, the market shares developed from the fuzzy utilities allows for categorisation under the normal case scenario, worst case scenario and the best case scenario. Thus, the market share models under uncertainties can be addressed by developing these fuzzy utilities from Tanaka's FR.

Thus, for the next step after obtaining the importance of product attributes for the tablet P. Cs, a model was developed to relate the importance of the tablet P.C attributes to the overall tablet P.C ratings using Tanaka's FR. The FR model shown in Equation 6.1 was used to estimate the utilities of the new tablet P.Cs. The market shares for the new tablet P. Cs were determined using the alpha values and utilities of the new competitive tablet P.C. The estimated fuzzy utilities and the market shares of new tablet P. Cs are shown in Tables 6.11 and 6.12.

Since the market share models were developed with the availability of past sales data, the number of adoptions of any of the tablet P.Cs could be determined. In this study, the Apple tablet P.C was chosen to demonstrate how the adoption of the Apple tablet P.C could be forecasted using the modified Bass model. The new market shares of Apple tablet P.C in conjunction with the market potential of Apple tablet in USA, was used to forecast the quarterly adoptions of Apple tablet users in USA starting from the first quarter of 2020. Table 6.13 shows the fuzzy market demand of the Apple tablet P.C. determined from the market shares under the three scenarios and market potential of the apple tablet P.C in the USA market.

Table 0.8 Sentiment scores of product attributes of new tablet P.C

Product	Sentiment Scores							
	TM	SZ	SC	PP	DS	CM	BL	PR
Apple	0.736	0.738	0.435	3.24	2.026	1.77	-1.05	-0.36
Amazon	0.573	0.531	0.831	1.53	0.801	1.26	0.05	-0.59
Samsung	0.397	0.637	0.952	1.79	0.88	1.05	0.31	-0.32
Asus	0.511	0.432	0.523	0.79	0.55	0.65	0.64	0.15
Acer	0.233	0.378	0.222	0.62	1.05	1.03	0.22	-0.5

Table 0.9 Frequencies of product attributes of new tablet P.C

Product	Frequencies (%)							
	TM	SZ	SC	PP	DS	CM	BL	PR
Apple	28.2	23.5	27.2	33.2	23.5	28.9	17.6	20.3

Amazon	14.8	18.5	13.5	12.2	16.1	17.36	18.6	15.2
Samsung	30.8	25.9	23.5	19.2	24.2	15.6	29.8	27.6
Asus	13.3	15.8	19.7	21.3	19.3	19.4	15.6	21.3
Acer	12.8	16.3	16	14.1	16.8	18.7	18.4	15.6

Table 0.10 Importance of product attributes of new tablet P.C

	Importance							
	TM	SZ	SC	PP	DS	CM	BL	PR
Apple	2.37	2.68	2.28	3	2.67	2	3.02	3.16
Amazon	2.43	2.52	2.27	1.77	2.36	2.07	2.8	2.66
Samsung	3.59	3.08	2.7	2.26	2.78	2.2	3.52	3.3
Asus	2.35	2.54	2.53	2.48	0.25	2.46	2.43	2.69
Acer	2.33	2.58	2.66	2.38	2.22	2.24	2.69	2.71

Table 0.11 Fuzzy utility of new tablet P.C

Product (Centre)	Utility value
Ua	47.8083
Ub	20.9168
Uc	28.6961
Ud	19.4425
Ue	14.4908
Product (Left)	Utility value
Ua	32.0393
Ub	20.1569
Uc	27.1950
Ud	18.6254
Ue	14.3432
Product (Right)	Utility value
Ua	63.5772
Ub	38.3258
Uc	53.5798
Ud	34.3627
Ue	26.5467

Table 0.12 Market share of new tablet P.C

	Actual market share (%)	Centre (%)	Left (%)	Right (%)
Apple	71.5800	72.1690	71.5800	86.1270
Amazon	9.0233	12.1650	12.0000	0.5110
Samsung	10.4667	14.6600	11.5900	0.1460
Asus	0.6567	0.3529	0.0030	13.1857
Acer	0.5267	0.6326	4.0710	0.0130

Table 0.13 Fuzzy Market demand of Apple iPads

Tablet users in USA Apple iPad (Centre value)	Tablet users in USA Apple iPad (Left value)	Tablet users in USA Apple iPad (Right value)
14282740	11264650	15895200

Apple tablet P.C units sold in USA from 2016 were used to determine the parameters of the modified Bass model. The Apple tablet P.C units sold were used as data points and using Equation (6.15) the parameters of the modified Bass model were determined by solving an optimization problem using the nonlinear least squares. The parameters of the modified Bass model are shown in Table 6.14.

Table 0.14 Parameters of Bass model

Parameter	Value
m	1050
p	0.00136
q_o	0.477179
q_m	0.477179
γ	4.18872

Using the parameters p, q_o, q_m, the market demand (m) and the number of units previously adopted n(t), the number of Apple tablet P.C units adopted were determined, and the results are shown in Table 6.15 for the three scenarios namely, the normal case scenario, worse case scenarios and the best case scenarios corresponding to the centre, left and right value of the fuzzy demand, respectively. The forecasts of the adoptions were made for 10 quarterly periods. To evaluate the effectiveness of the proposed model, the non-modified Bass model shown in Equation 6.16 is used in this study to also forecasts demand of Apple tablet P.C and the results are compared with the proposed method. The results of the non-modified Bass model forecast is shown in Table 6.16. Fig. 6.4 compares the product adoptions based on the proposed model and the non-modified Bass model.

$$(t) = \frac{dN(t)}{dt} = [p + \frac{q}{m}N(t)][m - N(t)] \qquad (6.17)$$

Table 0.15 Results of sales forecasting under three scenarios using the modified Bass model

Quarter	Centre Normal case scenario forecast (Proposed Bass model)	Left Worst case scenario forecast (Proposed Bass model)	Right Best case scenario forecast (Proposed Bass model)
1	18158	18010	21670
4	26786	26568	31967
8	39471	39149	47105
12	49528	49124	59107

16	60223	59732	71871
20	69950	69379	83479
24	79492	78844	94813
28	88469	87747	105500
32	97070	96278	115783
36	105212	104354	125496

Table 0.16 Forecasted adoptions based on the non-modified Bass model

Quarter	Adoptions
1	24657
4	31992
8	41468
12	46429
16	50679
20	53385
24	55427
28	56819
32	57825
36	58528

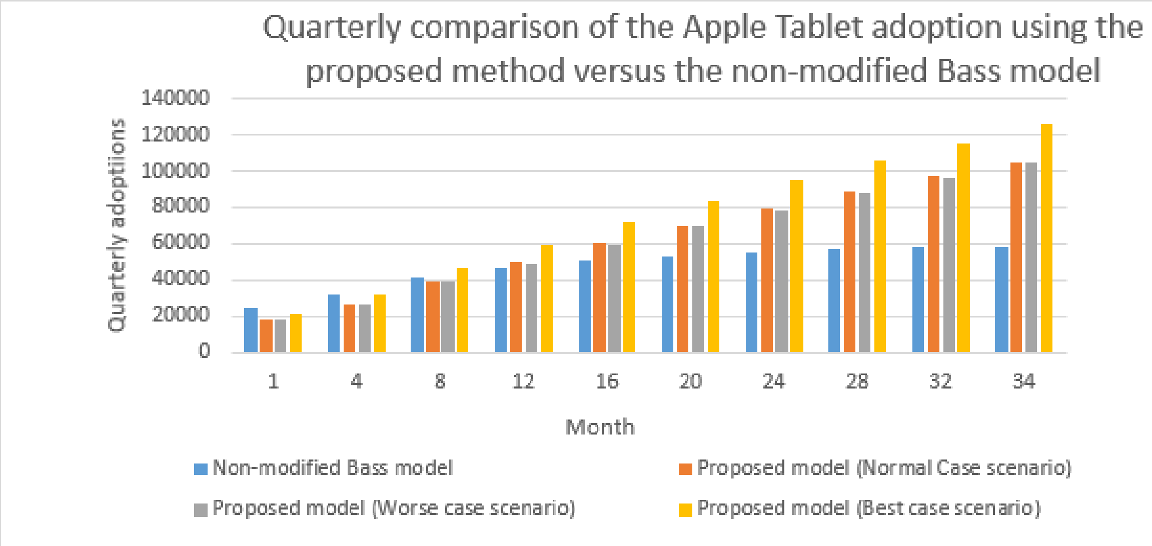

Figure 0.5 Comparison of the proposed method and the non-modified Bass model

The fuzzy time series method was also compared with the proposed method. With fuzzy time series method, the variable required was the past sales data. The results of the forecasts are shown in Table 6.17. Figure 6.6 shows the comparison between fuzzy time series and the proposed method.

Table 0.17 Forecasted adoptions based on the fuzzy time series method

Quarter	Fuzzy time series method
1	26578
4	26578
8	26578
12	26578
16	26578
20	26578
24	26578
28	26578
32	26578
34	26578

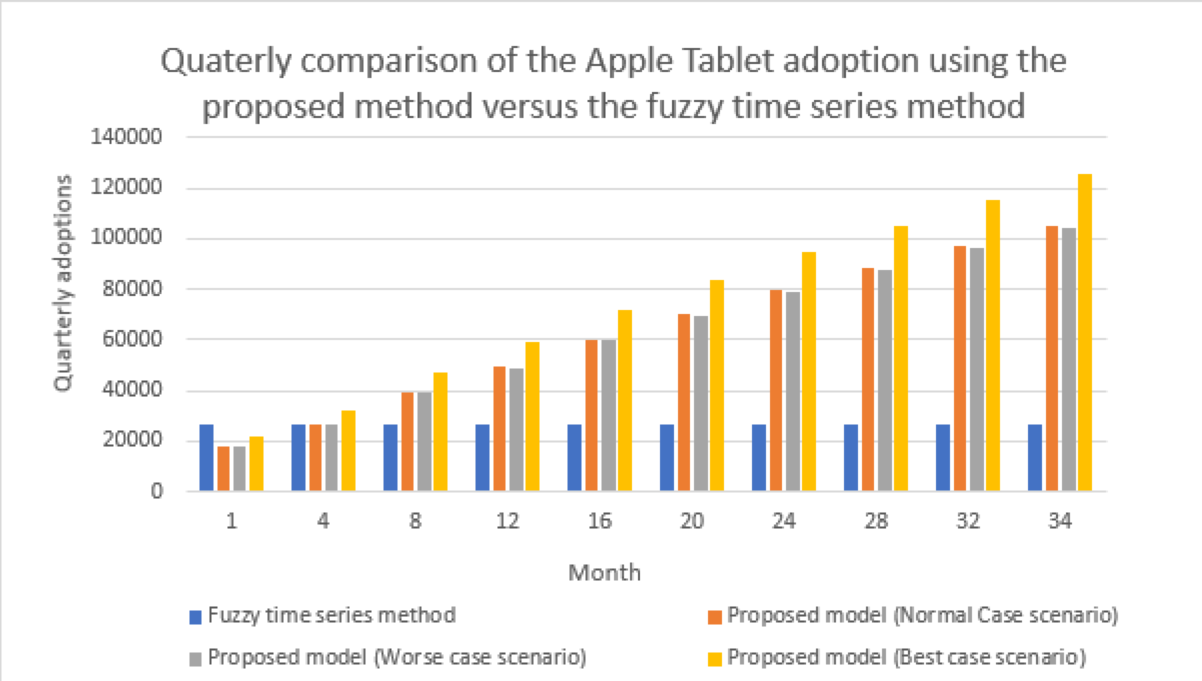

Figure 0.6 Comparison of the proposed method and the fuzzy time series method

Based on Figure 6.5, the Apple tablet P.C adoption for the three scenarios of the proposed modified Bass model showed a higher rise in demand contrary to the non-modified Bass model that showed a steady rise in adoptions. The trend in the proposed method could possibly be attributed to the covid-19 pandemic that forces people to stay or work from home thus increasing the demand for Apple tablet P.Cs. The worst case scenario and the normal case scenario of the proposed method showed similar adoption, hence providing a better means for forecasting the demand tablet P.C since the demand gap in the two scenarios is small. Also, the non-modified Bass model does not address the fuzziness in the product adoption forecast. A one-way ANOVA test was conducted to determine if any significance difference existed between the proposed model and the non-modified Bass model. The null hypothesis was defined as no significant difference existing between the proposed model and the non-modified Bass model. The results of the one-way ANOVA test in Table 6.18 shows that there is no statistically significant difference between the proposed model and the non-modified Bass model with the values F (3,36) = 1.662 and p = 0.192. As the p-value exceeded 0.05, the null hypothesis was not rejected. This implies close forecasting results between the proposed model and the non-modified Bass model. Thus, the proposed model can be used for forecasting the demand for a product with the advantage of providing the forecast demand the in three scenarios.

Table 0.18 Results of the one-way ANOVA test between the proposed method and non-modified Bass model

ANOVA					
Quarterly Forecast	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3931464448	3	1310488149	1.662	0.192
Within Groups	28381500109	36	788375003		
Total	32312964558	39			

Again, the proposed method was compared with the fuzzy time series method. The forecasted adoption based on the fuzzy time series, as seen in Figure 6.6 showed a stable adoption over time. This could be attributed using the historical sales data of the Apple tablet P.C which shows a decline in adoption before the advent of the COVID-19 pandemic. The forecasts as a result, reflect the past adoption patterns of the Apple tablet P.Cs. However, the adoption forecasts based on the proposed model showed a steadily rise in Apple tablet P.C adoption. This could also be attributed to the integration of the sentiment scores obtained from online reviews in the Bass model. One-way ANOVA testing was conducted to determine whether there was a significant difference between the proposed method and the fuzzy time series method. The null hypothesis was defined as there being no significant difference between the proposed model and the fuzzy time series method. The results of the one-way ANOVA test show that there is statistically a significant difference between the proposed model and the fuzzy time series with $F(3,36) = 5.968$ and $p = 0.002$. With the p-value less than 0.05, the null hypothesis was rejected.

The results of the one-way ANOVA Test are shown in Table 6.19. With the null hypothesis rejected, it can be inferred that when sentiment scores are not considered in the forecasting of the Apple tablet P.C adoption using the fuzzy time series method, the forecasts do not represent the reality of consumer adoption, hence leading to the inability of the fuzzy time series method to show the change in adoption during the COVID-19 pandemic.

Table 0.19 Results of one-way ANOVA test between the proposed method and the fuzzy time series method

ANOVA					
Quarterly Forecast	Sum of Squares	df	Mean Square	F	Sig.

Between Groups	13507628288	3	4502542763	5.968	0.002
Within Groups	27159512534	36	754430903.7		
Total	40667140822	39			

6.4 Summary

The Bass model has been used to forecast the demand/adoption for both new and existing products. However, in recent times the advent of big data obtained from consumers from the internet has been influencing how consumers purchase or adopt products. In this chapter, fuzzy and discrete choice modelling analysis methods for forecasting product adoption using online reviews were presented. In the proposed approach, online reviews are obtained from ecommerce websites and then processed using opinion mining to extract products attributes and their associated sentiment scores and frequencies of occurrence in the reviews. Using the fuzzy inference system (FIS), the importance weight of each product attributes was determined. Fuzzy utilities were generated from competitive products using FR to model relationships between ratings and sentiment scores of product attributes of competitive products. A market share model was developed using the MNL. Next, sentiment scores were integrated in the Bass model. By using the market share model developed and the modified Bass model, the demand for a product can be determined. A case study to forecast the adoption of Apple tablet P.Cs was conducted based on the proposed approach. The forecast results were then compared to the non-modified Bass model and the fuzzy time series. The proposed approach had similar forecasts to the non-modified Bass model indicating that proposed method can be used to forecast product adoption. Moreover, the proposed method presented the product adoption in three scenarios to allow manufacturers to plan adequately with consideration of the scenarios presented in the proposed method.

Chapter 7 Discussion

This chapter discusses the proposed methodologies, implementation, and results.

7.1 Discussion on the proposed methodology for a multigene genetic programming based fuzzy regression approach for modelling customer satisfaction based on online reviews

In this research, the methodology for modelling customer satisfaction based on online reviews was developed using the proposed MGGP-FR. Online reviews of products were used because they were user-generated contents and did not rely on a predefined set of questions to elicit customer needs as seen in traditional questionnaire based surveys methods. Moreover, the volume and up to date-nature of online reviews from the internet makes it easy to determine customer needs in a timely manner compared to questionnaire-based surveys. The MGGP-FR was adopted to develop the fuzzy non-linear customer satisfaction model to deal with uncertainties and non-linearities in modelling customer satisfaction. To develop the chromosomes of the MGGP-FR, the maximum tree depths greater than 4 were tested, but bloating phenomena were experienced thus the maximum depth of 4 was set with three genes of each individual. Also, to obtain a better forecasting accuracy, the value of h in the FR was selected based on the value that gave the least error. The validation tests conducted in section 4.3 showed that the proposed method had better prediction performance when compared to the FR, GP, and GP-FR method. The proposed method has the potential to be applied to other customer satisfaction modelling scenarios where there exist high levels of fuzziness and non-linearities.

Product manufacturers can consider proposed MGGP-FR when resources are scarce or limited. The MGGP-FR is used to enhance prediction accuracy. Redundant product attributes or features, which have little contribution in enhancing customer satisfaction, are eliminated. Moreover, the proposed method for this study phase will be most suited and applicable to food blenders, computers, mobiles, children's toys, headphones etc. These are products used by consumers regularly and can warrant their desire and need to express their sentiments on online platforms. The scope of application of the proposed MGGP-FR for customer satisfaction modelling can be used to identify and develop new product features integrated into existing products. This is because customer concerns from online reviews on existing products are required to develop the customer satisfaction model. Nevertheless, with the adoption of the proposed method, companies can expand their product family in order to increase revenue and profits.

7.2 Discussion on determining and forecasting the importance of product attributes using online customer reviews and Google Trends

As consumers' needs become more dynamic with rapid changes in technologies, this study proposed a methodology to determine and forecast the importance of products using online reviews and Google Trends. Online reviews were adopted for this study because new and updated user-generated content related to products can easily be obtained from online platforms. The online review platforms provide a channel for consumers to relay their true feelings and concerns regarding a product. Moreover, the Google Trends data adopted in this study was used to consider consumer search behaviour in product attributes weightage. There are some situations where consumers are only interested in certain product attributes and a product's related information. These investigations by consumers are sometimes carried out in search engines. Since Google's search engine has the largest market share of all the search engines, especially in the USA, this study extracted consumer search data related to product attributes from Google Trends.

This study looked at how consumers perceived the importance or relevance of certain product attributes in existing products. However, the determination of the importance of product attributes improves existing products and helps product manufacturers design entirely new products with desired and improved features. Since online reviews and Google Trends are two different metrics, there is the need to allocate value to each metric to determine how they can contribute to product attribute importance determination. Hence, the Shapley value was used to determine the value of the two online metrics in the product attributes importance determination. To predict which product attributes will be more critical in the future based on consumer search behaviour and online reviews, a fuzzy rough set time series method is used to forecast the importance of product attributes. The rough set enabled the forecasts to be made from if-then rules by eliminating irrelevant attributes and identifying structural relationships in the data.

The proposed method in this study can be applied to products with very dynamic markets such as fast fashions, electronic devices, products, etc to avoid obsolescence. The validation results of this study in section 5.3 indicate that the proposed method can give a better prediction result than fuzzy time series, fuzzy k medoid clustering time, and ANFIS. This is necessary for product manufacturers to mitigate the problems of wrongly predicting future consumers' needs. Thus, the real key to ensuring the success of a new or existing product is to ensure that accurate information of consumers' desires for a product at different times period is conveyed at the right time.

7.3 Discussion on a fuzzy and discrete choice modelling analysis method for forecasting product adoption using online reviews

This research explored online reviews in forecasting the adoption of new products. For many extant studies, market share models for different product profiles have been developed from discrete choice models such as the MNL. The over-reliance on questionnaire-based surveys to develop these market share models results in the delay in determining market shares of different products. In addition, the problems of answering predefined questions in the surveys is due to the misinterpretation by the respondents.

However, this study integrated online reviews in MNL to mitigate the effects of using the conventional approach of developing market share models of the MNL. The online reviews conveyed the real sentiments from consumers since they were users' input and did not involve any predefined questions.

Also, to address the issue of uncertainty in determining the market shares of products, this study developed a fuzzy market share model to determine the market shares of products in three scenarios: the worst-case scenario, the typical case scenario, and the best-case scenario. Product manufacturers can adopt this market share determination strategy to plan which product profiles will most likely generate the largest market share. The proposed method can be complemented with questionnaire-based surveys to develop a new product.

The Bass model in this study was modified to accommodate how online reviews drive product adoption. This was done by modifying the coefficient of imitation in the Bass model. However, no modification was made to the coefficient of innovation of the Bass model. The forecast for product adoption was done by using the fuzzy market share model developed in this research to determine the market potential of a product and using the market potential in the modified Bass model to forecast the demand for a product. Because the market share model was a fuzzy model, the final product adoption forecasts addressed fuzziness and uncertainties by forecasting in different scenarios. This can provide substantial managerial implications in an unpredictable market, especially for technology-related products. This research is also appropriate for short to medium-term forecasting; and the results of this study are shown in Figure 6.5. There were significant changes in adoptions patterns between the modified Bass model and non-modified Bass model from the 20^{th} period. It should be noted that this study is applicable to both new and existing products. Some variants of the Bass model include factors such as price, which is not considered in this study and could be considered in future work.

Chapter 8 Conclusions, Limitations and Future Work

This chapter presents the conclusions and limitations and highlights this research's significant contribution. Suggestions for future work are presented.

8.1 Conclusions

Many scholars and researchers are calling for the active participation of consumers in the product development process. This call is driven by the need for firms to keep innovating to remain competitive while maintaining key competence in their field. Consumers' participation in the product development process can be facilitated by a large volume of data generated from the internet user, such as blogs, e-commerce platforms, online reviews websites, etc. However, the lack of rigorous and sustainable methods that firms can adopt to remain competitive remains a barrier for firms to improve their business model in a data-driven approach. A survey by New Venture Partners, a venture capitalist firm in the USA, showed that businesses recorded to be data-driven dropped from 37% in 2017 to 31% in 2019. The results from the survey indicate the lack of understanding about how to integrate consumers' online data into their operations.

However, the outcomes of this study present a guide for product manufacturers to plan and design products that can potentially increase the loyalty of consumers to their brands. Moreover, this study's adoption can enable firms to make efficient uses of resources by making informed decisions on resource allocation. In particular, the novel MGGP-FR presented in this study can eliminate product attributes that cannot enhance customer satisfaction while maintaining the relevant product attributes that will increase customer satisfaction. Although there was a time lag between data collection and the time for predicting customer satisfaction with the MGGP-FR, the FR addressed the uncertainty in the prediction due to time lag.

The proposed methodology for determining the importance and future importance of product attributes provides a reliable approach for product manufacturers to keep up with the dynamic needs of consumers. This is particularly important for product manufacturers in the technology aspect to maintain their relevance and business success. The fuzzy rough set time series used in forecasting the future importance of product attributes generate forecasting rules that are easily interpretable and understandable; hence product manufacturers can adopt this method in their product planning actions.

Regarding the proposed market share model and the modified Bass model in Chapter 6, product adoptions forecasts can be conducted for unpredictable markets for specific products. Three scenarios of the forecast can enable product manufacturers to plan their capacity and inventory to manage the resources efficiently. Since consumers' online reviews were used to develop the product adoption forecasting model, product manufacturers should be confident in adopting this method since it conveys consumers' true feelings towards a product. The development of an entirely new product can supplement the proposed method with questionnaire-based surveys and develop new ideas for a product.

The major contribution of the research are summarised as follows:

- A novel methodology for modelling customer satisfaction from online reviews using a MGGP-FR is proposed to address the fuzziness and non-linearities in customer satisfaction modelling. Moreover, by conducting opinion mining on online reviews, customers' needs, and their associated sentiments could be easily obtained for the customer satisfaction modelling. The proposed method outperformed existing methods when compared.

- An improved method of determining the importance of product attributes as well predicting the future importance of product attributes from online reviews and Google Trends is presented. The Shapley value and Choquet integral were used to determine the importance of product attributes, whiles the fuzzy rough set time series was used to predict the future importance of product attributes. The proposed methods had better forecasting accuracy when compared to existing methods. The successful development of the proposed method can be applied in the development of product designs that can eliminate the wastage of resources on product attributes that have value to consumers. The method can also facilitate the modern approach of integrating online data from consumers in new product development process in order to come up with optimal product designs specifications without compromising on the quality and customer satisfaction.

- A new approach for forecasting product adoption under uncertainty using a new market share model and a modified Bass model was developed and presented in this study. Online reviews were integrated into the market share and Bass model to forecast product adoption in different scenarios. The results of the forecasting were compared to the original Bass model and the fuzzy times series method. From the ANOVA test, the forecasts of the original Bass model were similar to the proposed model, hence indicating that the proposed method could be used to forecast the demand for a product with the advantage of generating demand forecast in different scenarios. The results of the forecasts using the proposed method highlights the flexibility of the Bass model in being extended to accommodate different conditions and scenarios. Besides, the proposed method presented enhances the decision making process of manufacturers on how to meet the demands of consumers. More importantly, with limited past sales data, product manufacturers can make demand forecasts for both new and existing products under conditions of uncertainty.

8.2 Limitation of studies

The limitations of this research are highlighted as follows:

- In all the three methods proposed, extracting of product attributes and their associated sentiment scores from online reviews were limited to product attributes that were explicitly mentioned in the reviews. However, some review bear implicit expressions which were excluded from the reviews processing because the sentiment analysis tool used was limited in its ability to identify the meaning of these expressions. This could be improved by using an improved sentiment analysis algorithm or tool that can detect implicit expressions in order to obtain more categories of product attributes.

- In chapter five, the proposed method for predicting the future importance of product attributes required the selection of parameters for the fuzzy rough set time series method based on expert knowledge. This has the potentially to introduce bias in the selection of the parameters.
- Lastly, in chapter six, the model developed for forecasting the demand for a product was limited and was more suitable for a specific geographic location. However, for the model be adaptable to any location, other factors, or variables and a wide scope of consumer review data need to be considered in the Bass model.

8.3 Future work

Although this research presented improved methods of integrating consumer data from online reviews and google search engine in product development process, further improvements can made by addressing the following areas in future work:

- In this research, consumer requirements obtained from online reviews were used to develop customer satisfaction model. However, opportunities exist to convert these customer requirements into technical product attributes. Future studies can focus on methods that can be used to select the technical product attributes settings. This will show the potential of enabling product manufacturers to design products with optimal technical product attributes settings without compromising on customer satisfaction

- This research mainly focused on a single line of products in customer satisfaction modelling and product attributes importance determination. If a family of product is considered, our proposed methods will be adaptable on a wide scope of products without limiting them to single line of products. Moreover, by considering product families in future studies, product manufacturers can potentially increase their market share when a wide scope of products is considered. Similar to the goal of increasing the market share for product manufacturers, future work can also focus on how to leverage online reviews for market segmentation. By determining the different market segments from online reviews, new ideas for new products can be obtained from the different market segments by product manufacturers.

- In our current research, the product attributes and consumers requirements were extracted from online reviews with explicit expression. Thus, reviews with implicit or latent expression were excluded from the sentiment analysis due to the sentiment analysis tools adopted. Future work can also focus on the development of improved algorithms that can detect and include implicit expressions in the reviews in the sentiment analysis process. This future study is worth considering as consumers opinions tend to be textually complex but harbour large amounts of useful information that can assist product manufacturers in designing products with customer satisfaction.

- To minimise the manual selection of parameters in the fuzzy rough set time series method in Chapter 5, an evolutionary algorithm will be used in future studies to address the manual selection of the parameters by the user in order to limit any bias.

Appendix A. Universe of discourse for product attributes of an electric hairdryer.

Table A.1. Universe of discourse for product attribute “Control settings”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 2]	[0 2]	[0 2]	[-4 -2]	[-5 -3]	[0 2]
U2	[2 4]	[2 4]	[2 4]	[-2 0]	[-3 -1]	[2 4]
U3	[4 6]	[4 6]	[4 6]	[0 2]	[-1 1]	[4 6]
U4	[6 8]	[6 8]	[6 8]	[2 4]	[1 3]	[6 8]
U5	[8 10]	[8 10]	[8 10]	[4 6]	[3 5]	[8 10]

Table A.2. Universe of discourse for product attribute “Airflow”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1	[0 1]	[-3 -1]	[-9 -6]	[0 2]
U2	[2 3]	[2 3]	[2 3]	[-1 1]	[-6 -3]	[2 4]
U3	[3 4]	[3 4]	[3 4]	[1 3]	[-3 0]	[4 6]
U4	[4 5]	[4 5]	[4 5]	[3 5]	[0 3]	[6 8]
U5	[5 6]	[5 6]	[5 6]	[5 7]	[3 6]	[8 10]
U6	[6 7]	[6 7]	[6 7]	[7 9]	[6 9]	
U7	[7 8]	[7 8]	[7 8]			

Table A.3. Universe of discourse for product attribute “Weight”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-4 -2]	[-6 -4]]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-2 0]	[-4 -2]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[0 2]	[-2 0]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[2 4]	[0 2]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[4 6]	[2 4]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[6 8]	[4 6]	
U7	[6 7]	[6 7]	[6 7]	[8 10]		
U8	[7 8]	[7 8]	[7 8]			

Table A.4. Universe of discourse for product attribute “Usability”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-2 0]	[-5 -3]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[0 2]	[-3 -1]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[2 4]	[-1 1]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[4 6]	[1 3]	[6 8]
U5	[4 5]	[4 5]	[4 5]		[3 5]	[8 10]
U6	[5 6]	[5 6]	[5 6]		[5 7]	
U7	[6 7]	[6 7]	[6 7]			
U8	[7 8]	[7 8]	[7 8]			

Table A.5. Universe of discourse for product attribute “Noise”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-3 -2]	[-6 -4]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-2 -1]	[-4 -2]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[-1 0]	[-2 0]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[0 1]	[0 2]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[1 2]	[2 4]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[2 3]	[4 6]	
U7	[6 7]	[6 7]	[6 7]	[3 4]		
U8	[7 8]	[7 8]	[7 8]	[4 5]		

Table A.6. Universe of discourse for product attribute “Price”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 2]	[0 2]	[0 2]	[-6 -2]	[-8 -6]	[0 2]
U2	[2 4]	[2 4]	[2 4]	[-4 -2]	[-6 -4]	[2 4]
U3	[4 6]	[4 6]	[4 6]	[-2 0]	[-4 -2]	[4 6]
U4	[6 8]	[6 8]	[6 8]	[0 2]	[-2 0]	[6 8]
U5	[8 10]	[8 10]	[8 10]	[2 4]	[0 2]	[8 10]
U6				[4 6]	[2 4]	
U7					[4 6]	

Table A.7. Universe of discourse for product attribute “Easy to use”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-4 -2]	[- 6 -4]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-2 0]	[-4 -2]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[0 2]	[-2 0]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[2 4]	[0 2]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[4 6]	[2 4]	[8 10]
U6	[5 6]	[5 6]	[5 6]		[4 6]	
U7	[6 7]	[6 7]	[6 7]			
U8	[7 8]	[7 8]	[7 8]			

Table A.8. Universe of discourse for product attribute “Power”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-3 -1]	[-9 -6]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-1 1]	[-6 -3]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[1 3]	[-3 0]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[3 5]	[0 3]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[5 7]	[3 6]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[7 9]	[6 9]	
U7	[6 7]	[6 7]	[6 7]			
U8	[7 8]	[7 8]	[7 8]			

Table A.9. Universe of discourse for product attribute “Comfortable to hold”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-3 -2]	[-8 -6]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-2 -1]	[-6 -4]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[-1 0]	[-4 -2]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[0 1]	[-2 0]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[1 2]	[0 2]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[2 3]	[2 4]	
U7	[6 7]	[6 7]	[6 7]	[3 4]	[4 6]	
U8	[7 8]	[7 8]	[7 8]	[4 5]	[6 8]	
				[5 6]		
				[6 7]		

Table A.10. Universe of discourse for product attribute “Durability”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-5 -3]	[-6 -4]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-3 -1]	[-4 -2]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[-1 1]	[-2 0]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[1 3]	[0 2]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[3 5]	[2 4]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[5 7]	[4 6]	
U7	[6 7]	[6 7]	[6 7]			
U8	[7 8]	[7 8]	[7 8]			

Table A.11. Universe of discourse for product attribute “Portability”

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-6 -4]	[-6 -4]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-4 -2]	[-4 -2]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[-2 0]	[-2 0]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[0 2]	[0 2]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[2 4]	[2 4]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[4 6]	[4 6]	
U7	[6 7]	[6 7]	[6 7]			
U8	[7 8]	[7 8]	[7 8]			

Table A.12. Universe of discourse for product attribute “Efficiency

	LAG1	LAG2	LAG3	M	S	Importance
U1	[0 1]	[0 1]	[0 1]	[-4 -3]	[-7 -5]	[0 2]
U2	[1 2]	[1 2]	[1 2]	[-3 -1]	[-5 -3]	[2 4]
U3	[2 3]	[2 3]	[2 3]	[-1 1]	[-3 -1]	[4 6]
U4	[3 4]	[3 4]	[3 4]	[1 3]	[-1 1]	[6 8]
U5	[4 5]	[4 5]	[4 5]	[3 5]	[1 3]	[8 10]
U6	[5 6]	[5 6]	[5 6]	[5 7]	[3 5]	
U7	[6 7]	[6 7]	[6 7]		[5 7]	
U8	[7 8]	[7 8]	[7 8]			
		[8 9]				

Appendix B. Data for product attributes

Table B.1. Data for the product attribute "Control settings"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	3.506	3.506
3.506	0.000	0.000	1.282	-2.224
4.788	3.506	0.000	-0.049	-1.331
4.739	4.788	3.506	2.748	2.798
7.487	4.739	4.788	-1.963	-4.711
5.525	7.487	4.739	-1.068	0.895
4.457	5.525	7.487	-0.512	0.556
3.944	4.457	5.525	-2.266	-1.754
1.678	3.944	4.457	0.142	2.408
1.820	1.678	3.944	-0.084	-0.225
1.736	1.820	1.678	0.539	0.622
2.275	1.736	1.820	2.191	1.652
4.466	2.275	1.736	-1.798	-3.989
2.668	4.466	2.275	1.194	2.992
3.862	2.668	4.466	-2.305	-3.498
1.558	3.862	2.668		
	1.558	3.862		
		1.558		

Table B.2. Data for the product attribute “Airflow”

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	7.018	7.018
7.018	0.000	0.000	-1.548	-8.566
5.470	7.018	0.000	1.200	2.748
6.670	5.470	7.018	1.091	-0.109
7.760	6.670	5.470	-2.314	-3.405
5.447	7.760	6.670	-0.300	2.014
5.147	5.447	7.760	-1.855	-1.555
3.293	5.147	5.447	-0.429	1.426
2.864	3.293	5.147	0.921	1.350
3.785	2.864	3.293	2.634	1.713
6.420	3.785	2.864	-1.612	-4.246
4.808	6.420	3.785	-0.907	0.705
3.900	4.808	6.420	0.109	1.017
4.010	3.900	4.808	0.257	0.148
4.267	4.010	3.900	-0.248	-0.505
4.020	4.267	4.010		
	4.020	4.267		
		4.020		

Table B.3. Data for the product attribute “weight”

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	5.492	5.492
5.492	0.000	0.000	0.673	-4.820
6.165	5.492	0.000	-1.218	-1.891
4.947	6.165	5.492	2.174	3.392
7.121	4.947	6.165	-3.027	-5.201
4.094	7.121	4.947	1.410	4.437
5.504	4.094	7.121	-0.803	-2.214
4.701	5.504	4.094	-2.062	-1.258
2.639	4.701	5.504	1.496	3.558
4.135	2.639	4.701	1.475	-0.021
5.609	4.135	2.639	-3.376	-4.851
2.234	5.609	4.135	0.290	3.666
2.524	2.234	5.609	1.778	1.488
4.302	2.524	2.234	0.252	-1.526
4.554	4.302	2.524	-0.333	-0.585
4.221	4.554	4.302		
	4.221	4.554		
		4.221		

Table B.4. Data for the product attribute "Usability"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	5.667	5.667
5.667	0.000	0.000	1.541	-4.125
7.208	5.667	0.000	-0.846	-2.388
6.362	7.208	5.667	1.152	1.999
7.514	6.362	7.208	-1.822	-2.974
5.692	7.514	6.362	-1.193	0.629
4.499	5.692	7.514	-2.048	-0.855
2.451	4.499	5.692	0.265	2.313
2.716	2.451	4.499	1.347	1.082
4.063	2.716	2.451	-0.947	-2.294
3.116	4.063	2.716	-0.818	0.129
2.298	3.116	4.063	1.334	2.152
3.632	2.298	3.116	-0.766	-2.100
2.866	3.632	2.298	-1.860	-1.094
1.005	2.866	3.632	1.832	3.692
2.837	1.005	2.866		
	2.837	1.005		
		2.837		

Table B.5. Data for the product attribute "Noise"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	4.754	4.754
4.754	0.000	0.000	-0.746	-5.500
4.009	4.754	0.000	2.470	3.216
6.479	4.009	4.754	0.751	-1.719
7.230	6.479	4.009	-1.558	-2.309
5.672	7.230	6.479	-1.426	0.131
4.245	5.672	7.230	-2.341	-0.914
1.905	4.245	5.672	-0.761	1.580
1.143	1.905	4.245	3.093	3.854
4.237	1.143	1.905	-2.683	-5.777
1.553	4.237	1.143	0.291	2.975
1.844	1.553	4.237	0.725	0.433
2.569	1.844	1.553	-0.801	-1.526
1.768	2.569	1.844	1.457	2.258
3.225	1.768	2.569	-1.503	-2.960
1.721	3.225	1.768		
	1.721	3.225		
		1.721		

Table B.6. Data for the product attribute "Price"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	4.929	4.929
4.929	0.000	0.000	4.079	-0.965
8.893	4.929	0.000	-4.752	-7.897
4.959	8.893	4.929	0.300	3.488
4.514	4.959	8.893	-1.116	-0.880
3.189	4.514	4.959	-0.776	0.484
2.347	3.189	4.514	-0.749	0.119
1.623	2.347	3.189	-0.071	0.643
1.544	1.623	2.347	-0.499	-0.378
1.086	1.544	1.623	0.456	0.683
1.312	1.086	1.544	0.421	0.354
1.891	1.312	1.086	1.127	0.550
3.020	1.891	1.312	-1.245	-2.384
1.765	3.020	1.891	-0.756	0.497
1.008	1.765	3.020	-0.916	-0.169
0.082	1.008	1.765		
	0.082	1.008		
		0.082		

Table B.7. Data for the product attribute "Easy to use"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	4.015	4.015
4.015	0.000	0.000	0.373	-3.642
4.387	3.793	0.000	-1.175	-1.547
3.213	4.381	4.015	3.280	4.455
6.493	3.200	4.387	1.386	-1.894
7.879	3.696	3.213	-3.675	-5.061
4.204	7.274	6.493	0.050	3.724
4.254	3.797	7.879	-2.110	-2.159
2.144	4.163	4.204	2.689	4.798
4.833	2.075	4.254	-2.557	-5.246
2.276	4.304	2.144	0.209	2.766
2.485	2.275	4.833	-0.739	-0.948
1.746	2.101	2.276	0.420	1.160
2.166	1.746	2.485	-0.171	-0.591
1.995	2.131	1.746	1.186	1.357
3.181	1.992	2.166		
	3.154	1.995		
		3.181		

Table B.8. Data for the product attribute “Power”

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	7.210	7.210
7.210	0.000	0.000	-0.876	-8.086
6.334	7.210	0.000	-2.116	-1.240
4.219	6.334	7.210	2.264	4.380
6.483	4.219	6.334	0.684	-1.580
7.167	6.483	4.219	-2.527	-3.211
4.640	7.167	6.483	-1.061	1.466
3.579	4.640	7.167	-0.484	0.577
3.095	3.579	4.640	0.709	1.193
3.804	3.095	3.579	-1.644	-2.353
2.161	3.804	3.095	1.594	3.237
3.755	2.161	3.804	1.535	-0.059
5.290	3.755	2.161	-0.228	-1.763
5.062	5.290	3.755	-0.008	0.220
5.053	5.062	5.290	-2.099	-2.091
2.954	5.053	5.062		
	2.954	5.053		
		2.954		

Table B.9. Data for the product attribute “Comfortable to hold”

LAG1	LAG2	LAG3	Moment	slope
0.000	0.000	0.000	6.479	6.479
6.479	0.000	0.000	-1.316	-7.795
5.163	6.479	0.000	-1.840	-0.524
3.323	5.163	6.479	2.698	4.538
6.021	3.323	5.163	1.363	-1.336
7.384	6.021	3.323	-1.291	-2.653
6.093	7.384	6.021	-0.994	0.297
5.099	6.093	7.384	-1.851	-0.857
3.249	5.099	6.093	2.163	4.013
5.411	3.249	5.099	0.096	-2.067
5.508	5.411	3.249	-2.215	-2.311
3.293	5.508	5.411	-0.430	1.785
2.862	3.293	5.508	1.121	1.551
3.983	2.862	3.293	0.903	-0.218
4.886	3.983	2.862	-2.457	-3.360
2.429	4.886	3.983		
	2.429	4.886		
		2.429		

Table B.10. Data for the product attribute "Durability"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	5.494	5.494
5.494	0.000	0.000	-0.473	-5.967
5.021	5.494	0.000	-1.056	-0.583
3.965	5.021	5.494	3.102	4.158
7.067	3.965	5.021	-0.312	-3.415
6.755	7.067	3.965	-4.276	-3.964
2.478	6.755	7.067	0.133	4.409
2.611	2.478	6.755	-0.344	-0.477
2.267	2.611	2.478	-0.117	0.227
2.150	2.267	2.611	3.050	3.167
5.200	2.150	2.267	-0.123	-3.173
5.076	5.200	2.150	-0.334	-0.211
4.743	5.076	5.200	-1.792	-1.458
2.951	4.743	5.076	-0.431	1.361
2.519	2.951	4.743	2.072	2.503
4.591	2.519	2.951		
	4.591	2.519		
		4.591		

Table B.11. Data for the product attribute "Portability"

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	4.524	4.524
4.524	0.000	0.000	1.127	-3.397
5.650	4.524	0.000	-1.196	-2.323
4.454	5.650	4.524	1.182	2.378
5.637	4.454	5.650	1.979	0.797
7.616	5.637	4.454	-1.459	-3.438
6.157	7.616	5.637	-1.880	-0.422
4.277	6.157	7.616	-0.616	1.265
3.661	4.277	6.157	0.371	0.987
4.033	3.661	4.277	2.218	1.847
6.251	4.033	3.661	-3.631	-5.849
2.619	6.251	4.033	-1.915	1.716
0.704	2.619	6.251	1.049	2.965
1.753	0.704	2.619	1.281	0.231
3.034	1.753	0.704	-1.848	-3.128
1.186	3.034	1.753		
	1.186	3.034		
		1.186		

Table B.12. Data for the product attribute “Efficiency”

LAG1	LAG2	LAG3	Moment	Slope
0.000	0.000	0.000	6.325	6.325
6.325	0.000	0.000	2.534	-3.791
8.860	6.325	0.000	-0.138	-2.672
8.722	8.860	6.325	-1.504	-1.367
7.218	8.722	8.860	-0.230	1.274
6.987	7.218	8.722	-1.006	-0.775
5.982	6.987	7.218	0.187	1.192
6.168	5.982	6.987	-2.117	-2.304
4.051	6.168	5.982	1.716	3.834
5.767	4.051	6.168	-2.788	-4.504
2.979	5.767	4.051	-0.288	2.500
2.691	2.979	5.767	0.527	0.815
3.218	2.691	2.979	0.811	0.283
4.029	3.218	2.691	0.753	-0.058
4.782	4.029	3.218	-0.987	-1.740
3.795	4.782	4.029		
	3.795	4.782		
		3.795		

Appendix C. Rules generated for forecasting the product attributes

Table C.1. Rules generated for "Control settings"

LAG1(A)(A5) AND LAG2(B)(B4) => I(P6)
LAG1(A)(A6) AND LAG2(B)(B5) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B6) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B4) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B4) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B3) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) => I(P1) OR I(P2)
LAG1(A)(A2) AND LAG2(B)(B1) => I(P7)
LAG1(A)(A7) AND LAG2(B)(B2) => I(P1)
LAG1(A)(A5) AND LAG2C(C6) => I(P6)
LAG1(A)(A6) AND LAG2C(C4) => I(P4)
LAG1(A)(A4) AND LAG2C(C5) => I(P4)
LAG1(A)(A4) AND LAG2C(C6) => I(P3)
LAG1(A)(A3) AND LAG2C(C4) => I(P1)
LAG1(A)(A1) AND LAG2C(C4) => I(P1)
LAG1(A)(A1) AND LAG2C(C3) => I(P1)
LAG1(A)(A1) AND LAG2C(C1) => I(P2)
LAG1(A)(A2) AND LAG2C(C1) => I(P7)
LAG1(A)(A7) AND LAG2C(C1) => I(P1)
LAG1(A)(A5) AND V(V4) => I(P6)
LAG1(A)(A6) AND V(V3) => I(P4)
LAG1(A)(A4) AND V(V3) => I(P4)
LAG1(A)(A4) AND V(V4) => I(P3)
LAG1(A)(A3) AND V(V3) => I(P1)
LAG1(A)(A1) AND V(V4) => I(P1) OR I(P2)
LAG1(A)(A2) AND V(V6) => I(P7)
LAG1(A)(A7) AND V(V2) => I(P1)
LAG1(A)(A5) AND S(S1) => I(P6)
LAG1(A)(A6) AND S(S6) => I(P4)
LAG1(A)(A4) AND S(S5) => I(P4)
LAG1(A)(A4) AND S(S4) => I(P3)
LAG1(A)(A3) AND S(S5) => I(P1)
LAG1(A)(A1) AND S(S5) => I(P1)
LAG1(A)(A1) AND S(S4) => I(P1)
LAG1(A)(A1) AND S(S6) => I(P2)
LAG1(A)(A2) AND S(S5) => I(P7)
LAG1(A)(A7) AND S(S6) => I(P1)

Table C.2 Rules generated for "Airflow"

LAG1(A)(A2) AND V(V1) => I(P2)
LAG1(A)(A2) AND V(V3) => I(P4) OR I(P1)
LAG1(A)(A4) AND V(V4) => I(P2)
LAG1(A)(A2) AND V(V5) => I(P3) OR I(P2)
LAG1(A)(A3) AND V(V3) => I(P2)
LAG1(A)(A2) AND V(V4) => I(P3)
LAG1(A)(A1) AND V(V4) => I(P2)
S(S3) => I(P2) OR I(P4)
S(S4) => I(P2)
S(S5) => I(P3) OR I(P2)
S(S1) => I(P2)
S(S2) => I(P1)
LAG1(A)(A2) AND LAG2(B)(B3) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND S(S3) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B2) AND S(S4) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B4) AND S(S5) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) AND S(S1) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) AND S(S5) => I(P3) OR I(P2)
LAG1(A)(A3) AND LAG2(B)(B2) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND S(S2) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B2) AND S(S4) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND S(S3) => I(P4)
LAG1(A)(A4) AND LAG2C(C2) AND S(S4) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND S(S5) => I(P3) OR I(P2)
LAG1(A)(A3) AND LAG2C(C4) AND S(S1) => I(P2)
LAG1(A)(A3) AND LAG2C(C3) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND S(S2) => I(P1)
LAG1(A)(A1) AND LAG2C(C2) AND S(S4) => I(P2)
LAG1(A)(A2) AND V(V1) AND S(S3) => I(P2)
LAG1(A)(A2) AND V(V3) AND S(S3) => I(P4)
LAG1(A)(A4) AND V(V4) AND S(S4) => I(P2)
LAG1(A)(A2) AND V(V5) AND S(S5) => I(P3) OR I(P2)
LAG1(A)(A3) AND V(V3) AND S(S1) => I(P2)
LAG1(A)(A2) AND V(V4) AND S(S5) => I(P3)
LAG1(A)(A3) AND V(V3) AND S(S3) => I(P2)

Table C.3 Rules generated for “Weight”

LAG1(A)(A6) AND LAG2(B)(B3) => I(P6)
LAG1(A)(A6) AND LAG2(B)(B6) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B6) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B4) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B3) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B1) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) => I(P2)
LAG1(A)(A6) AND LAG2C(C6) => I(P6)
LAG1(A)(A6) AND LAG2C(C3) => I(P4)
LAG1(A)(A4) AND LAG2C(C6) => I(P3)
LAG1(A)(A3) AND LAG2C(C6) => I(P1)
LAG1(A)(A1) AND LAG2C(C4) => I(P1)
LAG1(A)(A1) AND LAG2C(C3) => I(P2)
LAG1(A)(A2) AND LAG2C(C1) => I(P3)
LAG1(A)(A3) AND LAG2C(C1) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) => I(P2)
LAG2(B)(B3) AND LAG2C(C6) => I(P6)
LAG2(B)(B6) AND LAG2C(C3) => I(P4)
LAG2(B)(B6) AND LAG2C(C6) => I(P3)
LAG2(B)(B4) AND LAG2C(C6) => I(P1)
LAG2(B)(B3) AND LAG2C(C4) => I(P1)
LAG2(B)(B1) AND LAG2C(C3) => I(P2)
LAG2(B)(B1) AND LAG2C(C1) => I(P3)
LAG2(B)(B2) AND LAG2C(C1) => I(P2)
LAG2(B)(B3) AND LAG2C(C2) => I(P2)
LAG2(B)(B2) AND LAG2C(C3) => I(P2)
LAG2(B)(B3) AND S(S1) => I(P6)
LAG2(B)(B6) AND S(S5) => I(P4)
LAG2(B)(B6) AND S(S3) => I(P3)
LAG2(B)(B4) AND S(S3) => I(P1)
LAG2(B)(B3) AND S(S4) => I(P1) OR I(P2)
LAG2(B)(B1) AND S(S3) => I(P2)
LAG2(B)(B1) AND S(S4) => I(P3)
LAG2(B)(B2) AND S(S4) => I(P2)
LAG2(B)(B2) AND S(S3) => I(P2)

Table C.4 Rules generated for "Usability"

LAG1(A)(A4) AND LAG2(B)(B6) AND V(V2) => I(P6)
LAG1(A)(A6) AND LAG2(B)(B4) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B6) AND V(V1) => I(P5)
LAG1(A)(A5) AND LAG2(B)(B4) AND V(V3) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B5) AND V(V2) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B4) AND V(V1) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) AND V(V3) => I(P5)
LAG1(A)(A5) AND LAG2(B)(B3) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B5) AND V(V1) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND V(V3) => I(P3)
LAG1(A)(A4) AND V(V2) AND S(S2) => I(P6)
LAG1(A)(A6) AND V(V4) AND S(S3) => I(P4)
LAG1(A)(A4) AND V(V1) AND S(S6) => I(P5)
LAG1(A)(A5) AND V(V3) AND S(S2) => I(P4)
LAG1(A)(A4) AND V(V2) AND S(S6) => I(P2)
LAG1(A)(A2) AND V(V1) AND S(S3) => I(P3)
LAG1(A)(A3) AND V(V3) AND S(S3) => I(P5)
LAG1(A)(A5) AND V(V3) AND S(S6) => I(P2)
LAG1(A)(A2) AND V(V1) AND S(S4) => I(P2)
LAG1(A)(A2) AND V(V3) AND S(S2) => I(P3)
LAG2(B)(B6) AND LAG2C(C5) => I(P6)
LAG2(B)(B4) AND LAG2C(C6) => I(P4)
LAG2(B)(B6) AND LAG2C(C4) => I(P5)
LAG2(B)(B5) AND LAG2C(C4) => I(P2)
LAG2(B)(B4) AND LAG2C(C5) => I(P3)
LAG2(B)(B2) AND LAG2C(C4) => I(P5)
LAG2(B)(B3) AND LAG2C(C2) => I(P2)
LAG2(B)(B5) AND LAG2C(C3) => I(P2)
LAG2(B)(B2) AND LAG2C(C5) => I(P3)
LAG2(B)(B6) AND S(S2) => I(P6)
LAG2(B)(B4) AND S(S3) => I(P4) OR I(P3)
LAG2(B)(B6) AND S(S6) => I(P5)
LAG2(B)(B4) AND S(S2) => I(P4)
LAG2(B)(B5) AND S(S6) => I(P2)
LAG2(B)(B2) AND S(S3) => I(P5)
LAG2(B)(B3) AND S(S6) => I(P2)
LAG2(B)(B5) AND S(S4) => I(P2)
LAG2(B)(B2) AND S(S2) => I(P3)

Table C.5 Rules generated for “Noise”

LAG1(A)(A4) => I(P4) OR I(P3)
LAG1(A)(A3) => I(P2)
LAG1(A)(A2) => I(P2)
LAG2(B)(B4) AND LAG2C(C3) => I(P4)
LAG2(B)(B4) AND LAG2C(C4) => I(P3) OR I(P2)
LAG2(B)(B3) AND LAG2C(C4) => I(P2)
LAG2(B)(B2) AND LAG2C(C3) => I(P2)
LAG2(B)(B2) AND LAG2C(C2) => I(P2)
LAG2(B)(B4) AND V(V2) => I(P4)
LAG2(B)(B4) AND V(V3) => I(P3)
LAG2(B)(B4) AND V(V1) => I(P2)
LAG2(B)(B3) AND V(V1) => I(P2)
LAG2(B)(B2) AND V(V1) => I(P2)
LAG2(B)(B2) AND V(V3) => I(P2)
LAG2(B)(B2) AND V(V2) => I(P2)
LAG2(B)(B4) AND S(S1) => I(P4) OR I(P3)
LAG2(B)(B4) AND S(S3) => I(P2)
LAG2(B)(B3) AND S(S1) => I(P2)
LAG2(B)(B2) AND S(S3) => I(P2)
LAG2(B)(B2) AND S(S2) => I(P2)
LAG2(B)(B2) AND S(S1) => I(P2)
LAG2C(C3) AND V(V2) => I(P4)
LAG2C(C4) AND V(V3) => I(P3)
LAG2C(C4) AND V(V1) => I(P2)
LAG2C(C3) AND V(V1) => I(P2)
LAG2C(C2) AND V(V3) => I(P2)
LAG2C(C2) AND V(V1) => I(P2)
LAG2C(C2) AND V(V2) => I(P2)
LAG2C(C3) AND S(S1) => I(P4)
LAG2C(C4) AND S(S1) => I(P3) OR I(P2)
LAG2C(C4) AND S(S3) => I(P2)
LAG2C(C3) AND S(S3) => I(P2)
LAG2C(C2) AND S(S2) => I(P2)
LAG2C(C2) AND S(S3) => I(P2)
LAG2C(C2) AND S(S1) => I(P2)
LAG1(A)(A4) AND LAG2C(C3) => I(P4)
LAG1(A)(A4) AND LAG2C(C4) => I(P3)

Table C.0.6. Rules generated for “Price”

LAG1(A)(A4) AND LAG2C(C3) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2C(C3) AND V(V3) => I(P4)
LAG1(A)(A4) AND LAG2C(C4) AND V(V3) => I(P4)
LAG1(A)(A4) AND LAG2C(C4) AND V(V2) => I(P3)
LAG1(A)(A3) AND LAG2C(C4) AND V(V1) => I(P1)
LAG1(A)(A1) AND LAG2C(C4) AND V(V2) => I(P3)
LAG1(A)(A3) AND LAG2C(C3) AND V(V4) => I(P1)
LAG1(A)(A1) AND LAG2C(C1) AND V(V1) => I(P1)
LAG1(A)(A1) AND LAG2C(C3) AND V(V2) => I(P2)
LAG1(A)(A2) AND LAG2C(C1) AND V(V3) => I(P1)
LAG1(A)(A4) AND S(S1) => I(P4)
LAG1(A)(A4) AND S(S4) => I(P4)
LAG1(A)(A4) AND S(S2) => I(P4)
LAG1(A)(A4) AND S(S3) => I(P3)
LAG1(A)(A3) AND S(S2) => I(P1)
LAG1(A)(A1) AND S(S3) => I(P3)
LAG1(A)(A3) AND S(S4) => I(P1)
LAG1(A)(A1) AND S(S4) => I(P1)
LAG1(A)(A1) AND S(S1) => I(P2)
LAG1(A)(A2) AND S(S4) => I(P1)
LAG2(B)(B3) AND V(V4) => I(P4)
LAG2(B)(B4) AND V(V3) => I(P4)
LAG2(B)(B4) AND V(V2) => I(P3)
LAG2(B)(B4) AND V(V1) => I(P1)
LAG2(B)(B3) AND V(V2) => I(P3)
LAG2(B)(B1) AND V(V4) => I(P1)
LAG2(B)(B3) AND V(V1) => I(P1)
LAG2(B)(B1) AND V(V2) => I(P2)
LAG2(B)(B1) AND V(V3) => I(P1)
V(V4) AND S(S1) => I(P4)
V(V3) AND S(S4) => I(P4) OR I(P1)
V(V3) AND S(S2) => I(P4)
V(V2) AND S(S3) => I(P3)
V(V1) AND S(S2) => I(P1)
V(V4) AND S(S4) => I(P1)
V(V1) AND S(S4) => I(P1)
V(V2) AND S(S1) => I(P2)
LAG1(A)(A4) AND V(V4) => I(P4)

Table C.7. Rules generated for "Easy to use"

LAG1(A)(A2) AND V(V3) => I(P4) OR I(P2)
LAG1(A)(A4) AND V(V5) => I(P4)
LAG1(A)(A4) AND V(V4) => I(P3)
LAG1(A)(A3) AND V(V3) => I(P2)
LAG1(A)(A2) AND V(V4) => I(P2)
LAG1(A)(A2) AND S(S1) => I(P4)
LAG1(A)(A4) AND S(S3) => I(P4)
LAG1(A)(A4) AND S(S5) => I(P3)
LAG1(A)(A3) AND S(S3) => I(P2)
LAG1(A)(A2) AND S(S3) => I(P2)
LAG1(A)(A2) AND S(S4) => I(P2)
LAG1(A)(A2) AND S(S5) => I(P2)
LAG2(B)(B3) AND V(V3) => I(P4)
LAG2(B)(B2) AND V(V5) => I(P4)
LAG2(B)(B4) AND V(V4) => I(P3)
LAG2(B)(B4) AND V(V3) => I(P2)
LAG2(B)(B3) AND V(V4) => I(P2)
LAG2(B)(B2) AND V(V4) => I(P2)
LAG2(B)(B2) AND V(V3) => I(P2)
LAG2(B)(B3) AND S(S1) => I(P4)
LAG2(B)(B2) AND S(S3) => I(P4) OR I(P2)
LAG2(B)(B4) AND S(S5) => I(P3)
LAG2(B)(B4) AND S(S3) => I(P2)
LAG2(B)(B3) AND S(S3) => I(P2)
LAG2(B)(B2) AND S(S4) => I(P2)
LAG2(B)(B2) AND S(S5) => I(P2)
LAG2C(C4) AND S(S1) => I(P4)
LAG2C(C3) AND S(S3) => I(P4)
LAG2C(C2) AND S(S5) => I(P3) OR I(P2)
LAG2C(C4) AND S(S3) => I(P2)
LAG2C(C3) AND S(S4) => I(P2)
LAG2C(C2) AND S(S4) => I(P2)
LAG2C(C2) AND S(S3) => I(P2)
V(V3) AND S(S1) => I(P4)

Table C.8. Rules generated for “Power”

LAG1(A)(A2) AND LAG2(B)(B5) AND LAG2C(C3) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) AND LAG2C(C5) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) AND LAG2C(C2) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND LAG2C(C3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND LAG2C(C2) => I(P2) OR I(P1)
LAG1(A)(A1) AND LAG2(B)(B2) AND LAG2C(C2) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) AND LAG2C(C2) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) AND LAG2C(C1) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B1) AND LAG2C(C1) => I(P1)
LAG1(A)(A2) AND LAG2C(C3) AND S(S4) => I(P3)
LAG1(A)(A3) AND LAG2C(C5) AND S(S7) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND S(S6) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND S(S4) => I(P2) OR I(P1)
LAG1(A)(A1) AND LAG2C(C2) AND S(S4) => I(P1)
LAG1(A)(A1) AND LAG2C(C1) AND S(S4) => I(P2)
LAG1(A)(A2) AND LAG2C(C1) AND S(S4) => I(P1)
LAG1(A)(A2) AND LAG2(B)(B5) AND S(S4) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) AND S(S7) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) AND S(S6) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND S(S4) => I(P2) OR I(P1)
LAG1(A)(A1) AND LAG2(B)(B2) AND S(S4) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) AND S(S4) => I(P1) OR I(P2)
LAG1(A)(A2) AND LAG2(B)(B1) AND S(S4) => I(P1)
LAG1(A)(A2) AND V(V1) AND S(S4) => I(P3)
LAG1(A)(A3) AND V(V4) AND S(S7) => I(P2)
LAG1(A)(A2) AND V(V3) AND S(S6) => I(P2)
LAG1(A)(A2) AND V(V3) AND S(S3) => I(P2)
LAG1(A)(A2) AND V(V3) AND S(S4) => I(P2) OR I(P1)
LAG1(A)(A1) AND V(V3) AND S(S4) => I(P1)
LAG1(A)(A1) AND V(V5) AND S(S4) => I(P1) OR I(P2)
LAG1(A)(A2) AND V(V5) AND S(S4) => I(P1)
LAG1(A)(A2) AND LAG2C(C3) AND V(V1) => I(P3)
LAG1(A)(A3) AND LAG2C(C5) AND V(V4) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND V(V3) => I(P2) OR I(P1)

Table C.9. Rules generated for “Comfortable to hold”

LAG1(A)(A3) AND LAG2(B)(B2) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B3) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B4) AND V(V1) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B3) AND V(V2) => I(P2)
LAG1(A)(A3) AND LAG2(B)(B2) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B3) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) AND V(V1) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND V(V2) => I(P2)
LAG2(B)(B3) AND S(S1) => I(P3)
LAG2(B)(B2) AND S(S4) => I(P4) OR I(P3)
LAG2(B)(B3) AND S(S5) => I(P3)
LAG2(B)(B4) AND S(S3) => I(P3)
LAG2(B)(B3) AND S(S3) => I(P2)
LAG2(B)(B3) AND S(S4) => I(P3)
LAG2(B)(B2) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2C(C4) AND S(S1) => I(P3)
LAG1(A)(A3) AND LAG2C(C3) AND S(S4) => I(P4) OR I(P3)
LAG1(A)(A4) AND LAG2C(C2) AND S(S5) => I(P3)
LAG1(A)(A3) AND LAG2C(C3) AND S(S3) => I(P3)
LAG1(A)(A3) AND LAG2C(C4) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND S(S4) => I(P3)
LAG1(A)(A3) AND LAG2C(C2) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND S(S3) => I(P2)
LAG1(A)(A2) AND V(V2) AND S(S1) => I(P3)
LAG1(A)(A3) AND V(V4) AND S(S4) => I(P4)
LAG1(A)(A4) AND V(V3) AND S(S5) => I(P3)
LAG1(A)(A3) AND V(V1) AND S(S3) => I(P3)
LAG1(A)(A3) AND V(V2) AND S(S3) => I(P2)
LAG1(A)(A2) AND V(V2) AND S(S4) => I(P3)
LAG1(A)(A3) AND V(V3) AND S(S4) => I(P3)
LAG1(A)(A3) AND V(V3) AND S(S3) => I(P2)
LAG1(A)(A2) AND V(V1) AND S(S3) => I(P2)
LAG1(A)(A2) AND V(V2) AND S(S3) => I(P2)
V(V2) AND S(S1) => I(P3)
V(V4) AND S(S4) => I(P4)
V(V3) AND S(S5) => I(P3)
V(V1) AND S(S3) => I(P3) OR I(P2)
V(V2) AND S(S3) => I(P2)

Table C.10. Rules generated for “Durability”

LAG2(B)(B3) AND LAG2C(C3) AND V(V3) => I(P3)
LAG2(B)(B3) AND LAG2C(C3) AND V(V4) => I(P4)
LAG2(B)(B4) AND LAG2C(C3) AND V(V3) => I(P3)
LAG2(B)(B4) AND LAG2C(C4) AND V(V3) => I(P2)
LAG2(B)(B3) AND LAG2C(C4) AND V(V3) => I(P2)
LAG2(B)(B2) AND LAG2C(C3) AND V(V4) => I(P3)
LAG2(B)(B2) AND LAG2C(C2) AND V(V4) => I(P2)
LAG2(B)(B3) AND LAG2C(C2) AND V(V2) => I(P1)
LAG2(B)(B2) AND LAG2C(C3) AND V(V3) => I(P2)
LAG2(B)(B1) AND LAG2C(C2) AND V(V4) => I(P2)
LAG2(B)(B2) AND LAG2C(C1) AND V(V4) => I(P1)
LAG2(B)(B3) AND V(V3) AND S(S3) => I(P3)
LAG2(B)(B3) AND V(V4) AND S(S3) => I(P4)
LAG2(B)(B3) AND V(V4) AND S(S6) => I(P4)
LAG2(B)(B4) AND V(V3) AND S(S5) => I(P3)
LAG2(B)(B4) AND V(V3) AND S(S6) => I(P2)
LAG2(B)(B3) AND V(V3) AND S(S4) => I(P2)
LAG2(B)(B2) AND V(V4) AND S(S5) => I(P3)
LAG2(B)(B2) AND V(V4) AND S(S4) => I(P2)
LAG2(B)(B3) AND V(V2) AND S(S5) => I(P1)
LAG2(B)(B2) AND V(V3) AND S(S3) => I(P2)
LAG2(B)(B1) AND V(V4) AND S(S5) => I(P2)
LAG2(B)(B2) AND V(V4) AND S(S6) => I(P1)
LAG1(A)(A3) AND LAG2C(C3) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2C(C3) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2C(C3) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2C(C3) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2C(C4) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2C(C4) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND V(V4) => I(P3)
LAG1(A)(A3) AND LAG2C(C2) AND V(V4) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND V(V2) => I(P1)
LAG1(A)(A1) AND LAG2C(C3) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND V(V4) => I(P2)
LAG1(A)(A2) AND LAG2C(C1) AND V(V4) => I(P1)
LAG2C(C3) AND V(V3) AND S(S3) => I(P3) OR I(P2)
LAG2C(C3) AND V(V4) AND S(S3) => I(P4)

Table C.11. Rules generated for "Portability"

LAG1(A)(A2) AND LAG2(B)(B3) AND V(V5) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B2) AND V(V3) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B4) AND V(V1) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B4) AND V(V3) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B2) AND V(V3) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) AND V(V4) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B1) AND V(V5) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B3) AND V(V3) => I(P3) OR I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND V(V4) => I(P3)
LAG2(B)(B3) AND LAG2C(C3) AND V(V5) => I(P4)
LAG2(B)(B2) AND LAG2C(C3) AND V(V3) => I(P4)
LAG2(B)(B4) AND LAG2C(C2) AND V(V1) => I(P2)
LAG2(B)(B4) AND LAG2C(C4) AND V(V3) => I(P1)
LAG2(B)(B2) AND LAG2C(C4) AND V(V3) => I(P1)
LAG2(B)(B1) AND LAG2C(C2) AND V(V4) => I(P2)
LAG2(B)(B1) AND LAG2C(C1) AND V(V5) => I(P3)
LAG2(B)(B2) AND LAG2C(C1) AND V(V3) => I(P3)
LAG2(B)(B3) AND LAG2C(C2) AND V(V3) => I(P3)
LAG2(B)(B3) AND LAG2C(C3) AND V(V3) => I(P2)
LAG2(B)(B2) AND LAG2C(C3) AND V(V4) => I(P3)
LAG1(A)(A2) AND LAG2(B)(B3) AND S(S2) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B2) AND S(S4) => I(P4)
LAG1(A)(A4) AND LAG2(B)(B4) AND S(S5) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B4) AND S(S3) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B2) AND S(S3) => I(P1)
LAG1(A)(A1) AND LAG2(B)(B1) AND S(S5) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B1) AND S(S4) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B2) AND S(S4) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B3) AND S(S5) => I(P3)
LAG1(A)(A3) AND LAG2(B)(B3) AND S(S3) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B3) AND S(S4) => I(P2)
LAG1(A)(A2) AND LAG2(B)(B2) AND S(S4) => I(P3)
LAG1(A)(A2) AND LAG2C(C3) AND V(V5) => I(P4)
LAG1(A)(A4) AND LAG2C(C3) AND V(V3) => I(P4)
LAG1(A)(A4) AND LAG2C(C2) AND V(V1) => I(P2)

Table C.12. Rules generated for “Usability”

LAG2(B)(B3) AND LAG2C(C3) AND V(V3) => I(P3)
LAG2(B)(B3) AND LAG2C(C3) AND V(V4) => I(P4)
LAG2(B)(B4) AND LAG2C(C3) AND V(V3) => I(P3)
LAG2(B)(B4) AND LAG2C(C4) AND V(V3) => I(P2)
LAG2(B)(B3) AND LAG2C(C4) AND V(V3) => I(P2)
LAG2(B)(B2) AND LAG2C(C3) AND V(V4) => I(P3)
LAG2(B)(B2) AND LAG2C(C2) AND V(V4) => I(P2)
LAG2(B)(B3) AND LAG2C(C2) AND V(V2) => I(P1)
LAG2(B)(B2) AND LAG2C(C3) AND V(V3) => I(P2)
LAG2(B)(B1) AND LAG2C(C2) AND V(V4) => I(P2)
LAG2(B)(B2) AND LAG2C(C1) AND V(V4) => I(P1)
LAG2(B)(B3) AND V(V3) AND S(S3) => I(P3)
LAG2(B)(B3) AND V(V4) AND S(S3) => I(P4)
LAG2(B)(B3) AND V(V4) AND S(S6) => I(P4)
LAG2(B)(B4) AND V(V3) AND S(S5) => I(P3)
LAG2(B)(B4) AND V(V3) AND S(S6) => I(P2)
LAG2(B)(B3) AND V(V3) AND S(S4) => I(P2)
LAG2(B)(B2) AND V(V4) AND S(S5) => I(P3)
LAG2(B)(B2) AND V(V4) AND S(S4) => I(P2)
LAG2(B)(B3) AND V(V2) AND S(S5) => I(P1)
LAG2(B)(B2) AND V(V3) AND S(S3) => I(P2)
LAG2(B)(B1) AND V(V4) AND S(S5) => I(P2)
LAG2(B)(B2) AND V(V4) AND S(S6) => I(P1)
LAG1(A)(A3) AND LAG2C(C3) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2C(C3) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2C(C3) AND V(V4) => I(P4)
LAG1(A)(A4) AND LAG2C(C3) AND V(V3) => I(P3)
LAG1(A)(A3) AND LAG2C(C4) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2C(C4) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2C(C3) AND V(V4) => I(P3)
LAG1(A)(A3) AND LAG2C(C2) AND V(V4) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND V(V2) => I(P1)
LAG1(A)(A1) AND LAG2C(C3) AND V(V3) => I(P2)
LAG1(A)(A2) AND LAG2C(C2) AND V(V4) => I(P2)
LAG1(A)(A2) AND LAG2C(C1) AND V(V4) => I(P1)
LAG2C(C3) AND V(V3) AND S(S3) => I(P3) OR I(P2)
LAG2C(C3) AND V(V4) AND S(S3) => I(P4)

References

Afzal, W., & Torkar, R. (2011). On the application of genetic programming for software engineering predictive modeling: A systematic review. *Expert Systems with Applications*, *38*(9), 11984–11997. https://doi.org/10.1016/J.ESWA.2011.03.041

Aggarwal, C. C., & Zhai, C. (2012). A Survey of Text Classification Algorithms. In *Mining Text Data* (pp. 163–222). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-3223-4_6

Akao, Y., & Mazur, G. H. (2003, February 1). The leading edge in QFD: Past, present and future. *International Journal of Quality & Reliability Management*. MCB UP Ltd. https://doi.org/10.1108/02656710310453791

Akbari-Alashti, H., Bozorg Haddad, O., Fallah-Mehdipour, E., & Mariño, M. A. (2014). Multi-reservoir real-time operation rules: a new genetic programming approach. *Proceedings of the Institution of Civil Engineers - Water Management*, *167*(10), 561–576. https://doi.org/10.1680/wama.13.00021

Alfonso, L. (2013). A Technical Note on the Use of Choquet Integral to Analyze Consumer Preferences: Application to Meat Consumption. *Journal of Sensory Studies*, *28*(6), 467–473. https://doi.org/10.1111/joss.12069

Anderson, E. W., Fornell, C., & Mazvancheryl, S. K. (2004). Customer

Satisfaction and Shareholder Value. *Journal of Marketing*, *68*(4), 172–185. https://doi.org/10.1509/jmkg.68.4.172.42723

Ankit, & Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. *Procedia Computer Science*, *132*, 937–946. https://doi.org/10.1016/J.PROCS.2018.05.109

Araujo, L. (2006). Multiobjective Genetic Programming for Natural Language Parsing and Tagging (pp. 433–442). Springer, Berlin, Heidelberg. https://doi.org/10.1007/11844297_44

Atsalakis, G. (2014). New technology product demand forecasting using a fuzzy inference system. *Operational Research*, *14*(2), 225–236. https://doi.org/10.1007/s12351-014-0160-y

Au, K. F., Choi, T. M., & Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*, *114*(2), 615–630. https://doi.org/10.1016/j.ijpe.2007.06.013

Aydoğan, S., Günay, E. E., Akay, D., & Okudan Kremer, G. E. (2020). Concept design evaluation by using Z-axiomatic design. *Computers in Industry*, *122*, 103278. https://doi.org/10.1016/j.compind.2020.103278

Barone, S., Lombardo, A., & Tarantino, P. (2007). A weighted logistic regression for conjoint analysis and kansei engineering. In *Quality and Reliability Engineering International* (Vol. 23, pp. 689–706).

https://doi.org/10.1002/qre.866

Basfirinci, C., & Mitra, A. (2015). A cross cultural investigation of airlines service quality through integration of Servqual and the Kano model. *Journal of Air Transport Management*, *42*, 239–248. https://doi.org/10.1016/J.JAIRTRAMAN.2014.11.005

Bass, F. M. (1969). A new product growth model for consumer durables. Mangeniient Sci. 15 215-227.. 1980. The relationship between diffusion rates, experience curves, and demand elasticities for consumer durable technological innovations. *J. Bus*.

Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass Model Fits without Decision Variables. *Marketing Science*, *13*(3), 203–223. https://doi.org/10.1287/mksc.13.3.203

Battistoni, E., Fronzetti Colladon, A., Scarabotti, L., & Schiraldi, M. M. (2013). Analytic hierarchy process for new product development. *International Journal of Engineering Business Management*, *5*(1), 42. https://doi.org/10.5772/56816

Beliakov, G., James, S., & Wu, J. Z. (2020). Learning Fuzzy Measures. In *Studies in Fuzziness and Soft Computing* (Vol. 382, pp. 205–239). Springer Verlag. https://doi.org/10.1007/978-3-030-15305-2_8

Besner, M. (2019). Axiomatizations of the proportional Shapley value. *Theory* *and Decision*, *86*(2), 161–183. https://doi.org/10.1007/s11238-019-09687-7

Bi, J. W., Liu, Y., Fan, Z. P., & Cambria, E. (2019). Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *International Journal of Production Research*, *57*(22), 1–21. https://doi.org/10.1080/00207543.2019.1574989

Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google Trends data - An analysis of the effects on tourism demand forecasting in Germany. *Tourism Management*, *75*, 1–12. https://doi.org/10.1016/j.tourman.2019.04.015

Busacca, B., & Padula, G. (2005). Understanding the relationship between attribute performance and overall satisfaction Theory, measurement and implications. *Marketing Intelligence and Planning*. https://doi.org/10.1108/02634500510624110

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. https://doi.org/10.1109/MIS.2013.30

Carriere-Swallow, Y., & Labbe, F. (2013). Nowcasting with Google Trends in an Emerging Market. *JOURNAL OF FORECASTING*, *32*(4), 289–298. https://doi.org/10.1002/for.1252

Chan, K. Y., Kwong, C. K., & Dillon, T. S. (2012). Development of Product

Design Models Using Fuzzy Regression Based Genetic Programming (pp. 111–128). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-27476-3_6

Chan, K. Y., Kwong, C. K., & Wong, T. C. (2011). Modelling customer satisfaction for product development using genetic programming. *Journal of Engineering Design*, *22*(1), 55–68. https://doi.org/10.1080/09544820902911374

Chan, L. K., Kao, H. P., Ng, A., & Wu, M. L. (1999). Rating the importance of customer needs in quality function deployment by fuzzy and entropy methods. *International Journal of Production Research*, *37*(11), 2499–2518. https://doi.org/10.1080/002075499190635

Chang, C. C., Chen, P.-L., Chiu, F.-R., & Chen, Y.-K. (2009). Application of neural networks and Kano's method to content recommendation in web personalization. *Expert Systems with Applications*, *36*(3), 5310–5316. https://doi.org/10.1016/J.ESWA.2008.06.067

Chen, C.-C., & Chuang, M.-C. (2008). Integrating the Kano model into a robust design approach to enhance customer satisfaction with product design. *International Journal of Production Economics*, *114*(2), 667–681. https://doi.org/10.1016/J.IJPE.2008.02.015

Chen, J., Zeng, Z., Jiang, P., & Tang, H. (2016). Application of multi-gene genetic programming based on separable functional network for landslide displacement prediction. *Neural Computing and Applications*, *27*(6), 1771–1784. https://doi.org/10.1007/s00521-015-1976-y

Chen, L.-H., & Chen, C.-N. (2014). Normalisation models for prioritising design requirements for quality function deployment processes. *International Journal of Production Research*, *52*(2), 299–313. https://doi.org/10.1080/00207543.2013.812813

Chen, L. F. (2012). A novel approach to regression analysis for the classification of quality attributes in the Kano model: An empirical test in the food and beverage industry. *Omega*, *40*(5), 651–659. https://doi.org/10.1016/j.omega.2011.12.004

Chen, L. S., Liu, C. H., Hsu, C. C., & Lin, C. Sen. (2010). C-kano model: A novel approach for discovering attractive quality elements. *Total Quality Management and Business Excellence*, *21*(11), 1189–1214. https://doi.org/10.1080/14783363.2010.529347

Chen, W., Hoyle, C., Wassenaar, H. J., Chen, W., Hoyle, C., & Wassenaar, H. J. (2013). Fundamentals of Analytical Techniques for Modeling Consumer Preferences and Choices. In *Decision-Based Design* (pp. 35–77). Springer London. https://doi.org/10.1007/978-1-4471-4036-8_3

Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., & Ma, Z. (2018, April 21). A novel feature extraction methodology for sentiment analysis of product reviews. *Neural Computing and Applications*, pp. 1–18. https://doi.org/10.1007/s00521-018-3477-2

Chen, Y., Fung, R. Y. K., & Tang, J. (2005). Fuzzy expected value modelling approach for determining target values of engineering characteristics in QFD. *International Journal of Production Research*, *43*(17), 3583–3604. https://doi.org/10.1080/00207540500032046

Chen, Y, Tang, J., Fung, R. Y. K., & Ren, Z. (2004). Fuzzy regression-based mathematical programming model for quality function deployment. *International Journal of Production Research*, *42*(5), 1009–1027. https://doi.org/10.1080/00207540310001619623

Chen, Yizeng, Fung, R. Y. K., & Tang, J. (2006). Rating technical attributes in fuzzy QFD by integrating fuzzy weighted average method and fuzzy expected value operator. *European Journal of Operational Research*, *174*(3), 1553–1566. https://doi.org/10.1016/j.ejor.2004.12.026

Cherif, M. S., Chabchoub, H., & Aouni, B. (2010). Integrating customer's preferences in the QFD planning process using a combined benchmarking and imprecise goal programming model. *International Transactions in Operational Research*, *17*(1), 85–102. https://doi.org/10.1111/j.1475-3995.2009.00718.x

Cherkassky, V. (2011). Fuzzy Inference Systems: A Critical Review. In *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications* (pp. 177–197). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-58930-0_10

Chien, C. F., Chen, Y. J., & Peng, J. T. (2010). Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and product life cycle. *International Journal of Production Economics*, *128*(2), 496–509. https://doi.org/10.1016/j.ijpe.2010.07.022

Childs, P. R. N., & Childs, P. R. N. (2014). Specification. *Mechanical Design Engineering Handbook*, 25–49. https://doi.org/10.1016/B978-0-08-097759-1.00002-2

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *ECONOMIC RECORD*, *88*(1, SI), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Chong, Y. T., & Chen, C. H. (2010). Customer needs as moving targets of product development: A review. *International Journal of Advanced Manufacturing Technology*, *48*(1–4), 395–406. https://doi.org/10.1007/s00170-009-2282-6

Chow, C. K. W. (2015). On-time performance, passenger expectations and satisfaction in the Chinese airline industry. *Journal of Air Transport*

Management, *47*, 39–47. https://doi.org/10.1016/J.JAIRTRAMAN.2015.04.003

Das, S., Singh, B., Kushwah, S., & Johri, P. (2016). Opinion based on Polarity and Clustering for Product Feature Extraction. *International Journal of Information Engineering and Electronic Business*, *8*(5), 36–43. https://doi.org/10.5815/ijieeb.2016.05.05

Dawson, D., & Askin, R. G. (1999). Optimal new product design using quality function deployment with empirical value functions. *Quality and Reliability Engineering International*, *15*(1), 17–32. https://doi.org/10.1002/(SICI)1099-1638(199901/02)15:1<17::AID-QRE203>3.0.CO;2-J

Delice, E. K., & Güngör, Z. (2011). Asi mixed integer goal programming model for discrete values of degn requirements in QFD. *International Journal of Production Research*, *49*(10), 2941–2957. https://doi.org/10.1080/00207541003720343

Ding, X., & Liu, B. (2007). The utility of linguistic rules in opinion mining. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 07*, 811. https://doi.org/10.1145/1277741.1277921

Dosoula, N., Griep, R., Ridder, R. Den, Slangen, R., Schouten, K., & Frasincar,

F. (2016). Detection of multiple implicit features per sentence in consumer review data. In *Communications in Computer and Information Science* (Vol. 615, pp. 289–303). Springer, Cham. https://doi.org/10.1007/978-3-319-40180-5_20

Dou, R., Zhang, Y., & Nan, G. Application of combined Kano model and interactive genetic algorithm for product customization, 29 Journal of Intelligent Manufacturing § (2016). https://doi.org/10.1007/s10845-016-1280-4

Eggers, F., & Eggers, F. (2011). Where have all the flowers gone? Forecasting green trends in the automobile industry with a choice-based conjoint adoption model. *Technological Forecasting and Social Change*, *78*(1), 51–62. https://doi.org/10.1016/j.techfore.2010.06.014

El Zarwi, F., Vij, A., & Walker, J. L. (2017). A discrete choice framework for modeling and forecasting the adoption and diffusion of new transportation services. *Transportation Research Part C: Emerging Technologies*, *79*, 207–223. https://doi.org/10.1016/j.trc.2017.03.004

Erzurum Cicek, Z. I., & Kamisli Ozturk, Z. (2021). Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting. *Applied Soft Computing*, *102*, 107091. https://doi.org/10.1016/j.asoc.2021.107091

Ettredge, M., Gerdes, J., & Karuga, G. (2005, November 1). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*. ACM PUB27 New York, NY, USA. https://doi.org/10.1145/1096000.1096010

Fan, Z.-P. P. Z.-P. Z.-P. P., Che, Y.-J. J., & Chen, Z.-Y. Y. Z.-Y. Z.-Y. Y. (2017). Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *Journal of Business Research*, *74*(1, SI), 90–100. https://doi.org/10.1016/j.jbusres.2017.01.010

Fang, X., Shen, Y., Zhou, J., Pantelous, A. A., & Zhao, M. (2020). QFD-Based Product Design for Multisegment Markets: A Fuzzy Chance-Constrained Programming Approach. *IEEE Transactions on Engineering Management*, 1–15. https://doi.org/10.1109/tem.2020.3009163

Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, *170*, 97–135. https://doi.org/10.1016/j.ijpe.2015.09.010

Feng, J., Cai, S., & Ma, X. (2018, January 12). Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm. *Cluster Computing*, pp. 1–19. https://doi.org/10.1007/s10586-017-1626-5

Feng, Y., Li, G., Sun, X., & Li, J. (2019). Forecasting the number of inbound tourists with Google Trends. In *Procedia Computer Science* (Vol. 162, pp. 628–633). Elsevier B.V. https://doi.org/10.1016/j.procs.2019.12.032

Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & Javier González-Castaño, F. (2016). *Unsupervised method for sentiment analysis in online texts*. *Expert Systems with Applications* (Vol. 58). Pergamon. https://doi.org/10.1016/j.eswa.2016.03.031

Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian Ensemble Learning. *Decision Support Systems*, *68*, 26–38. https://doi.org/10.1016/j.dss.2014.10.004

Fung, R. Y. K., Tu, Y., Tang, J., Wang, D., Tu, Y., & Wang, D. (2002). Product design resources optimization using a non-linear fuzzy quality function deployment model. *International Journal of Production Research*, *40*(3), 585–599. https://doi.org/10.1080/00207540110061634

Galetto, M., Franceschini, F., Maisano, D. A., & Mastrogiacomo, L. (2018). Engineering characteristics prioritisation in QFD using ordinal scales: A robustness analysis. *European Journal of Industrial Engineering*, *12*(2), 151–174. https://doi.org/10.1504/EJIE.2018.090617

Gandomi, A. H., & Alavi, A. H. (2012). A new multi-gene genetic programming approach to nonlinear system modeling. Part I: Materials and structural engineering problems. *Neural Computing and Applications*, *21*(1), 171–187. https://doi.org/10.1007/s00521-011-0734-z

Garg, A. A., Garg, A. A., Tai, K., Barontini, S., & Stokes, A. (2014). A Computational Intelligence-Based Genetic Programming Approach for the Simulation of Soil Water Retention Curves. *Transport in Porous Media*, *103*(3), 497–513. https://doi.org/10.1007/s11242-014-0313-8

Garg, A., & Lam, J. S. L. (2015). Improving environmental sustainability by formulation of generalized power consumption models using an ensemble based multi-gene genetic programming approach. *Journal of Cleaner Production*, *102*, 246–263. https://doi.org/10.1016/J.JCLEPRO.2015.04.068

Garg, A., Vijayaraghavan, V., Wong, C. H., Tai, K., Sumithra, K., Gao, L., & Singru, P. M. (2014). Combined CI-MD approach in formulation of engineering moduli of single layer graphene sheet. *Simulation Modelling Practice and Theory*, *48*, 93–111. https://doi.org/10.1016/j.simpat.2014.07.008

Garg, A, Tai, K., & Savalani, M. M. (2014). State-of-the-art in empirical modelling of rapid prototyping processes. *Rapid Prototyping Journal*, *20*(2), 164–178. https://doi.org/10.1108/RPJ-08-2012-0072

Garg, Akhil, Bhalerao, Y., & Tai, K. (2013). Review of empirical modelling techniques for modelling of turning process. *International Journal of Modelling, Identification and Control*, *20*(2), 121. https://doi.org/10.1504/IJMIC.2013.056184

Garg, Ankit, Garg, A., & Lam, J. S. L. (2015). Evolving Functional Expression of Permeability of Fly Ash by a New Evolutionary Approach. *Transport in Porous Media*, *107*(2), 555–571. https://doi.org/10.1007/s11242-015-0454-4

Geng, X., Chu, X., & Zhang, Z. (2010). A new integrated design concept evaluation approach based on vague sets. *Expert Systems with Applications*, *37*(9), 6629–6638. https://doi.org/10.1016/j.eswa.2010.03.058

González, M. E., Quesada, G., & Bahill, A. T. (2003). Improving Product Design Using Quality Function Deployment: The School Furniture Case in Developing Countries. *Quality Engineering*, *16*(1), 45–56. https://doi.org/10.1081/QEN-120020770

Google Trends. (n.d.). Retrieved March 26, 2021, from https://support.google.com/trends/answer/4365533

Govindarajan, M. (2013). Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm. *International Journal of Advanced Computer …*, *135*, 9–17.

https://doi.org/10.1016/j.knosys.2017.07.015

Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, *89*(3), 445–456. https://doi.org/10.1016/0377-2217(95)00176-X

Grigoroudis, E., & Siskos, Y. (2002). Preference disaggregation for measuring and analysing customer satisfaction: The MUSA method. *European Journal of Operational Research*, *143*(1), 148–170. https://doi.org/10.1016/S0377-2217(01)00332-0

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/J.TOURMAN.2016.09.009

Gupta, R., Lau, C. K. M., Plakandaras, V., & Wong, W. K. (2019). The role of housing sentiment in forecasting U.S. home sales growth: evidence from a Bayesian compressed vector autoregressive model. *Economic Research-Ekonomska Istrazivanja*, *32*(1), 2554–2567. https://doi.org/10.1080/1331677X.2019.1650657

Güven, İ., & Şimşir, F. (2020). Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Computers and Industrial Engineering*, *147*, 106678. https://doi.org/10.1016/j.cie.2020.106678

Hao, H., Zhang, Q., Wang, Z., & Zhang, J. (2018). Forecasting the number of end-of-life vehicles using a hybrid model based on grey model and artificial neural network. *Journal of Cleaner Production*, *202*, 684–696. https://doi.org/10.1016/j.jclepro.2018.08.176

He, L., Chen, W., Hoyle, C., & Yannou, B. (2012). Choice modeling for usage context-based design. *Journal of Mechanical Design, Transactions of the ASME*, *134*(3). https://doi.org/10.1115/1.4005860

He, W., Shen, J., Tian, X., Li, Y., Akula, V., Yan, G., & Tao, R. (2015). Gaining competitive intelligence from social media data. *Industrial Management & Data Systems*, *115*(9), 1622–1636. https://doi.org/10.1108/IMDS-03-2015-0098

Higgins, A., Paevere, P., Gardner, J., & Quezada, G. (2012). Combining choice modelling and multi-criteria analysis for technology diffusion: An application to the uptake of electric vehicles. *Technological Forecasting and Social Change*, *79*(8), 1399–1412. https://doi.org/10.1016/j.techfore.2012.04.008

Hirschberg, J., & Manning, C. D. (2015, July 17). Advances in natural language processing. *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.aaa8685

Hong, Y., Lu, J., Yao, J., Zhu, Q., & Zhou, G. (2012). What Reviews Are Satisfactory: Novel Features for Automatic Helpfulness Voting. *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 495–504. https://doi.org/10.1145/2348283.2348351

Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2019). Google Trends data for analysing tourists' online search behaviour and improving demand forecasting: the case of Åre, Sweden. *Information Technology and Tourism*, *21*(1), 45–62. https://doi.org/10.1007/s40558-018-0129-4

Hosseini, S. M., Fard, M. Y., & Baboli, S. M. (2011). A New Method For Stock Price Index Forecasting Using Fuzzy Time Series. *Australian Journal of Basic and Applied Sciences*, *5*(12), 894–898. Retrieved from https://pdfs.semanticscholar.org/ddeb/db5e4be25dcb705340d5029f057b94f034d4.pdf

Hoyle, C., Chen, W., Wang, N., & Koppelman, F. S. (2010). Integrated bayesian hierarchical choice modeling to capture heterogeneous consumer preferences in engineering design. *Journal of Mechanical Design, Transactions of the ASME*, *132*(12). https://doi.org/10.1115/1.4002972

Hu, X., & Wu, B. (2009). Classification and summarization of pros and cons for customer reviews. In *Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology -*

Workshops, WI-IAT Workshops 2009 (Vol. 3, pp. 73–76). IEEE. https://doi.org/10.1109/WI-IAT.2009.234

Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E. W. T., & Liu, M. (2015). Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review. *Applied Soft Computing*, *36*, 534–551. https://doi.org/10.1016/J.ASOC.2015.07.008

Huang, J., You, X. Y., Liu, H. C., & Si, S. L. (2019). New approach for quality function deployment based on proportional hesitant fuzzy linguistic term sets and prospect theory. *International Journal of Production Research*, *57*(5), 1283–1299. https://doi.org/10.1080/00207543.2018.1470343

Innovations, D., Everett, E., & Google, M. R. (2013). Diffusion of Innovations, 4th Edition - Everett M. Rogers - Google Books Page 1 of 1, 2013. Retrieved from https://books.google.com.hk/books?hl=en&lr=&id=v1ii4QsB7jIC&oi=fnd&pg=PR15&ots=DMTwwJRo8O&sig=f7AKpjFcTHLjGYimgpWqw6GTrng&redir_esc=y#v=onepage&q&f=false

Irhami, E. A., & Farizal, F. (2021). Forecasting the Number of Vehicles in Indonesia Using Auto Regressive Integrative Moving Average (ARIMA) Method. In *Journal of Physics: Conference Series* (Vol. 1845, p. 12024). IOP Publishing Ltd. https://doi.org/10.1088/1742-6596/1845/1/012024

Jia, W., Liu, Z., Lin, Z., Qiu, C., & Tan, J. (2016). Quantification for the importance degree of engineering characteristics with a multi-level hierarchical structure in QFD. *International Journal of Production Research*, *54*(6), 1627–1649. https://doi.org/10.1080/00207543.2015.1041574

Jiang, H., Kwong, C. K., Ip, W. H., & Chen, Z. (2013). Chaos-based fuzzy regression approach to modeling customer satisfaction for product design. *IEEE Transactions on Fuzzy Systems*, *21*(5), 926–936. https://doi.org/10.1109/TFUZZ.2012.2236841

Jiang, H., Kwong, C. K., Okudan Kremer, G. E., & Park, W. Y. (2019). Dynamic modelling of customer preferences for product design using DENFIS and opinion mining. *Advanced Engineering Informatics*, *42*, 100969. https://doi.org/10.1016/j.aei.2019.100969

Jiang, H., Kwong, C. K., & Yung, K. L. (2017). *A methodology for predicting future importance of customer needs based on online customer reviews*. *Journal of Mechanical Design* (Vol. 139). https://doi.org/10.1115/1.4037348

Jiang, Y., Huang, S., Zhang, Y., & Feng, Z. (2016). Multi-gene genetic programming based modulation classification using multinomial logistic regression. In *2016 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)* (pp. 352–357).

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep Convolution Neural Networks for Twitter Sentiment Analysis. *IEEE Access*, *6*, 23253–23260. https://doi.org/10.1109/ACCESS.2017.2776930

Jin, J., Ji, P., & Gu, R. (2016). Identifying comparative customer requirements from product online reviews for competitor analysis. *Engineering Applications of Artificial Intelligence*, *49*, 61–73. https://doi.org/10.1016/J.ENGAPPAI.2015.12.005

Jin, J., Liu, Y., Ji, P., & Liu, H. (2016). Understanding big consumer opinion data for market-driven product design. *International Journal of Production Research*, *54*(10), 3019–3041. https://doi.org/10.1080/00207543.2016.1154208

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the international conference on Web search and web data mining - WSDM '08* (p. 219). https://doi.org/10.1145/1341531.1341560

Johnson, J. (2021). • Search engine market share worldwide | Statista. Retrieved April 14, 2021, from https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Jun, D. Bin, & Park, Y. S. (1999). A Choice-Based Diffusion Model for Multiple Generations of Products. *Technological Forecasting and Social Change*, *61*(1), 45–58. https://doi.org/10.1016/S0040-1625(98)00049-3

Jun, S. P., & Park, D. H. (2017). Visualization of brand positioning based on consumer web search information: Using social network analysis. *Internet Research, 27*(2), 381–407. https://doi.org/10.1108/IntR-02-2016-0037

Jun, S. P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Kala, R. (2012). Multi-robot path planning using co-evolutionary genetic programming. *Expert Systems with Applications*, *39*(3), 3817–3831. https://doi.org/10.1016/J.ESWA.2011.09.090

Kanakaraj, M., & Guddeti, R. M. R. (2015). NLP based sentiment analysis on Twitter data using ensemble classifiers. In *2015 3rd International Conference on Signal Processing, Communication and Networking, ICSCN 2015* (pp. 1–5). IEEE. https://doi.org/10.1109/ICSCN.2015.7219856

Kang, M., Ahn, J., & Lee, K. (2018). Opinion mining using ensemble text hidden Markov models for text classification. *Expert Systems with Applications*, *94*, 218–227. https://doi.org/10.1016/j.eswa.2017.07.019

KANO, N., SERAKU, N., TAKAHASHI, F., & TSUJI, S. (1984). Attractive Quality and Must-Be Quality. *Journal of The Japanese Society for Quality Control*, *14*(2), 147–156. https://doi.org/10.20684/quality.14.2_147

Khan, A. A., Basri, S., Dominic, P. D. D., & Amin, F. E. (2013). Communication risks and best practices in global software development during requirements change management: A systematic literature review protocol. *Research Journal of Applied Sciences, Engineering and Technology, 6*(19), 3514–3519. https://doi.org/10.19026/rjaset.6.3554

Khan, F. H., Qamar, U., & Bashir, S. (2016). Multi-Objective Model Selection (MOMS)-based Semi-Supervised Framework for Sentiment Analysis. *Cognitive Computation, 8*(4), 614–628. https://doi.org/10.1007/s12559-016-9386-8

Kim, K., & Lee, J. (2014). Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. *Pattern Recognition, 47*(2), 758–768. https://doi.org/10.1016/j.patcog.2013.07.022

Kim, S.-M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing - EMNLP '06* (p. 423). https://doi.org/10.3115/1610075.1610135

Kim, T., & Yoo, T. (n.d.). Methodology for extracting the delighter in Kano model using big data analysis, *31*(5–6). https://doi.org/10.1080/14783363.2018.1442715

Kim, Y., Dwivedi, R., Zhang, J., & Jeong, S. R. (2016). Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online Information Review*, *40*(1), 42–61. https://doi.org/10.1108/OIR-03-2015-0068

Kobayakawa, T. S., Kumano, T., Tanaka, H., Okazaki, N., Kim, J.-D., & Tsujii, J. (2009). Opinion classification with tree kernel SVM using linguistic modality analysis. *Proceeding of the 18th ACM Conference on Information and Knowledge Management CIKM 09*, 1791. https://doi.org/10.1145/1645953.1646231

Kohli, A. K., & Jaworski, B. J. (1990). Market Orientation: The Construct, Research Propositions, and Managerial Implications. *Journal of Marketing*, *54*(2), 1–18. https://doi.org/10.1177/002224299005400201

Krapivin, M., Autayeu, A., Marchese, M., Blanzieri, E., & Segata, N. (2010). Keyphrases extraction from scientific documents: Improving machine learning approaches with natural language processing. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 6102 LNCS, pp. 102–111). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13654-2_12

Krutz, G. W., Iwashita, M., & Doster, D. H. (1986). Forecasting farm equipment sales in a declining market. In *SAE Technical Papers*. SAE International. https://doi.org/10.4271/861248

Kück, M., & Freitag, M. (2021). Forecasting of customer demands for production planning by local k-nearest neighbor models. *International Journal of Production Economics*, *231*, 107837. https://doi.org/10.1016/j.ijpe.2020.107837

Kuklys, W. (2002). Stated choice methods: analysis and application, Jordan J. Louviere, David A. Hensher and Joffre D. Swait, Cambridge University Press, ISBN: 0-521-78830-7. *Journal of Applied Econometrics*, *17*(6), 701–704. https://doi.org/10.1002/jae.701

Kumar, V., & Kumar, R. (2019). The demand forecasting: A comparative review of conventional and non-conventional techniques. *International Journal of Mechanical and Production Engineering Research and Development*, *9*(1), 253–262. https://doi.org/10.24247/ijmperdfeb201924

Kwong, C. K., & Bai, H. (2002). A fuzzy AHP approach to the determination of importance weights of customer requirements in quality function deployment. *Journal of Intelligent Manufacturing*, *13*(5), 367–377. https://doi.org/10.1023/A:1019984626631

Kwong, C. K., Chen, Y., Bai, H., & Chan, D. S. K. (2007). A methodology of determining aggregated importance of engineering characteristics in QFD. *Computers and Industrial Engineering*, *53*(4), 667–679. https://doi.org/10.1016/j.cie.2007.06.008

Kwong, C. K., Chen, Y., Chan, K. Y., & Luo, X. (2010). A generalised fuzzy least-squares regression approach to modelling relationships in QFD. *Journal of Engineering Design*, *21*(5), 601–613. https://doi.org/10.1080/09544820802563234

Kwong, C. K., Wong, T. C., & Chan, K. Y. (2009). A methodology of generating customer satisfaction models for new product development using a neuro-fuzzy approach. *Expert Systems with Applications*, *36*(8), 11262–11270. https://doi.org/10.1016/j.eswa.2009.02.094

Lai, X., Tan, K. C., & Xie, M. (2007). Optimizing product design using quantitative quality function deployment: A case study. In *Quality and Reliability Engineering International* (Vol. 23, pp. 45–57). https://doi.org/10.1002/qre.819

Laili, Y., Tao, F., & Zhang, L. (2015). A Hybrid RCO for Dual Scheduling of Cloud Service and Computing Resource in Private Cloud. In *Configurable Intelligent Optimization Algorithm* (pp. 257–287). Springer, Cham. https://doi.org/10.1007/978-3-319-08840-2_9

Lau, R. Y. K., Lai, C. L., Bruza, P. B., & Wong, K. F. (2011). Leveraging web 2.0 data for scalable semi-supervised learning of domain-specific sentiment lexicons. In *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11* (p. 2457). https://doi.org/10.1145/2063576.2063991

Lee, A. J. T., Yang, F. C., Chen, C. H., Wang, C. S., & Sun, C. Y. (2016). *Mining perceptual maps from consumer reviews*. *Decision Support Systems* (Vol. 82). Elsevier B.V. https://doi.org/10.1016/j.dss.2015.11.002

Lee, J., & Cho, Y. (2009). Demand forecasting of diesel passenger car considering consumer preference and government regulation in South Korea. *Transportation Research Part A: Policy and Practice*, *43*(4), 420–429. https://doi.org/10.1016/j.tra.2008.11.007

Lee, J., Cho, Y., Lee, J. D., & Lee, C. Y. (2006). Forecasting future demand for large-screen television sets using conjoint analysis with diffusion model. *Technological Forecasting and Social Change*, *73*(4), 362–376. https://doi.org/10.1016/j.techfore.2004.12.002

Lee, Y. C., & Huang, S. Y. (2009). A new fuzzy concept approach for Kano's model. *Expert Systems with Applications*, *36*(3 PART 1), 4479–4484. https://doi.org/10.1016/j.eswa.2008.05.034

Lee, Y., Park, J., & Cho, S. (2020, November 1). Extraction and prioritization of product attributes using an explainable neural network. *Pattern Analysis and Applications*. Springer. https://doi.org/10.1007/s10044-020-00878-5

Li, S.-T., & Tsai, F.-C. (2013). A fuzzy conceptualization model for text mining with application in opinion polarity classification. *Knowledge-Based Systems*, *39*, 23–33. https://doi.org/10.1016/J.KNOSYS.2012.10.005

Li, S., Tang, D., & Wang, Q. (2019). Rating engineering characteristics in open design using a probabilistic language method based on fuzzy QFD. *Computers and Industrial Engineering*, *135*, 348–358. https://doi.org/10.1016/j.cie.2019.06.008

Li, X., Li, J., & Wu, Y. (2015). A Global Optimization Approach to Multi-Polarity Sentiment Analysis. *PLOS ONE*, *10*(4), e0124672. https://doi.org/10.1371/journal.pone.0124672

Li, Z., Tian, Z. G., Wang, J. W., Wang, W. M., & Huang, G. Q. (2018). Dynamic mapping of design elements and affective responses: a machine learning based method for affective design. *Journal of Engineering Design*, *29*(7), 358–380. https://doi.org/10.1080/09544828.2018.1471671

Liang, X., Xie, L., & Yan, H. (2015). Self-Restraining Bass Models. *Journal of Forecasting*, *34*(6), 472–477. https://doi.org/10.1002/for.2346

Lim, S., & Tucker, C. S. (2016). A Bayesian Sampling Method for Product Feature Extraction From Large-Scale Textual Data. *Journal of Mechanical Design*, *138*(6), 061403. https://doi.org/10.1115/1.4033238

Lin, F.-H., Tsai, S.-B., Lee, Y.-C., Hsiao, C.-F., Zhou, J., Wang, J., & Shang, Z. (2017). Empirical research on Kano's model and customer satisfaction. *PLOS ONE*, *12*(9), e0183888. https://doi.org/10.1371/journal.pone.0183888

Liu *, S.-T. (2005). Rating design requirements in fuzzy quality function deployment via a mathematical programming approach. *International Journal of Production Research*, *43*(3), 497–513. https://doi.org/10.1080/00207540420002700395

Liu, C., Tang, L., & Shan, W. (2018). An extended HITS algorithm on bipartite network for features extraction of online customer reviews. *Sustainability (Switzerland)*, *10*(5), 1425. https://doi.org/10.3390/su10051425

Liu, H. T., & Wang, C. H. (2010). An advanced quality function deployment model using fuzzy analytic network process. *Applied Mathematical Modelling*, *34*(11), 3333–3351. https://doi.org/10.1016/j.apm.2010.02.024

Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: A review. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2013/738675

Liu, Yan, & Cirillo, C. (2018). A generalized dynamic discrete choice model for green vehicle adoption. *Transportation Research Part A: Policy and Practice*, *114*, 288–302. https://doi.org/10.1016/j.tra.2018.01.034

Liu, Yuanyuan, Zhou, J., & Chen, Y. (2014). Using fuzzy non-linear regression to identify the degree of compensation among customer requirements in QFD. *Neurocomputing*, *142*, 115–124. https://doi.org/10.1016/j.neucom.2014.01.053

Lu, C. J., Lee, T. S., & Lian, C. M. (2012). Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. *Decision Support Systems*, *54*(1), 584–596. https://doi.org/10.1016/j.dss.2012.08.006

Lu, C. J., & Wang, Y. W. (2010). Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics*, *128*(2), 603–613. https://doi.org/10.1016/j.ijpe.2010.07.004

Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., & Anzanello, M. J. (2020). Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management*, *83*, 101760. https://doi.org/10.1016/J.JAIRTRAMAN.2019.101760

Ma, J., & Kim, H. M. (2015). Product family architecture design with predictive, data-driven product family design method. *Research in Engineering Design*, *27*(1), 5–21. https://doi.org/10.1007/s00163-015-0201-4

Madzík, P. (2016). Increasing accuracy of the Kano model – a case study. *Total Quality Management & Business Excellence*, *29*(3–4), 1–23. https://doi.org/10.1080/14783363.2016.1194197

Mahendhiran, P. D., & Kannimuthu, S. (2018). Deep Learning Techniques for

Polarity Classification in Multimodal Sentiment Analysis, *17*. https://doi.org/10.1142/S0219622018500128

Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web*, *20*(2), 135–154. https://doi.org/10.1007/s11280-015-0381-x

Martinez, A. R. (2010a). *Natural language processing*. *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 2). John Wiley & Sons, Ltd. https://doi.org/10.1002/wics.76

Martinez, A. R. (2010b, May 1). Natural language processing. *Wiley Interdisciplinary Reviews: Computational Statistics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/wics.76

Maynard, J. A., Talavera, A., Forero, L., & Pacheco, M. A. C. (2018). Estimating the Geological Properties in Oil Reservoirs Through Multi-Gene Genetic Programming. In *2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings* (pp. 1–5). IEEE. https://doi.org/10.1109/CEC.2018.8477910

McAllister, M. N. (1996). *Fuzzy Logic with Engineering Applications (Timothy Ross)*. *SIAM Review* (Vol. 38). Wiley. Retrieved from http://books.google.com/books?hl=en&lr=&id=nhz1f9j6_SMC&oi=fnd&p

g=PR7&dq=ross+fuzzy&ots=vfr-0vLov5&sig=stuh2F1LrUt6ZVWre7c-76kRf6U

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*(4), 303–328. https://doi.org/10.1016/0047-2727(74)90003-6

Miller, J. F., & Mohid, M. (2013). Function optimization using cartesian genetic programming. In *Proceeding of the fifteenth annual conference companion on Genetic and evolutionary computation conference companion - GECCO '13 Companion* (p. 147). New York, New York, USA: ACM Press. https://doi.org/10.1145/2464576.2464646

Mostard, J., Teunter, R., & De Koster, R. (2011). Forecasting demand for single-period products: A case study in the apparel industry. *European Journal of Operational Research*, *211*(1), 139–147. https://doi.org/10.1016/j.ejor.2010.11.001

Nazari-Shirkouhi, S., & Keramati, A. (2017). Modeling customer satisfaction with new product design using a flexible fuzzy regression-data envelopment analysis algorithm. *Applied Mathematical Modelling*, *50*, 755–771. https://doi.org/10.1016/j.apm.2017.01.020

Netisopakul, P., & Leenawong, C. (2017). Multiple linear regression using gradient descent: A case study on Thailand car sales. *Advanced Science*

Letters, *23*(6), 5195–5198. https://doi.org/10.1166/asl.2017.7340

Neto, J. C., Silva, M. M., & Santos, S. M. (2016). A time series model for estimating the generation of lead acid battery scrap. *Clean Technologies and Environmental Policy*, *18*(6), 1931–1943. https://doi.org/10.1007/s10098-016-1121-3

Nijssen, Ed J, & Lieshout, K. F. M. (1995). Awareness, use and effectiveness of models and methods for new product development. *European Journal of Marketing*, *29*(10), 27–44. https://doi.org/10.1108/03090569510098483

Nijssen, Edwin J., & Frambach, R. T. (2000). Determinants of the adoption of new product development tools by industrial firms. *Industrial Marketing Management*, *29*(2), 121–131. https://doi.org/10.1016/S0019-8501(98)00043-1

Nosrati, M., Karimi, R., Mohammadi, M., & Malekian, K. (2013). Internet Marketing or Modern Advertising! How? Why? *International Journal of Economy, Management and Social Sciences*, *2*(3). Retrieved from www.waprogramming.com

Nuo, L., Hao, C., & Jian-qiang, H. (2019). Application of Multigene Genetic Programming for Estimating Elastic Modulus of Reservoir Rocks - IEEE Conference Publication. In *2019 Symposium on Piezoelectrcity,Acoustic Waves and Device Applications (SPAWDA)* (p. 4). Harbin, China, China:

IEEE. https://doi.org/https://doi.org/10.1109/SPAWDA.2019.8681879

Orlowska, E. (1998). *Incomplete information : rough set analysis*.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, *11*(5), 341–356. https://doi.org/10.1007/BF01001956

Peters, G. (1994). Fuzzy linear regression with fuzzy intervals. *Fuzzy Sets and Systems*, *63*(1), 45–55. https://doi.org/10.1016/0165-0114(94)90144-9

Polpinij, J., & Ghose, A. K. (2008). An Ontology-Based Sentiment Classification Methodology for Online Consumer Reviews. In *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (pp. 518–524). IEEE. https://doi.org/10.1109/WIIAT.2008.68

Prusa, J. D., & Khoshgoftaar, T. M. (2017). Improving deep neural network design with new text data representations. *Journal of Big Data*, *4*(1), 7. https://doi.org/10.1186/s40537-017-0065-8

Quan, C., & Ren, F. (2014). Unsupervised product feature extraction for feature-oriented opinion determination. *Information Sciences*, *272*, 16–28. https://doi.org/10.1016/j.ins.2014.02.063

R. Yager, R. (2016). Uncertainty modeling using fuzzy measures. *Knowledge-Based Systems*, *92*, 1–8. https://doi.org/10.1016/J.KNOSYS.2015.10.001

Raharjo, H., Brombacher, A. C., & Xie, M. (2008). Dealing with subjectivity in early product design phase: A systematic approach to exploit Quality Function Deployment potentials. *Computers and Industrial Engineering*, *55*(1), 253–278. https://doi.org/10.1016/j.cie.2007.12.012

Rahman, M. A., Sarker, B. R., & Escobar, L. A. (2011). Peak demand forecasting for a seasonal product using Bayesian approach. *Journal of the Operational Research Society*, *62*(6), 1019–1028. https://doi.org/10.1057/jors.2010.58

Rashid, M. M., Tamaki, J., Ullah, A. M. M. S., & Kubo, and A. (2011). A Kano Model Based Linguistic Application for Customer Needs Analysis. *International Journal of Engineering Business Management*, *3*(2), 29–36. Retrieved from https://doaj.org/article/b8a56e17dba745249e986335174fa944

Resende, C. B., Heckmann, C. G., & Michalek, J. J. (2011). Robust design for profit maximization under uncertainty of consumer choice model parameters using the delta method. In *Proceedings of the ASME Design Engineering Technical Conference* (Vol. 5, pp. 421–434). American Society of Mechanical Engineers Digital Collection. https://doi.org/10.1115/DETC2011-48409

Rochford, L. (1991). Generating and screening new products ideas. *Industrial Marketing Management*, *20*(4), 287–296. https://doi.org/https://doi.org/10.1016/0019-8501(91)90003-X

Sabir, S. S. (2020). Does product design stimulate customer satisfaction? Mediating role of affect. *Asia Pacific Journal of Marketing and Logistics*, *32*(6), 1255–1268. https://doi.org/10.1108/APJML-03-2019-0216

Schouten, K., & Frasincar, F. (2014). Implicit feature extraction for sentiment analysis in consumer reviews. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8455 LNCS, pp. 228–231). Springer, Cham. https://doi.org/10.1007/978-3-319-07983-7_31

Searson, D., Willis, M., & Montague, G. (2007). Co-evolution of non-linear PLS model components. *Journal of Chemometrics*, *21*(12), 592–603. https://doi.org/10.1002/cem.1084

Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Seuk Wai, P., Wai Chung, Y., … Al-Garadi, M. A. (2018). Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges. *IEEE Access*, *6*, 37807–37827. https://doi.org/10.1109/ACCESS.2018.2851311

Shiau, T.-A., & Lee, C.-S. (2017). Measuring Network-Based Public Transit Performance Using Fuzzy Measures and Fuzzy Integrals, *9*(5), 695. https://doi.org/10.3390/su9050695

Song, Q., & Chissom, B. S. (1993). Forecasting enrollments with fuzzy time series - Part I. *Fuzzy Sets and Systems*, *54*(1), 1–9.

https://doi.org/10.1016/0165-0114(93)90355-L

Song, Q., & Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series — part II. *Fuzzy Sets and Systems*, *62*(1), 1–8. https://doi.org/10.1016/0165-0114(94)90067-1

Song, Y., Lee, S., Zo, H., & Lee, H. (2015). A hybrid Bass-Markov model for the diffusion of a dual-type device-based telecommunication service: The case of WiBro service in Korea. *Computers and Industrial Engineering*, *79*, 85–94. https://doi.org/10.1016/j.cie.2014.10.020

Strachan, G. C., Koshiyama, A. S., Dias, D. M., Vellasco, M. M. B. R., & Pacheco, M. A. C. (2014). Quantum-inspired multi-gene linear genetic programming model for regression problems. In *Proceedings - 2014 Brazilian Conference on Intelligent Systems, BRACIS 2014* (pp. 152–157). IEEE. https://doi.org/10.1109/BRACIS.2014.37

Suef, M., Suparno, S., & Singgih, M. L. (2017). Categorizing product attributes efficiently in QFD-Kano: A case analysis in telecommunication. *TQM Journal*, *29*(3), 512–526. https://doi.org/10.1108/TQM-03-2015-0036

Sullivan, D. (2016). Google now handles at least 2 trillion searches per year. Retrieved February 12, 2020, from https://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information Fusion*, *36*, 10–25. https://doi.org/10.1016/j.inffus.2016.10.004

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Tanaka, H., Uejima, S., & Asai, K. (1982). Linear Regression Analysis With Fuzzy Model. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-12*(6), 903–907. https://doi.org/10.1109/tsmc.1982.4308925

Tebaldi, L., Pindari, S., & Bottani, E. (2019). Demand forecasting in an automotive company: An artificial neural network approach. In *31st European Modeling and Simulation Symposium, EMSS 2019* (pp. 162–167). https://doi.org/10.46354/i3m.2019.emss.024

Tom Coughlin. (2018). 175 Zettabytes By 2025. Retrieved July 21, 2021, from https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/?sh=447fab125459

Tuarob, S., & Tucker, C. S. (2015). Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data. *Journal of Computing and Information Science in Engineering*, *15*(3), 031003. https://doi.org/10.1115/1.4029562

Tubishat, M., Idris, N., & Abushariah, M. A. M. (2018). Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges. *Information Processing and Management*, *54*(4), 545–563. https://doi.org/10.1016/j.ipm.2018.03.008

Tucker, C. S., & Kim, H. M. (2011). Trend Mining for Predictive Product Design. *Journal of Mechanical Design*, *133*(11), 111008. https://doi.org/10.1115/1.4004987

Ulwick, A. W. (2002). Turn customer input into innovation. *Harvard Business Review*. Retrieved from www.hbr.org

Van De Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications*, *42*(11), 4999–5010. https://doi.org/10.1016/j.eswa.2015.02.007

van Steenbergen, R. M., & Mes, M. R. K. (2020). Forecasting demand profiles of new products. *Decision Support Systems*, *139*, 113401. https://doi.org/10.1016/j.dss.2020.113401

Vicenç, T., & Yasuo Narukawa. (2007). From the Weighted Mean to Fuzzy Integrals. In *Modeling Decisions* (pp. 147–196). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68791-7_6

Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior*. *Theory of Games and Economic Behavior* (Vol. 40).

https://doi.org/10.2307/3610940

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565–578. https://doi.org/10.1002/for.1213

Wang, C.-H., & Fong, H.-Y. (2016). Integrating fuzzy Kano model with importance-performance analysis to identify the key determinants of customer retention for airline services. *Journal of Industrial and Production Engineering*, *33*(7), 450–458. https://doi.org/10.1080/21681015.2016.1155668

Wang, C. H., & Chen, J. N. (2012). Using quality function deployment for collaborative product design and optimal selection of module mix. *Computers and Industrial Engineering*, *63*(4), 1030–1037. https://doi.org/10.1016/j.cie.2012.06.014

Wang, C. H., & Wang, J. (2014). Combining fuzzy AHP and fuzzy Kano to optimize product varieties for smart cameras: A zero-one integer programming perspective. *Applied Soft Computing Journal*, *22*, 410–416. https://doi.org/10.1016/j.asoc.2014.04.013

Wang, C. H., & Wu, C. W. (2014). Combining conjoint analysis with Kano model to optimize product varieties of smart phones: A VIKOR perspective. *Journal of Industrial and Production Engineering*.

https://doi.org/10.1080/21681015.2014.918566

Wang, Gang, Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*, *57*(1), 77–93. https://doi.org/10.1016/j.dss.2013.08.002

Wang, Guan, Xie, S., Liu, B., & Yu, P. S. (2012). Identify Online Store Review Spammers via Social Review Graph. *ACM Trans. Intell. Syst. Technol. Article*, *3*(21). https://doi.org/10.1145/2337542.2337546

Wang, J. (1999). Fuzzy outranking approach to prioritize design requirements in quality function deployment. *International Journal of Production Research*, *37*(4), 899–916. https://doi.org/10.1080/002075499191599

Wang, T., & Zhou, M. (2020). A method for product form design of integrating interactive genetic algorithm with the interval hesitation time and user satisfaction. *International Journal of Industrial Ergonomics*, *76*, 102901. https://doi.org/10.1016/j.ergon.2019.102901

Wang, W., Xu, H., & Wan, W. (2013). Implicit feature identification via hybrid association rule mining. *Expert Systems with Applications*, *40*(9), 3518–3531. https://doi.org/10.1016/j.eswa.2012.12.060

Wang, X.-Z., He, Y.-L., Dong, L.-C., & Zhao, H.-Y. (2011). Particle swarm optimization for determining fuzzy measures from data. *Information Sciences*, *181*(19), 4230–4252. https://doi.org/10.1016/J.INS.2011.06.002

Wang, X., Fang, H., & Song, W. (2020). Technical attribute prioritisation in QFD based on cloud model and grey relational analysis. *International Journal of Production Research*, *58*(19), 5751–5768. https://doi.org/10.1080/00207543.2019.1657246

Wang, Y.-J. J. (2014). A criteria weighting approach by combining fuzzy quality function deployment with relative preference relation. *Applied Soft Computing Journal*, *14*(PART C), 419–430. https://doi.org/10.1016/j.asoc.2013.10.001

Wang, Y. J. (2020). Combining quality function deployment with simple additive weighting for interval-valued fuzzy multi-criteria decision-making with dependent evaluation criteria. *Soft Computing*, *24*(10), 7757–7767. https://doi.org/10.1007/s00500-019-04394-5

Wang, Yanyu, Pei, L., & Wang, Z. (2017). The nls-based grey bass model for simulating new product diffusion. *International Journal of Market Research*, *59*(5), 655–670. https://doi.org/10.2501/IJMR-2017-045

Wang, Yue, & Tseng, M. M. (2011). Adaptive attribute selection for configurator design via Shapley value. In *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* (Vol. 25, pp. 185–195). Cambridge University Press. https://doi.org/10.1017/S0890060410000624

Wang, Z., & Klir, G. J. (2013a). *Fuzzy measure theory*. *IEEE Transactions on Fuzzy Systems* (Vol. 3). Plenum Press. https://doi.org/10.1109/TFUZZ.1995.481959

Wang, Z., & Klir, G. J. (2013b). *Fuzzy measure theory*. *IEEE Transactions on Fuzzy Systems* (Vol. 3). Springer US. https://doi.org/10.1109/TFUZZ.1995.481959

Wind, J., & Mahajan, V. (1997). Issues and Opportunities in New Product Development: An Introduction to the Special Issue. *Journal of Marketing Research*, *34*(1), 1–12. https://doi.org/10.1177/002224379703400101

Won, E. J. S., Oh, Y. K., & Choeh, J. Y. (2018). Perceptual mapping based on web search queries and consumer forum comments. *International Journal of Market Research*, *60*(4), 394–407. https://doi.org/10.1177/1470785317745971

Wu, C., Wu, F., Wu, S., Yuan, Z., & Huang, Y. (2018). A hybrid unsupervised method for aspect term and opinion target extraction. *Knowledge-Based Systems*, *148*(3), 66–73. https://doi.org/10.1016/j.knosys.2018.01.019

Wu, Q. (2010). Product demand forecasts using wavelet kernel support vector machine and particle swarm optimization in manufacture system. *Journal of Computational and Applied Mathematics*, *233*(10), 2481–2491. https://doi.org/10.1016/j.cam.2009.10.030

Wu, X., & Liao, H. (2021). Customer-oriented product and service design by a novel quality function deployment framework with complex linguistic evaluations. *Information Processing and Management*, *58*(2), 102469. https://doi.org/10.1016/j.ipm.2020.102469

Xi, L., Zhang, H., Li, S., & Cheng, J. (2020). Integrating fuzzy Kano model and fuzzy importance–performance analysis to analyse the attractive factors of new products. *International Journal of Distributed Sensor Networks*, *16*(5). https://doi.org/10.1177/1550147720920222

Xia, L., Wang, Z., Chen, C., & Zhai, S. (2016). The Electronic Library Research on feature-based opinion mining using topic maps. *The Electronic Library*, *34*(3), 435–456. Retrieved from https://doi.org/10.1108/EL-11-2014-0197

Xie, S., Wang, G., Lin, S., & Yu, P. S. (n.d.). Review Spam Detection via Time Series Pattern Discovery. Retrieved from http://delivery.acm.org/10.1145/2190000/2188164/p635-xie.pdf?ip=202.125.194.17&id=2188164&acc=ACTIVE SERVICE&key=CDD1E79C27AC4E65.5081EC796BE7F652.4D4702B0C3E38B35.4D4702B0C3E38B35&__acm__=1529945530_8d4a8fe0b4a82160b35ad85b9019fd62

Xu, H., Zhang, F., & Wang, W. (2015). Implicit feature identification in Chinese reviews using explicit topic mining model. *Knowledge-Based*

Systems, *76*, 166–175. https://doi.org/10.1016/j.knosys.2014.12.012

Xu, Q., Jiao, R. J., Yang, X., Helander, M., Khalid, H. M., & Opperud, A. (2009). An analytical Kano model for customer need analysis. *Design Studies*, *30*(1), 87–110. https://doi.org/10.1016/J.DESTUD.2008.07.001

Yan, H.-B., Ma, T., & Li, Y. (2013). A novel fuzzy linguistic model for prioritising engineering design requirements in quality function deployment under uncertainties. *International Journal of Production Research*, *51*(21), 6336–6355. https://doi.org/10.1080/00207543.2013.796423

Yan, Z., Xing, M., Zhang, D., & Ma, B. (2015, November 1). EXPRS: An extended pagerank method for product feature extraction from online consumer reviews. *Information and Management*. North-Holland. https://doi.org/10.1016/j.im.2015.02.002

Yang, C., Cheng, J., & Wang, X. (2019). Hybrid quality function deployment method for innovative new product design based on the theory of inventive problem solving and Kansei evaluation. *Advances in Mechanical Engineering*, *11*(5), 1–17. https://doi.org/10.1177/1687814019848939

Yang, Y., & Soh, C. K. (2000). Fuzzy Logic Integrated Genetic Programming for Optimization and Design. *Journal of Computing in Civil Engineering*, *14*(4), 249–254. https://doi.org/10.1061/(ASCE)0887-3801(2000)14:4(249)

Yeh, T. M., Pai, F. Y., & Liao, C. W. (2014). Using a hybrid MCDM methodology to identify critical factors in new product development. *Neural Computing and Applications*, *24*(3–4), 957–971. https://doi.org/10.1007/s00521-012-1314-6

Ying, C. shuo, Li, Y. L., Chin, K. S., Yang, H. T., & Xu, J. (2018). A new product development concept selection approach based on cumulative prospect theory and hybrid-information MADM. *Computers and Industrial Engineering*, *122*, 251–261. https://doi.org/10.1016/j.cie.2018.05.023

Yucesan, M., Gul, M., & Celik, E. (2017). Application of artificial neural networks using bayesian training rule in sales forecasting for furniture industry. *Drvna Industrija*, *68*(3), 219–228. https://doi.org/10.5552/drind.2017.1706

Yun, J. J., Jeon, J. H., Park, K. B., & Zhao, X. (2018). Benefits and costs of closed innovation strategy: Analysis of Samsung's Galaxy Note 7 explosion and withdrawal scandal. *Journal of Open Innovation: Technology, Market, and Complexity*, *4*(3). https://doi.org/10.3390/joitmc4030020

Zeng, L., & Li, F. (2013). A classification-based approach for implicit feature identification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8202 LNAI, pp. 190–202). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41491-6_18

Zhang, C., Tian, Y. X., & Fan, L. W. (2020). Improving the Bass model's predictive power through online reviews, search traffic and macroeconomic data. *Annals of Operations Research*, *295*(2), 881–922. https://doi.org/10.1007/s10479-020-03716-3

Zhang, Fan, Xu, H., & Bai, X. (2017). On the need of hierarchical emotion classification: Detecting the implicit feature using constrained topic model. *Intelligent Data Analysis*, *21*(6), 1393–1406. https://doi.org/10.3233/IDA-163181

Zhang, Fanglan, Yang, M., & Liu, W. (2014). Using integrated quality function deployment and theory of innovation problem solving approach for ergonomic product design. *Computers and Industrial Engineering*, *76*(1), 60–74. https://doi.org/10.1016/j.cie.2014.07.019

Zhang, R., & Tran, T. (2008). An entropy-based model for discovering the usefulness of online product reviews. In *Proceedings - 2008 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2008* (pp. 759–762). IEEE. https://doi.org/10.1109/WIIAT.2008.149

Zheng, X., Zhu, S., & Lin, Z. (2013). Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach. *Decision Support Systems*, *56*(1), 211–222. https://doi.org/10.1016/j.dss.2013.06.002

Zhong, S., Zhou, J., & Chen, Y. (2014). Determination of target values of engineering characteristics in QFD using a fuzzy chance-constrained modelling approach. *Neurocomputing*, *142*, 125–135. https://doi.org/10.1016/j.neucom.2014.01.052

Zhou, F., Jianxin Jiao, R., & Linsey, J. S. (2015). Latent Customer Needs Elicitation by Use Case Analogical Reasoning From Sentiment Analysis of Online Product Reviews. *Journal of Mechanical Design*, *137*(7), 071401. https://doi.org/10.1115/1.4030159

Zhou, F., Jiao, J. R., Chen, S., & Zhang, D. (2011). A case-driven ambient intelligence system for elderly in-home assistance applications. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *41*(2), 179–189. https://doi.org/10.1109/TSMCC.2010.2052456