Feature-based service and product feedback analysis model via social media data mining

By

Anish Kumar Varudharajulu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Engineering Management

Department of Mechanical Engineering

University of Alberta

ABSTRACT

The growth of an organization in the market relies on customer's satisfaction towards its products and services. Due to the dynamic nature of the internet, and increasing blogs, forums, and customer feedback, it usually remains a key issue in any industry to identify and extract data attributes manually. However, the effort required to collect the data and analyze it is cumbersome. To overcome these difficulties, organizations need to use big data processes to focus on customers by automating their information processing with good qualification to improve product quality, and customer satisfaction. Starting with discussions on the need for Concurrent and Collaborative Engineering (CCE) and the impact of Virtual Enterprise (VE) based product models on networked business and management, this research then describes a theoretical framework that creates a system linking the customers' perspective of product features directly to the enterprise management. It has been shown that customer feedback information sharing across the organization can be achieved and useful using state-of-the-art data mining technologies. This framework demonstrates how data mining could enhance enterprise management through improved information sharing, efficient feedback utilization, and improved feature performance tracking along the lifecycle of the product and services, as well as the reduced lead time for improvement. The significant components of the proposed framework are social media data collection module, the data preprocessing module where data is filtered and formatted to an acceptable format, and the module of classifying and clustering where the data collected is visualized after being processed and analyzed using big data analytical tools to identify product features and their performances. Finally, the ability of the framework to discover customer satisfaction towards the key service and product features is demonstrated in a real-life case study. Observations and insights from this research could provide prototyping experience and case studies for academic, business ventures, and industry practitioners to implement the discussed big data

techniques in related fields.

PREFACE

This thesis is the original work of Anish Kumar Varudharajulu who completed the thesis work under the supervision of Dr. Yongsheng Ma. The overall research direction was suggested by Dr. Yongsheng Ma; while the specific research topics, framework, proposed methodology, software application, and technical papers were developed by Anish Kumar Varudharajulu. The contents of this thesis have been published in the following journals or conference proceedings.

List of proceedings

1. A.K. Varudharajulu and Y. Ma, "A Survey on Big Data Process Models for E-Business, E-Management, E-Learning, and E-Education," published in *International Journal of Innovative Research in Computer and Communication (IJIRCCE)*, pp. 9523–9553, 2018.
2. A. K. Varudharajulu and Y. Ma, "Feature-based Restaurant Customer Reviews Process Model using Data Mining," published in *Proceedings of the International Conference on Computing and Big Data (ICCBD)*, pp. 32–37, 2018.
3. A. K. Varudharajulu and Y. Ma, "Feature-based Facebook Reviews Process Model for E-Management using Data Mining," published in *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E)*, 2019.
4. A. K. Varudharajulu and Y. Ma, "Data mining algorithms for a feature-based customer reviews process model with engineering informatics approach," accepted in *the 3rd International Conference on Artificial Intelligence, Automation and Control Technologies (AIACT)*, 2019.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere appreciation to the many wonderful people who have supported me during my studies. First, I would like to express my deepest gratitude to my supervisor, Dr. Yongsheng Ma, for his support, direction, encouragement, and support throughout this research.

I am grateful to MITACS Accelerate grant (IT09867, PI: YongshengMa) and NSERC Discovery grant (NSERC RGPIN 5641 Ma, PI: Yongsheng Ma) for their generous financial support of my research. Also, the University of Alberta has provided me with eight months of industrial experience as a Graduate Research Intern, with one of the leading Window and Door manufacturing company in Canada, which helped me in getting familiar with the industry.

I would like to express my thanks to all my colleagues in my research group, Gail, Richard, Jessica, Bernie, Rick, and to the all staff at the Department of Mechanical Engineering, University of Alberta, for their care and technical support. I would also like to thank my Big Boss family - Sushmitha, Rakshith, Saraswathi, Sathish, Anil, Harish, Ananthan, Vishal, and Ankit for their help and moral support.

Finally, my deepest gratitude to my family for all the love and encouragement. Especially, I thank my Dad, Varudharajulu; Sister, Rashmika Naidu; and Girlfriend, Pooja Chakrabarthy for always inspiring me and supporting me through everything. Also, I would like to dedicate this work to my mother, Veeralakshmi, for her endless love, hope, and dedication to make me study throughout my schooling.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
APIs	Application Programming Interfaces
ARFF	Attribute-Relation File Format
BI&A	Business Intelligence and Analytics
BoL	Beginning of Lifecycle
BoM	Bill of Materials
CAAQS	Canadian Ambient Air Quality Standards
CCE	Concurrent and Collaborative Engineering
CNC	Computer Numerical Control
CRM	Customer Relationship Management
DBMS	Database Management Systems
Durabuilt	Durabuilt Windows and Doors Inc.
EoL	End of Lifecycle
ERP	Enterprise Resource Planning
GPS	Global Positioning Systems
GUI	Graphical User Interface
I/O	Input or Output
ICT	Information and Communication Technology
MCC	Matthews's Correlation Coefficient
MNC	Multi National Company
MoL	Middle of Lifecycle
NPD	New Product Development
PB	PetaByte

PDM	Product Data Management
PLM	Product Lifecycle Management
R&D	Research and Development
RFID	Radio Frequency Identification
ROC	Receiver Operating Characteristic
SCM	Supply Chain Management
SDK	Software Development Kits
URL	Uniform Resource Locator
VE	Virtual Enterprise
WEKA	Waikato Environment for Knowledge Analysis
ZB	ZettaByte

CHAPTER 1. INTRODUCTION

1.1 Background

A study in 2004 [1] about 35 years of internet use concluded that any future technological advancement of the internet will be by the people, who utilize and add to make it prominent. Organizations need to keep abreast of the latest improvements and advancements of the internet to truly exploit its abilities and potential outcomes. Today we realize that a whole lot of data is created by people in various social media platforms.

Schools utilize the internet for E-Learning platforms and as an immense electronic library to facilitate E-Education, with untold potential outcomes. Engineering, medical, and other fields utilize the internet for E-Management and/or E-Business purposes to consult with partners a large portion of the world away in a time and cost-effective manner. Existing computers have great difficulty in storing and processing big data, and there is a need to consolidate the data from diverse sources. Data can be used as a strategic resource for making decisions at a low cost.

1.1.1 Data mining

The applications of machine learning and AI (Artificial Intelligence) in engineering were developed as early as the 1980s [2] [3]. Machine learning can be broadly applied to classes of tasks that may appear unrelated to the non-technical eye [4]. Data mining can be widely applied to any field; [5] analyzes soil data using data mining techniques to verify if the soil in the Salem district is good for agriculture; in the field of education, [6] uses data mining to gain valuable insights into students' learning experience using Facebook; [7] uses data mining in credit card risk assessment and fraud detection.

1.1.2 Big data

Big data focuses not only on the size of data in storage but also on other important attributes of big data, such as data variety and data velocity [8]. For example, the definition from Wikipedia illustrates that "big data" is a term for any collection of large and complex data sets which traditional data processing methods find difficult to analyze. In a straighter definition, big data just means data that it is too big, too fast, or too hard for existing tools to process [9]. Data mining techniques can complete a process of representing, analyzing, and extracting actionable patterns and trends from raw social media data [10] and use it in pattern recognition, decision making, and sentiment analysis.

The challenging properties of big data are as follows [11]:

1) Volume refers to the amount of data generated from various feedback in the form of surveys, emails and customer contact forms, usability tests, exploratory customer interviews, on-site activity, comment boxes, and instant feedback data in the websites of different organizations. There is a massive amount of feedback lying unused in social media and the internet in the form of customer reviews.
2) Variety (Scattered) refers to the different types of data such as blogs, music, videos, pictures, and geographical position.
3) Velocity (Dynamic) refers to the high speed of the data process, i.e., the entire information supply chain must be near real time.
4) Veracity refers to low reliability and disordered data.
5) Value refers to the ability to extract valuable information that can be used to improve services.

Data mining techniques are commonly used to address such challenges. Raju and Sravanthi [12] provide information regarding the application of the concept and techniques of web mining for social networks analysis and review the related literature about web mining and social networks analysis. They also provide inputs on how to use web mining; a general process of using web mining for social networks analysis has also been studied.

1.1.3 Big data repositories

The trend of big data has evolved mainly because of the available sources of data. We have several big data repositories [13] that can be used for data acquisition such as 1) Facebook Graph [14], although much of the information on users' Facebook profiles is private, a lot is not. Facebook provides the Graph API as a way of querying the huge amount of information that its users agree to share with the world; 2) UK Government open data portal, data from the UK Government, including the British national bibliography metadata on all UK books and publications since 1950 [15]; 3) The CIA World Factbook [16] provides information on the history, people, government, economy, energy, geography, communications, transportation, military, and transnational issues for 267 countries. The reference tab includes maps of the major world regions, as well as flags of the world, a physical map of the world, a political map of the world, a world oceans map, and a Standard Time Zones of the world map; 4) The US Government portal [17] has information on everything from climate to crime; 5) Google trends [18] provide statistics on search for any given term, since 2004, as well as epidemiology and population statistics; 6) The National Climatic Data Center [19] is a huge collection of environmental, meteorological, and climate data sets from the US National Climatic Data Center. It is the world's largest archive of weather data; 7) Likebutton [20] mines Facebook's public data, globally and from your own network, to give an overview of what people "Like" at the moment. All these data are available to conduct research, develop web/mobile applications, and design data visualizations.

1.1.4 Social Media

Social media is the main source of information. Nowadays people are more expressive on social sites, whether about oneself, organizations, or society. As an example, here are a few examples of what data about a person is scattered on social media. Twitter is an unending stream of encounters, suppositions, and assessments of the customer about everything from PCs to films which can be utilized by organizations. For LinkedIn, a major part of the revenue comes from providing access to information about its members to recruiters and sales professionals. On Pinterest, if a product has a high number of pins and re-pins, this generally tells the producer of the product that it is well liked by many members of the Pinterest community. Now that Pinterest lets marketers access the data, companies can view user comments on the product to learn why people like or dislike a particular product.

Google search, web, pictures, news, web journals, and so forth, monitors whatever we search. Googlebot visits all websites and detects links (SRC and HREF) on each page and adds them to its list of pages to crawl. Crawling is the process by which Googlebot discovers new and updated pages to be added to the Google index. In the aspect of site examination, Google Analytics tracks the activity of a site. YouTube has viewing, and subscription details of their clients. From electronic mail applications, the mail content of both sent and received messages are parsed and broken down. Contacts in Google Talk, Gmail, and Android are followed by Google. Google also has data from their applications like docs, spreadsheets, calendar, and so forth. All these are profitable information/records of clients. In Facebook, you can put advertisements using parameters like area, sex, age, likes and interests, relationship status, working environment and instruction of your intended interest group. Facebook is even tracking your mouse movements to recognize whether or not it is a robot. Another way they recognize it is by monitoring whether your computer is foregrounded or backgrounded. According to the study conducted in 2017 [21], two billion users makes Facebook the largest social app in terms of logged-in users, above YouTube's 1.5 billion, WeChat's 889 million, Twitter's 328 million, and Snapchat's estimated 255 million (see Figure 1-1).

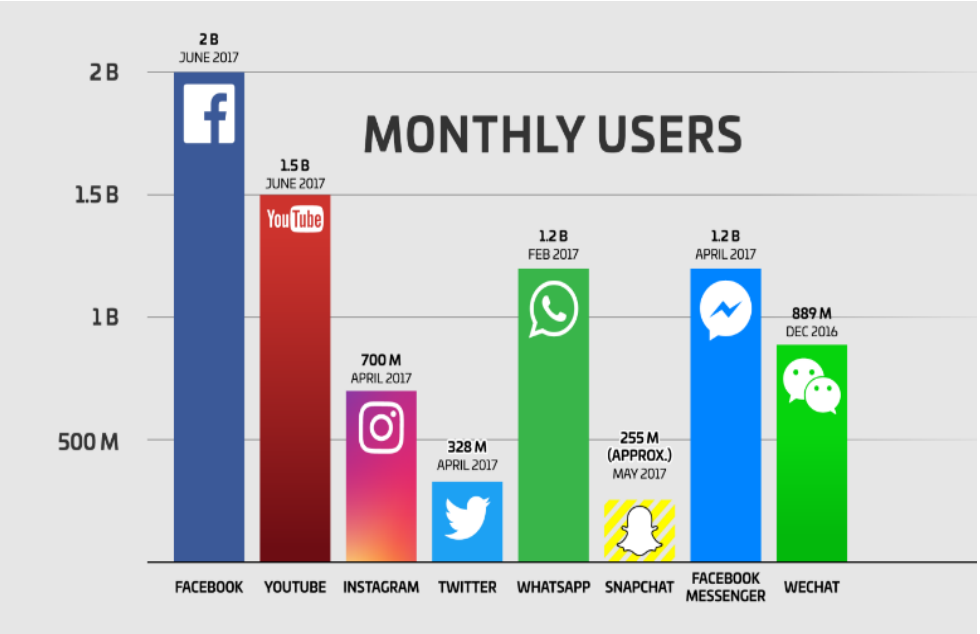

Figure 1-1 Social media users

The Facebook platform is a vital part of the big data ecosystem and can be significantly utilized with profound reach to users, consumers, businesses, governments, and non-profit organizations. Businesses can use Facebook as their business execution platform for product design, innovation, decision making, consumer or stakeholder relations management, and marketing [22]. For example, personal info like gender on Facebook can be used to pop up related clothing ads for male and female users.

1.2 Problem Statement, Scope of Work, and Research Objective

Traditional feedback systems used surveys or questionnaires to gather customer needs and interests manually. Now, the customers do not have the time and patience to write manual feedback. Also, the effort required to collect and analyze the existing social media feedback is cumbersome. This research resolves these issues by gathering customers' thoughts about the existing product features performances and future improvements they are expecting. We have developed a feedback analysis tool to gather feedback using customer reviews from Facebook about various product and service features. By doing so, we can generate positive assurance on quality management generically in the long run, and eventually create economic benefits in the forms of higher satisfaction rate, retaining customers, sustained business growth and reputable brand name in the sector.

The key deliverable of the research work is to develop a software prototype to:

1) Enable real-time extraction of customer comments from Facebook.
2) Pre-process the collected reviews and make them intelligible to WEKA.
3) Categorize the reviews in order to find which product or service feature the customer is concerned about.
4) Classify the positive and negative opinions of the customers about the different features under consideration.
5) Generate graphs for tracking performance of different product and service features in real-time.

Key objectives of the proposed feature-based feedback analysis model are as follows:

1) To enable real-time extraction of customer comments from social media and reflect uniformly the opinions of customers, and then provide a systematic basis for management review and decision making for the operation.
2) To create a quality improvement cycle (see Figure 1-2) with the help of a software toolkit so that the management can measure the effectiveness of process changes and new innovations on parts and other product features.

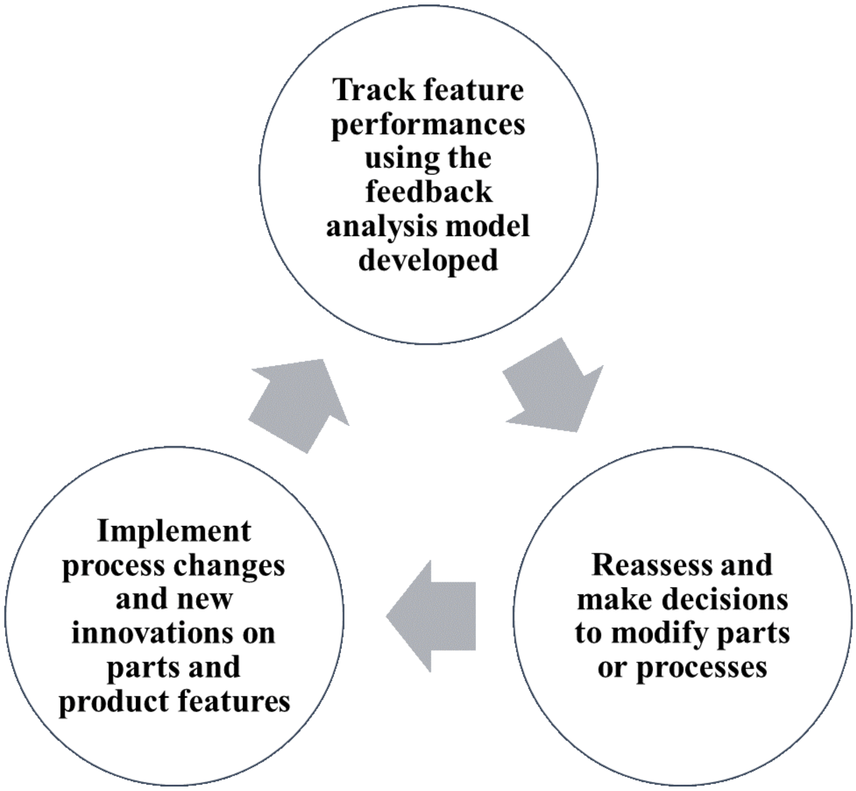

Figure 1-2 Product and service quality improvement continuous cycle

1.3 Thesis Organization

This chapter introduces the topic and research objective and provides a general context of data mining.

Chapter 2 summarizes the literature review gathered on this topic, discussions on the need for Concurrent and Collaborative Engineering (CCE), and the impact of Virtual Enterprise (VE) based data mining product models on networked business and management

Chapter 3 presents the methodology used in this research, which consists of three main elements: (1) Data collection, (2) Data pre-processing, and (3) Data classification and Data clustering. It also provides real-life empirical studies on various classification and clustering algorithms using WEKA.

Chapter 4 elaborates the different data mining algorithms that are considered for implementation in two different enterprises.

Chapter 5 discusses the implementation of the developed real-time software application using a case study on a famous South Indian restaurant in Edmonton to analyze the performance of different service features.

Chapter 6 discusses the implementation of the developed real-time software application on an engineering firm's products and processes.

Chapter 7 concludes the research with highlights of the contribution of the study and proposes some recommendations for future work.

CHAPTER 2. LITERATURE REVIEW

This chapter presents a critical review with respect to the need for CCE and data mining in various Virtual Enterprise (VE) based product and process models in order to clarify the point of departure for this research. Feature-based process models [23] that can convert big data customer feedback from the internet into valuable engineering feature information will enhance global competitiveness and make an organization to lead their market by interpreting and satisfying customers' expectations.

The applications of Information and Communication Technology (ICT), such as machine learning and artificial intelligence in engineering were developed as early as the 1980s [2] [3]. Machine learning can be broadly applied to classes of tasks that may appear unrelated to the non-technical eye [4]. In this era of ICT, Data mining is widely applied in many fields, [5] analyzes soil data using data mining techniques to verify if the soil in Salem district is good for agriculture using data collected from Farm Science Centre. In the field of education [6] uses data mining to gain valuable insights into students' learning experience using Facebook. Global competition is increasing with pressure on prices, smaller orders, shorter life cycles, geographically dispersed Research and Development (R&D) teams, more suppliers, more governmental regulations, and increasing material and energy costs. These new business drivers make manufacturers pursue a more competitive Virtual Enterprise (VE) and apply concurrent and collaborative manufacturing practices [24].

2.1 Concurrent and Collaborative Engineering Informatics

In this era, the traditional step by step process of product design and realization have been replaced by concurrent product development practices. For instance, a car manufacturer does not necessarily manufacture all parts in a step by step process. Several parts like chassis, gearbox, engine, and transmission are manufactured concurrently and sometimes in distributed locations. So, the need has arisen for collaborative systems to manage and monitor product feature data and feature performances across all departments of the manufacturer.

A simple example of CCE is the automotive industry. When the first automobile was introduced, it was purely a device capable of generating its own power to move. But modern cars have high tech systems such as lane changing assistance, reverse cameras, artificial intelligence software, dual zone climate control, and remote starters, all resulting from CCE by engineers in various fields, such as mechanical, electrical, and robotics.

There is also a need for collaboration between the four engineering dimensions - processes, information, organization, and technology in today's virtual organizations where the producers and suppliers work together in engineering processes and share their core competencies or when in Multi-National Companies (MNC) in diverse locations [25].

2.1.1 CCE in Product Design

The uncertainty in product design decision-making is primarily due to the variations in customer specifications. This issue is prevalent in most project-based companies, Engineer-to-Order, and mass customization units. A small change in the product specification can incur costs in the form of product development and validations, which can be significant if there is a need to involve suppliers in design decisions. This can ultimately lead to rework and loss of man-hours. So, there is a need for transforming from a product-based system to a product-service system as proposed by G. Pezzotta et al. [26] and customer focus. We can overcome this proactively by concurrently analyzing the customer interests using real-time web informatics systems, and customers' key needs can be known more accurately even before the design phase.

From a survey of the world's top 500 enterprises [27], it can be concluded that they usually search for user feedback on the internet during the New Product Development (NPD) phase. In some circumstances, such as the company's online community, users are even permitted to join in the design phase. Furthermore, feedback can be quickly involved in product design with the help of big data analysis capability. NPD is cross-functional and there is a need for more effective collaborative information systems.

Robin et al.[28] proposed a framework to manage collaborative design by capitalizing relationships between humans as well as knowledge exchanged between various stakeholders, but the model is limited to internal stakeholders only and does not include the key stakeholder i.e. customer. Integrating our software module to the above framework can enhance the efficiency of such collaborative systems.

2.1.2 CCE in Mass Customization

Let us consider the state of the art of a typical manufacturing company. Once an order is received, there is a need for a feasibility study of the various parts involved and modifications available. The engineering design criteria should be validated based on which suppliers can be selected to raise purchase orders. An estimate can be given to the customer only after all these processes are done. Thus, these activities must be carried out concurrently to reduce the total lead time [29] and to increase customer satisfaction by accurate quoting. Having an effective informatics system to gather feedback along with the existing product extension services [30], and analyze them to identify the category of parts i.e. personalized products in which customers are interested can aid in updating the configurator with the most generic parts and thus reduce the total lead time.

2.1.3 Collaborative Requirements Gathering

Requirement gathering is the basis for defining and managing the project scope, including the product scope. Kamrani and Nasr [31] suggested that the first step towards Integrated Product Development in a CCE environment is understanding and managing customer requirements. Saikia and Shilpi [32] developed a model to analyze text features in requirement engineering document in which they have analyzed the text features. There is a need to analyze engineering features in such documents.

Due to the ever-increasing complexity of products and their customization, there is an accumulation of massive volume of requirements. There is a need for collaborative requirement mining frameworks [33], which can visualize the key requirements and have computation capabilities. The software model can be used as a collaborative tool to distill customer requirements by analyzing the feedback reviews from Facebook.

2.1.4 Product Lifecycle Collaboration

A Product Lifecycle Management (PLM) system is supposed to support the information needs regarding products ‘‘from cradle to grave,” within which many engineering stages are involved, such as industrial design, conceptual design, detailed design, production, process planning, manufacturing, assembly, sales, maintenance, and recycle or destroy. Such lifecycle stages are inter-related and mutually constraining. They also involve different computer-aided application tools to carry out the processes [34].

There is a need for effective collaboration between all the stakeholders of a company, such as customers, engineers, suppliers, and manufacturers throughout the product lifecycle. Different kinds of data flow along the lifecycle in the form of customer feedback; it should be stored and then transferred to other forms, such as a number, picture, chart, light, temperature, and so on. Figuring out customers' needs accurately and quickly is an effective means for manufacturers to increase customers' approval and loyalty.

2.1.5 Collaborative Feedback Management

Customer feedback is managed through customer support systems where customers can raise their complaints. For MNCs, there will be a huge amount of complaint tickets and there is a need to categorize the tickets into their respective engineering features as proposed by Lloyd et al. [35]. Facebook reviews are now widely used by customers to complain and express their dissatisfaction. The companies also give immediate attention and resolve the issues on a case by case basis to ensure loyalty and brand name success in open forums. Thus, arises the need for analyzing the feature information in Facebook reviews.

Most traditional feedback systems [36] fail because the users are not interested and unwilling to provide feedback. So as an alternative to such a feedback system, customer reviews in social media can be used. Organizations need to automate their feedback processes. Customer complaints management can be done accurately and quickly since most reviews [37] are either positive feedback or complaints [38]. Although customer feedback in the form of social media reviews has important roles in company decision-making, it usually remains a key issue in any industry to track them manually.

To overcome these difficulties, we are proposing an automated software model to analyze customer review features in real-time and provide feedback on feature performances to aid decision-making. In the case study, the software module uses customer reviews from Facebook page and analyzes them using various classification [39] and clustering algorithms [40] to aid in decision making and to identify key areas to be improved as a continual improvement process for increasing the customer satisfaction [41].

2.2 Review of Virtual Enterprise based Product Models in Networked Business

E-Business platforms are the major driver of global economics and trade. Businesses now have a lot of data available on the internet which can be collected at a low cost to make a real-time analysis of customer preferences and interests. E-Business big data stored using ICT includes server data like server logs and cookies logs that are stored in the form of document files; query data, which has stored information of customer's online search; customer registration information, which has all the demographic information about the customer; and online market data about purchases of customers, merchandise and so on. These data can be used to facilitate personalized marketing, to enhance target marketing and precision marketing in Customer Relationship Management (CRM), and to discover potential customers and markets. [42].

2.2.1 Agriculture

E-business platforms can be used in agriculture for getting the latest updates on the national network of cooperative policy and real-time market rates of agricultural products from agricultural offices. It can aid farmers in finding the latest information on agricultural products supply or demand information [43] and to learn about the new planting and cultivating production technology from the agricultural research experts.

ICT has enabled rapid innovations in Global Positioning Systems (GPS), remote sensing, agricultural equipment data, agriculture departments' data, and social data like agriculture forums, crowdsourcing and mobile sourcing for site-specific management. Leonid et al. [44] studied the augmented approach to emerging technologies identification in the agriculture and food sector. Cloud-based systems can gather queries from farmers directly through preconfigured devices and provide the required information using big data analytics without taking much time [45]. There is a need to track data about the reservoirs' monthly water history and global surface water using ICT [46]. On careful analysis, these data can help the government to identify high-risk areas, for making decisions on water usage, infrastructure construction, and storing and distribution of water. Big data systems for water management can aid in preventing losses from water shortage, food shortage, livestock, and fish breeding. There are losses due to the fact that many governments give importance to the agriculture production process while they fail to do so in the post-production process. We can compensate for this by facilitating agricultural e-commerce [47].

Climate change has increased the loss in agricultural production systems. Information about the environmental factors like greenhouse gas emissions, humidity, moisture, wind, water, carbon dioxide collected from sensors deployed in agricultural fields, and computed tomography digital images [48] is a major source of big data. These data can be used in precision agriculture [49] to assist agricultural layout and production based on farmland environmental factors [50]. Sana et al. [51] monitor chili crop and gray mold disease through a wireless sensor network. A study [5] analyzes soil data using data mining techniques to verify if the soil in the Salem district is good for agriculture using data collected from the Farm Science Centre. Regression models can be used to forecast rainfall [52]. Such predictive systems can increase profitability by enhancing the productivity of crops, food security, land, equipment and labor management, climate change awareness, and mitigation. In addition, they can be used to develop customized and prescriptive farm specific decision support systems [53].

Information gap in supply and demand networks can hinder farmers' income [54]. There is also a need to integrate the information about deforestation, afforestation, land degradation, water contamination, and economic, social, and cultural trends to develop sustainable farming practices that do not degrade our environment [55] [56]. The data from heterogeneous systems mentioned in this section can be integrated into one framework and processed with the help of the tools and techniques discussed in Chapter III.

2.2.2 E-Commerce

Internet marketing is the core of product sales in many e-commerce companies. Social media data mining has been widely used for the purpose of e-commerce on sites such as Facebook and YouTube by tracking customer preferences through the pages and channels visited. Customer clickstream analysis and path analysis can be done to determine the frequently visited products [57]. Using this data, association rules can be created for product placement optimization; product recommendation in the websites and potentially profitable products can be identified. By tracking the user's region, age, and other personal information, we can determine the appropriate location for advertisements. Three types of data can be extracted: (1) Historical data of expenditure and purchase of similar products; (2) Market research data from forums, questionnaires, surveys, and newspapers; (3) Browser data like browsing history and preferred websites. These three types of data are massive in size. The combination of these data requires ICT processing technology to help make decisions about targeting customers using recommender systems. It is alluring for internet business elements with restricted databases to combine their recommender framework databases to improve the unwavering quality of suggestions for clients and amplify the accuracy of focused showcasing while at the same time safeguarding the security of client inclinations [58].

The technical gap in these recommender frameworks is clarified through the case of eBay. If a user expresses his preferences for a certain category like books by rating it, this data was overlooked when the framework processes suggestions for things in another category like movies. In addition, it is natural to expect that a person buying a Harry Potter novel is most likely to buy Harry Potter movies. The idea of relating customer preferences in one class to another using customer ratings and feedback is still at the base of research [59]. Implementing such decision-making systems to forecast demands will enable the procurement teams to maintain just-in-time inventory and thus reduce delays in production and/or distribution.

2.2.3 Business Intelligence and Analytics

Business Intelligence and Analytics (BI&A) is the knowledge-based information processing to make business-centric decisions related to e-commerce, e-marketplace, e-banking, healthcare, and information security. Decision tree learning algorithms can be used in BI&A for prediction modeling, It uses a decision tree as a predictive model that maps observations about an item to conclusions about the item's target value [60]. A decision tree can be used in E-Business for customer churn predictions to take proactive actions and retain valuable customers [61]. Since decision tree can analyze numerical as well as categorical data, they can be used in E-Learning for predicting student performance [62] and determining the risk of a student failing or dropping out of a course.

Substantial healthcare data should be changed into knowledge and information, which can control cost and maintain the high caliber of patient care. Without data mining, it is hard to understand the significance of information gathered inside medical databases like Healthdata.gov [63] which has 125 years of US healthcare data including claim-level Medicare data. For example, we can use age as the dependent attribute to predict blood groups that have high risks of sickle cell disease [64].

Trained machine learning algorithms can be used to facilitate predicting fake Facebook profiles using attributes like age, gender, and the number of Facebook friends [65]. Similar algorithms can be used in credit card risk assessment and fraud detection [7] by training them with vulnerability datasets of customers holding credit cards. Mined social data from Facebook profiles can be used to extract intellectual knowledge for human behavior prediction. Human behavioral pattern recognition can be used by organizations to distribute the work to deliver high performance. The same can be used for the purpose of focused marketing [66].

Key functions of BI&A involve the processing of unstructured content by information retrieval, opinion mining, social media analysis or social network analysis. Pippal et al. [67] studied how to use data mining to generate effective business strategies. Web applications have the ability to gather a significant amount of feedback from a diverse customer population for different businesses. There is a need for developing appropriate ICT frameworks to find the key node of E-Business big data from this feedback, relate it to the organization nodes and tasks and aid in the process of decision making [68].

2.3 Review of Virtual Enterprise based Product Models in Management

2.3.1 Product Lifecycle Management

The concept of PLM was developed in the early 21st century as an extension of Product Data Management (PDM). PLM is supposed to support the information flow of a product from its cradle, i.e., conception, design, manufacturing, marketing, and sales to the grave, i.e., use, service, maintenance, and destroy or recycle [34].

Despite the amount of research and number of studies conducted in terms of product information modeling, there still remain some limitations. Design and manufacturing related information are widely addressed, whereas Middle of Lifecycle (MoL), End of Lifecycle (EoL), and customer related information such as maintenance, usage, recycling, and service, are not fully covered. Predicting models can be used in PLM for the below scenario, if an item is at its EoL, it does not imply that each part of it is futile. The majority of the time, the remaining value of the parts is worth predicting to determine which to reuse. The foreseeing procedure is not simple and includes the upkeep of historical information such as part ID, and Radio Frequency Identification (RFID) from the Beginning of Lifecycle (BoL) and MoL periods [27]. Fu et al. [69] provide insights into the management of repair activities in electrical and electronic equipment using big data and ICT.

Product lifecycles have declined; the organization bringing the product to the market first is the most profitable provided the product meets the customer expectations. Although there is a lot of groundwork before manufacturing a product, the most essential steps are marketing analysis and product design [27]. The most important expectation of these steps is to meet customer demands in various forms like comments on social media, the pages customers' like on Facebook, and the websites they bookmark.

Organizations have shifted from pushing standardized mass-produced products to a pull system able to respond to customer expectations, i.e., from a product-oriented market to a customer-oriented market. A comprehensive framework is required to analyze customer demands which are scattered over the internet.

2.3.2 Supply Chain Management (SCM)

SCM is an enterprise-wide view of unifying the core business processes of marketing, sales, product development, operations and other capabilities of an allied group of businesses to respond to marketplace opportunities as a single business entity. Previously businesses leveraged their internal competencies and performance measurements and had distinctive business strategies but SCM shifts attention towards customers and suppliers as new sources of competitive advantage by looking beyond the frontiers of their own organization. Research in business-to-business e-commerce [70] investigates interdependence and coordination between multiple firms and finds that supply channel mutuality, process redesign, and coordination are all closely associated with firm performance.

The goal of supply chain management is to synchronize the requirements of the customer with the stream of materials from suppliers to effect an equilibrium between the conflicting objectives of high customer service, low inventory, and reduced cost [71]. In a study [72] to determine how firms rely on processes to capture insights into what customers value, it was discovered that firms are not involved to the extent they could to reap the benefits of emergent supply chain opportunities with enough lead time to develop sustainable products.

ICT has created the linkage between once separate companies into a single competitive supply chain system. Big data sharing between Walmart and P&G [73] have enabled more effective category management, continuous replenishment, and process coordination, which can collectively make the supply chain more efficient. ICT can enable networking of geographically dispersed teams, connectivity between companies, vendors and customers for integrating the inventory system to the supply chain. For effective management of a company’s supply channels, the velocity of feedback that comes from various functions like sales, marketing, customer service, maintenance, and insurance services should be captured, stored and retrieved whenever necessary. In addition, the continuous improvements and innovations in ICT can cause scope creep of any operation/service, irrespective of its function.

Several studies have been made to mitigate the bullwhip effect in SCM which causes excess or shortage of inventories, unstable production activities and unplanned variations in capacity planning. In [74], the author analyzes the big data properties discussed in chapter I and infers that velocity has the greatest impact on the bullwhip effect. Supply chain applications are still using traditional data from their Enterprise Resource Planning (ERP) system only. Such systems will soon be obsolete if they do not have the capability to integrate the increasing volume, variety, and velocity of customer-generated data. In [75], the author discusses how big data analytics can be used to extract useful information and aid in supply chain decision making. The next wave will be an era of AI and machine learning where most of the supply chain decisions will be taken by algorithms which can evolve themselves to learn from data and different use cases.

2.3.3 ICT in Collaborative engineering

In hierarchy-based organizations, people were not allowed to communicate with other teams. There were a lot of formalities to even get data from other departments and they have to go through their supervisors. It was assumed that if all the manpower works optimally then the whole company is functioning successfully. But their gains were relatively small [76].

Elisha Ondieki Makori [77] studied the application of the ICT and the factors fostering the extraction of information. It was found that 91% of respondents extract data for the purpose of communication and effective collaboration between teams. There is a need to exchange product information and requirements near real time. For example, if reducing production time is the target then all departments, such as the body shop, paint shop, foundry, engine manufacturing, transmission manufacturing, assembly lines, and quality control, should work collaboratively to achieve the target. Sometimes one or more parts may be manufactured in vendor locations and there is again a need to collaborate. The compatibility of software developed by different companies and the technology alliances formed between different automotive companies are key examples of collaborative engineering.

Manufacturers are facing many challenges like better product quality and lower costs; a variety of parts from different vendors are needed to be procured due to complex designs and diverse teams scattered over many locations. Thus, product design changes must be communicated effectively. To respond to new business requirements and to overcome the aforementioned challenges, implementation of collaborative process planning and manufacturing technologies as discussed in the study [24] should be done. Frutos and Borenstein [78] propose a collaborative information model using the internet to quickly and effectively provide an interface between customers and companies. To improve the performance of such ICT models, integration of the social media feedback into the inter-enterprise data near real-time and enabling the design team to be aware of the customers' expectations are crucial.

2.3.4 Product Design and Development

Twenty years ago, Alan Mullaly, CEO of Boeing 777, mentioned AI in product design and development; Computers don't design airplanes. We have not put the knowledge that is in the airplane designer's head into AI. Right now, the knowledge to design airplanes is in the designer's head.

There were several reasons for his comment:

1) Lack of Standards: they did not have proper standards which focused on customers like customer satisfaction guidelines for complaints handling in organizations (ISO 10002 [79]), and guidelines for monitoring and measuring customer satisfaction (ISO

10004 [80]). Also, there were no proper standards for big data programming [81] and databases [82]. For instance, quality management systems fundamentals and vocabulary (ISO 9000 [41]) were first published in 1987.

2) Data Repositories: they did not have data sharing systems that support the process through which explicit or tacit knowledge is communicated to other individuals. But now after the introduction of different Database Management Systems (DBMS), we are in the generation of big data. A report in 2010 [83] states that the data storage units of measurement in the world have reached to PetaByte (PB) and ZettaByte (ZB).
3) The knowledge-acquisition bottleneck: this is due to inadequate tools for knowledge-acquisition in an intuitive manner. There is a need for the development of machine learning process models for automated knowledge acquisition from social media and other big data repositories.

Traditionally managers use sales data and sales call reports to make decisions on what customers will like. Focusing on what customers currently need is entirely different from focusing on what they will need in the future. A study [84] suggests that creating an environment to get insights about customers and looking for small clues to judge what customers may value in the future and take decisions to modify existing processes and product design was considered innovative by the customers.

In a few companies, users in online communities are even permitted to join in the design phase. An effective software support tool that can aid the designer to make better decisions based on customer feedback needs efficient data representation schemes. Senthil et al. [85] studied the future of knowledge representation in product design systems and supporting tools providing such information. In a study [78], the authors present the design and implementation of an information system framework for agile interactions between companies and customers in a mass customization environment.

There is a need to effectively customize static designs. For the success of a mass customization system, the main factor is knowledge management since it allows collaboration between customers and vendors. With the assistance of ICT, marketing can be more precise and specific than ever, which presents great opportunities for customization. Furthermore, the comments from online forums can be quickly involved in product design with the help of big data analysis capability. With the assistance of big data, product development can be more precise and specific than ever, which presents great opportunities for feature customization.

CHAPTER 3. RESEARCH METHODOLOGY

In this chapter, tools and techniques of data mining are reviewed along with real-life empirical studies on how data mining can be used.

3.1 Data Scraping

The internet is the main sources of data. To make the model close to reality, data should be dynamically extracted from Facebook / Twitter or other websites, normalized, and used further for classification and clustering. The available Application Programming Interfaces (APIs) can be integrated with the source code of the web application to extract data from various websites like Facebook / Twitter periodically. This can reduce the manual effort to a great extent. The next topic in this section describes one of the widely used API of Facebook.

3.2 Facebook API

The Facebook Graph API is the primary way for apps to read and write to the Facebook social graph. Software Development Kits (SDK) [86] and products can interact with the graph API in some way, so understanding how the Graph API works is crucial. SDK is typically a set of software development tools that aid in creating applications for a certain software package, software framework, hardware platform, computer system, video game console, operating system, or similar development platform.

We can use the Facebook Graph API to Input or Output (I/O) data on Facebook's platform. It is an HTTP-based API that can be used to query data, post new stories, manage ads, upload photos, and perform a variety of other tasks that an app might implement. Graph API needs an access token to process the request and retrieve data. An access token can be generated using the app id and Secret when registered at http://developers.facebook.com.

3.3 Machine Learning

The main aim of data mining is to discover the properties of datasets, including machine learning. Data mining uses machine learning to do prediction analysis. There are two main types of machine learning algorithms used for a specific purpose.

1) Supervised learning
2) Unsupervised learning

3.3.1 Supervised Learning

In supervised learning, the prediction is made on the target attribute on analyzing the given set of other attributes. The predicted attribute is the dependent attribute and others are independent attributes. The predictions are made on labeled data.

Example: Diabetes prediction

Here, the independent attributes are

1) Number of times pregnant
2) Plasma glucose concentration - 2 hours in an oral glucose tolerance test
3) Diastolic blood pressure (mm Hg)
4) Triceps skinfold thickness (mm)
5) 2-Hour serum insulin (mu U/ml)
6) Body mass index (weight in kg/(height in m)^2)
7) Diabetes pedigree function
8) Age (years)

Dependent attribute (also alternatively used as a class label) is the presence of diabetes, i.e., tested negative or positive. In this case, the outcome is discrete, i.e., either positive or negative. In the case of stock prediction, the outcome is a continuous value. Another division in supervised learning is where the algorithms that predict discrete values are called classification algorithms and those that predict continuous values are regression algorithms. But in both cases, the prediction is made on labeled data. An empirical study on some of the classification algorithms is explained in later sections.

3.3.2 Unsupervised Learning

This type of learning deals with unlabeled data. Unlike supervised learning, it does not assign a class value for the test data instead it groups the data based on its features. There are many ways in which the data to form data can be grouped into clusters. An empirical study on some of the clustering algorithms is explained in later sections.

3.4 WEKA (Waikato Environment for Knowledge Analysis)

WEKA is an open source data mining tool, developed by the University of Waikato in New Zealand that has all data mining algorithms implemented using Java. WEKA would not only provide a tool for learning algorithms, but also a framework inside which researchers could implement new algorithms without having to be concerned with supporting infrastructure for data manipulation and scheme evaluation. Nowadays, WEKA is recognized as a landmark system in data mining and machine learning. WEKA is a collection of machine learning algorithms for data mining tasks. It contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization [87]. WEKA supports various algorithms for generating mining models required by researchers such as clustering and classification. In classification, training examples can be used to train a model that can classify the data samples into known classes.

WEKA has four different modes:

1) Simple CLI: It is an environment to provide a simple command-line interface that allows direct execution of WEKA commands.
2) Explorer: It is an environment for exploring data.
3) Experimenter: It is an environment for performing experiments and conducting statistical tests between learning schemes
4) Knowledge Flow: Presents a "data-flow" inspired interface to WEKA

The main interface in WEKA (see Figure 3-1) is the explorer which has a set of panels, each of which can be used to perform a different task. Data can be loaded in WEKA from various sources, including files, URLs, and databases. Supported file formats include Attribute-Relation File Format (ARFF) format, CSV, and C4.5 format. Once a dataset has been loaded, one of the other panels in the explorer can be used. Steps to use WEKA for classification and clustering is shown below:

Step 1: Launch WEKA explorer.

Figure 3-1 WEKA explorer

Step 2: Under the preprocess tab → click the open file → choose the dataset as shown in Figure 3-2.

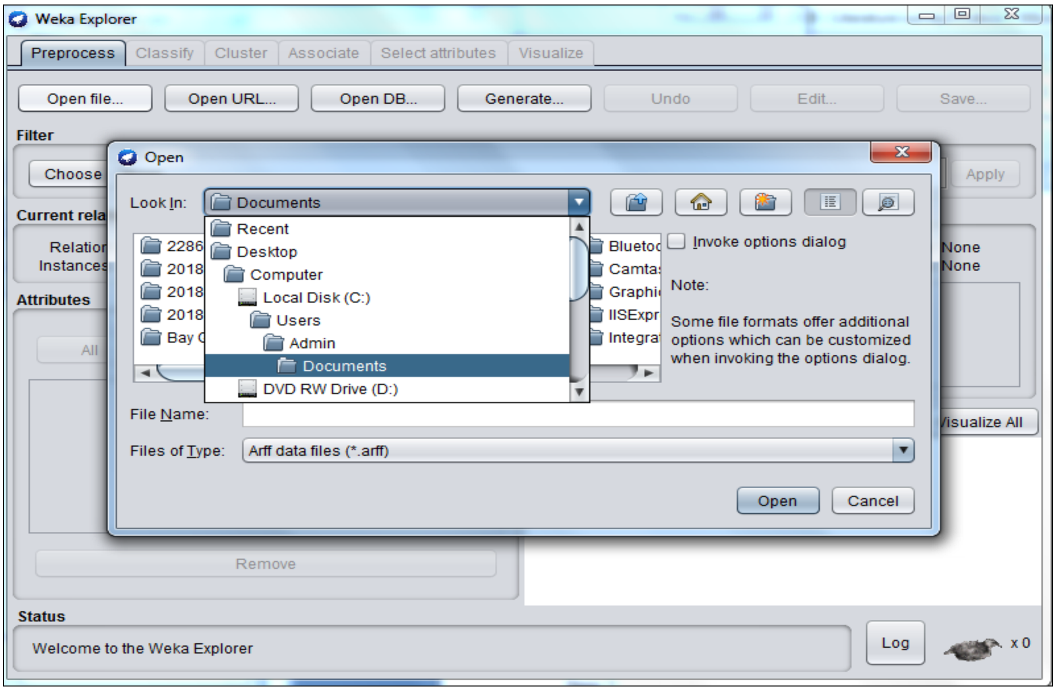

Figure 3-2 Load dataset in WEKA

Step 3a: Under classify tab, choose a classification algorithm as shown in Figure 3-3. The knowledge flow of the WEKA classifier is shown in Figure 3-4.

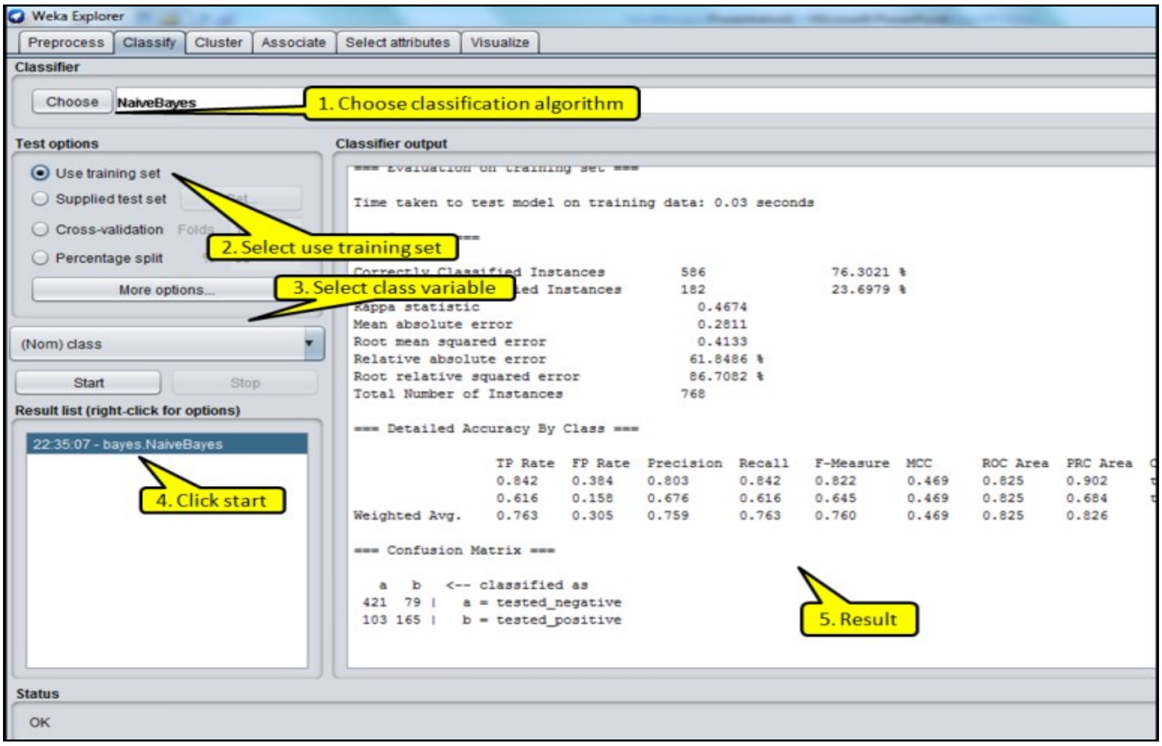

Figure 3-3 Classification window in WEKA

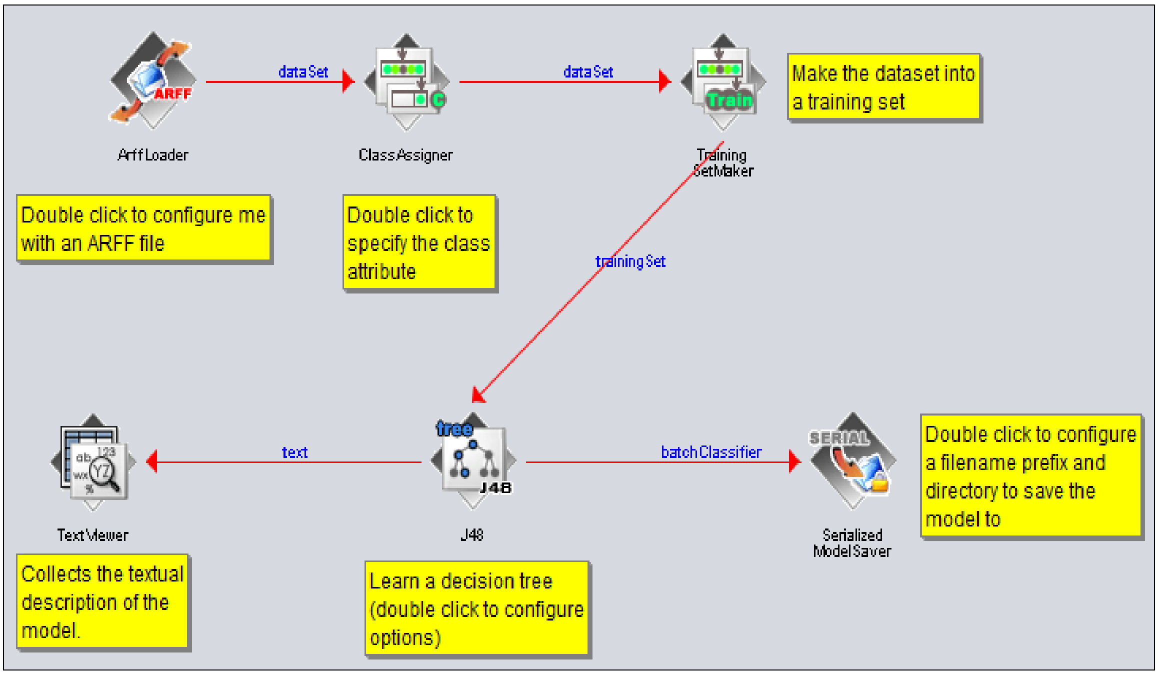

Figure 3-4 Knowledge flow of classifier in WEKA

Step 3b: Under the cluster tab, choose a clustering algorithm as shown in Figure 3-5. The knowledge flow of the clustering algorithm is shown in Figure 3-6.

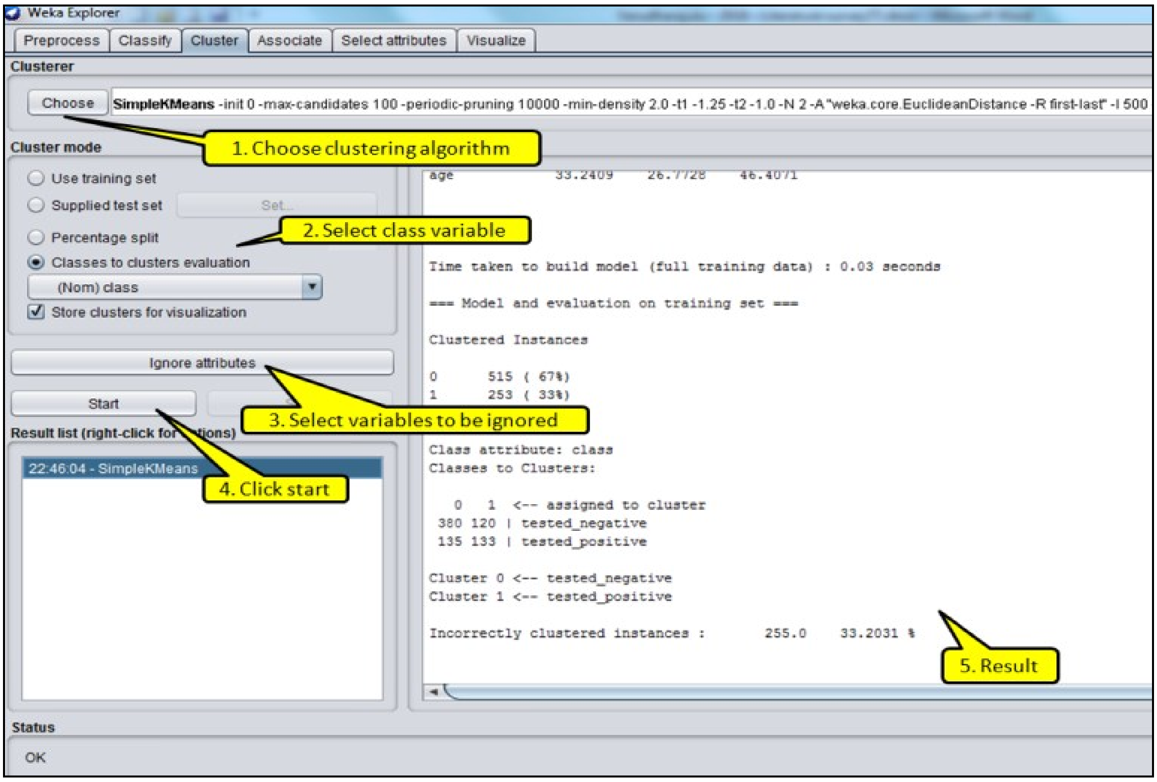

Figure 3-5 Clustering window in WEKA

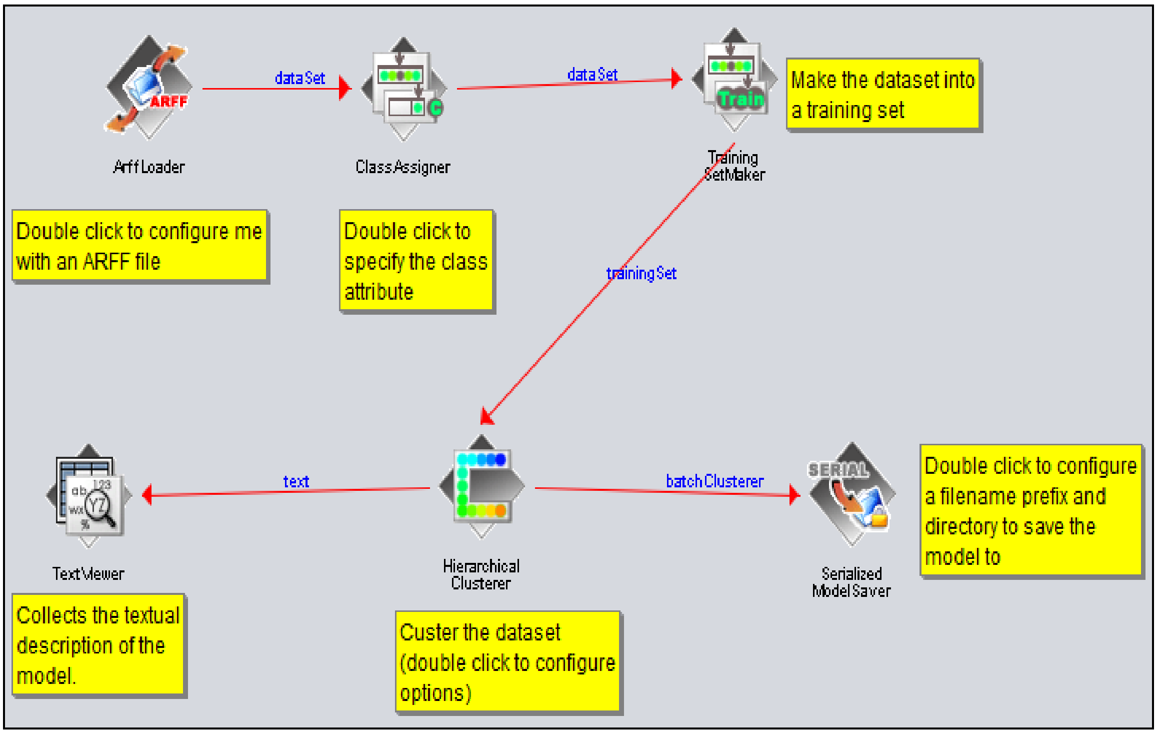

Figure 3-6 Knowledge flow of clustering in WEKA

3.5 Case Study

In this section, case studies done to understand the algorithms available in WEKA are discussed. Each algorithm under study was tested using different real-life data extracted from the data repositories discussed in Chapter 1. Since real life data is used as is without any pre-processing, the outcomes of these case studies may not be accurate. Also, improving the accuracy of classification and clustering is not the scope of this part of the work. The performance of classification and clustering algorithms on customer feedback reviews has been analyzed separately on Chapter 4. The output graphs and other models generated in this section are not used by the software model developed. The results from these case studies were published in the IJIRCCE Journal article [88].

Data mining as the name implies is the process of extracting meaningful information from a huge set of data. This section describes how data mining can be used for identifying unknown/unexpected patterns in product and service-based companies. Application of classification and clustering algorithms in complaint category identification, Computer Numerical Control (CNC) machining safety prediction, mechanical fitting failure detection, car price prediction, healthcare infrastructure planning, emission control and management, distribution channel optimization are discussed in this section.

Classification and Clustering algorithms

In this section, we will mainly concentrate on the performance of machine learning algorithms in natural language processing. Sample case studies were made on the algorithms listed below.

1) Naive Bayes classifier.
2) Decision tree classifier.

3) Support Vector Machine classifier.

4) K-nearest Neighbors classifier.

5) Simple K-means clustering.

6) Hierarchical clustering.

7) Farthest_First clustering.

3.5.1 Naive Bayes Classifier

The *Naive Bayes* algorithm is a simple probabilistic classifier that uses Bayes theorem to calculate the probability for each instance by counting the frequency of values in a given dataset. It assumes all attributes i.e., features to be independent given the value of the class label and this classifies it as Naïve.

Bayes Theorem [89]**.**

$P(C|X) = P(X|C) * P(C) / P(X)$

$P(C|X)$ is the posterior probability of a class C for a given attribute.

$P(C)$ is the prior probability of a class.

$P(X|C)$ is the likelihood of the probability of an attribute X for a given class.

$P(X)$ is the prior probability of an attribute.

The *Naive Bayes* algorithm is mainly used in areas where the input is text or image. In the software industry, it can be used for spam detection, image recognition, text mining, and sentimental analysis [90] [91] [92].

Case Study: Complaint category identification

In any sector, incident management is essential. To process and resolve the complaints effectively, the incidents have to be segregated and assigned to the concerned department.

From the complaints lodged by consumers, the product or service that is referred can be identified and assigned to the respective department for further analysis using Naïve Bayes algorithm. *The* bank customer complaint dataset [93], which was used to train Naïve Bayes algorithm had 88 instances. The data was classified based on the attribute "consumer complaint narrative" (see Table 3-1).

Table 3-1 List of attributes and its type

Attribute	Description	Datatype
Consumer complaint narrative	Complaint of the consumer	String
Product	The product which the complaint is about	Nominal

Class attribute: Product

Observations:

Let's take a single instance of the attribute "consumer complaint narrative" and see how *Naive Bayes* classifies and predicts which product the consumer is complaining about.

@data

'The loan amount is incorrect it includes $260000 of fraudulent charges',?

Text: "The loan amount is incorrect it includes $260000 of fraudulent charges".

1) Calculate P(Consumer Loan | The loan amount is incorrect it includes $260000 of fraudulent charges)

   P (Consumer Loan | Text) = P (Text | Consumer Loan) * P (Consumer Loan) / P (Text)

2) Calculate probability of all other class attributes.

   Since P (Text) is the same for the case, it is discarded. In comparison, the class value is assigned to the text for which the probability is higher. If P (Consumer Loan) is higher, that instance is classified as "Consumer Loan".

   To calculate the P (Text), multiply the probability of each word. For example, P(The loan amount is incorrect it includes $260000 of fraudulent charges) = P(The) * P(loan) * P(amount) * P(is) * P(incorrect) * P(it) * P(includes) * P($260000) * P(of) * P(fraudulent) * P(charges). Note that the probability of each word is calculated by the number of occurrences of that word in the whole training data divided by the total number of words.

```
=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:? 3:Consumer Loan        0.106
```

Figure 3-7 Naive Bayes classifier output

3) From the occurrence of each word in the training data, the probability of each word is calculated and applied to the Bayes formula. Based on the probability of each class value, the class value with the highest probability is assigned to the test data.
4) Here, the probability of "Consumer loan" is the highest, so the test data is assigned class value "Consumer loan" as shown in Figure 3-7. Note that to increase the accuracy of classification (see Figure 3-8), we need to tune the training data accordingly.

```
Correctly Classified Instances          86               97.7273 %
Incorrectly Classified Instances         2                2.2727 %
Kappa statistic                          0.9759
Mean absolute error                      0.0942
Root mean squared error                  0.2112
Relative absolute error                 94.784   %
Root relative squared error             94.7847 %
Total Number of Instances               88
```

Figure 3-8 Accuracy of the Naive Bayes classifier

3.5.2 Decision Tree Classifier

Decision Tree algorithms builds a decision tree after analyzing the training instances. In the decision tree, the topmost node is the root node and the leaf nodes are the class values. In general, the nodes are attributes and the branches are decisions. Classifying the test instances using the constructed decision tree starts from the root node and traverses to the leaf node based on the condition at each node and assigns a class value. Decision tree algorithms give better results in areas of research where data is more conditional. Some of its applications include intrusion detection [94] and breast cancer detection [95].

Case Study: CNC machining safety detection

In this case study, a decision tree classifier is used to predict whether the CNC machining operation can be completed safely. The training data [96] taken from previous experiments conducted in the CNC machine had 18 instances, which can be classified based on 5 attributes namely the workpiece material, feed rate, clamp pressure, tool condition and machining finalized described in Table 3-2.

Table 3-2 List of attributes and its type

Attribute	Description	Datatype
Material	Workpiece material	Nominal
Feed rate	The relative velocity of the cutting tool along the workpiece (in mm/s)	Numeric
Clamp pressure	The pressure at which the material is held in the vise	Numeric
Tool condition	A label for worn and unworn tools	Nominal
Machining finalized	Indicates if machining was completed safely	Nominal

Class attribute: Machining finalized

Observations:

A decision tree is built based on the input dataset's range of values for each attribute. All the leaves are class attribute values, i.e., yes or no, each node is an attribute and each branch hold a condition. Figure 3-9 shows that not all the attributes are covered in the tree, and only the attributes which contribute to the class classification are framed as nodes. This helps in conditionally classifying the test data. The algorithm checks all the attribute values of the test data and based on the values, it flows through the branches in the tree and finally end up in a class attribute value and assigns that class for the test data.

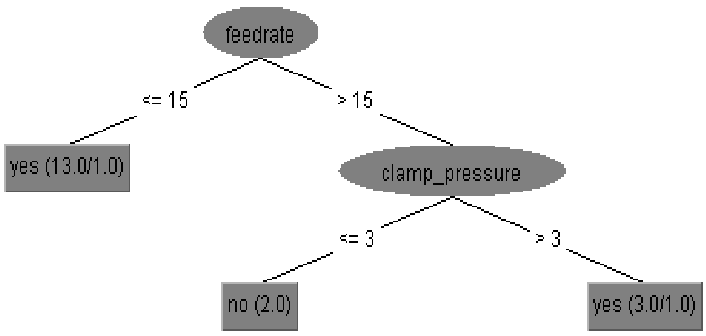

Figure 3-9 Tree generated by Decision Tree Classifier

Let’s take a sample test data with the following values,

@data

wax,20,4,unworn,?

The decision tree algorithm starts classifying the test data from the root of the tree. In Figure 3-9 since the value of the feed rate is greater than 15 it goes to the right branch. Next, the value of clamp pressure is greater than 3, so it goes to the right branch. Now it reached the class attribute value “yes”. So, from the test data, it can be predicted that the machining can be completed safely (i.e. Yes) as shown in Figure 3-10. The accuracy of the decision tree classifier was 88% as shown in Figure 3-11.

```
== Predictions on user test set

  inst#     actual  predicted
      1        1:?      1:yes
```

Figure 3-10 Decision Tree classifier output

```
Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances          16               88.8889 %
Incorrectly Classified Instances         2               11.1111 %
Kappa statistic                          0.6087
Mean absolute error                      0.1766
Root mean squared error                  0.2972
Relative absolute error                 48.9152 %
Root relative squared error             71.3245 %
Total Number of Instances               18
```

Figure 3-11 Accuracy of Decision Tree Classifier

3.5.3 Support Vector Machine Classifier

This algorithm plots each attribute's value in an n-dimensional space (n being the number of independent attributes). After plotting, a hyperplane is defined which divides the plotted data into the required number of groups (number of classes). The test data is, it is classified according to the part of the hyperplane it is in. *Support Vector Machine* algorithms work well for cases with two class values. It divides the two class values with a hyperplane. The right hyperplane is chosen considering the key points such as

1) The hyperplane should divide the two classes equally.
2) The hyperplane should be with the highest margin, where the margin is the distance between the hyperplane and the nearest point.

Case Study: Mechanical fitting failure detection

In this case study, a *Support Vector Machine* classifier is used to analyze the historical data of fitting failures. The analyzed data can be used in risk assessment while planning new pipeline construction and to identify the cause of fitting failures in existing pipelines. The Dataset [97] which was used to train *Support Vector Machine* algorithm had 120 instances with nine attributes listed in Table 3-3.

Table 3-3 List of attributes and its type

Attribute	Description	Datatype
Leak location a text	Indicates whether the leakage is above or below ground	Nominal
Leak location b text	Indicates whether the leakage is inside or outside	Nominal
Leak location c text	Indicates the type of connection: main to service, service to service, etc.	Nominal
Manufacturer	Manufacturer name	Nominal
Model number	Model number	Nominal
Fitting material text	Material used to connect one pipe to another: in this case, to connect the two pipes.	Nominal
First pipe material text	The material of the first pipe	Nominal
Second pipe material text	The material of the second pipe	Nominal
Leak cause text	Cause for the leakage	Nominal

Class attribute: Leak cause text

Observations:

Let’s take a sample test data with the following values:

@data

ABOVEGROUND,INSIDE,MAIN-TO-

SERVICE,PERFECTION,75313,STEEL,PLASTIC,PLASTIC,?

```
=== Predictions on user test set ===

    inst#     actual  predicted error prediction
        1        1:? 1:NATURAL FORCES        0.267
```

Figure 3-12 Support Vector Machine classifier output

1) From the classifier output in Figure 3-12, the cause of the leak is identified as "Natural Forces". The accuracy of the algorithm was more than 99% as shown in Figure 3-13.

```
Time taken to test model on training data: 0.1 seconds

=== Summary ===

Correctly Classified Instances         119               99.1667 %
Incorrectly Classified Instances         1                0.8333 %
Kappa statistic                          0.99
Mean absolute error                      0.2226
Root mean squared error                  0.3102
Relative absolute error                 80.1333 %
Root relative squared error             83.2239 %
Total Number of Instances              120
```

Figure 3-13 Accuracy of Support Vector Machine algorithm

3.5.4 K-nearest Neighbors Classifier

This algorithm puts the K-nearest Neighbors to a data point in a class. The value of k is given by the user. The choice of k value is crucial for this algorithm. It calculates the Euclidean distance of a test data point to all other points and takes the top k least values and assigns them a class that appears frequent.

Case Study: Car price prediction

We can use a K-nearest Neighbor algorithm to predict the cost of a car using its features as listed in Table 3-4. Note that the term attributes can refer to the features of a product in the manufacturing industry. The Dataset [98] which was used to test *K-nearest Neighbors* classifier had 199 instances with 16 car specification features as shown in Table 3-4.

Table 3-4 List of attributes and its type

Attributes	Description	Datatype
Make	Car manufacturer	Nominal
Fuel type	Gas or diesel	Nominal
Aspiration	Turbo or standard	Nominal
Num of doors	Number of doors	Nominal
Body style	Generic shape of the car	Nominal
Drive wheels	Indicates the axles to which the engine power is transmitted through the driveshaft	Nominal
Engine location	Location of the engine with respect to the passenger cabin	Nominal
Wheelbase	The distance between the center of the front and rear wheels	Numeric
Length	The total length of the vehicle	Numeric
Width	The total width of the vehicle	Numeric
Height	The total height of the car	Numeric
Curb weight	The weight of the vehicle without occupant or baggage	Numeric
Engine type	Indicates whether the engine has a single camshaft mechanism or dual camshaft mechanism to operate the intake and exhaust valves	Nominal
Num of cylinders	Number of cylinders in the engine	Nominal
Engine size	Displacement in cubic inch	Numeric
Price	The market cost of the vehicle	Numeric

Class attribute: Price

Observations:

Let's take a sample test data with the following values:

@data

audi,gas,turbo,two,convertible,rwd,front,88.6,168.8,64.1,48.8,2548,dohc,four,130,?

honda,gas,turbo,four,sedan,fwd,front,99.8,176.6,66.2,54.3,2337,ohc,four,109,?

```
=== Predictions on user test set ===

    inst#     actual  predicted      error
        1          ?  14997.5            ?
        2          ?   8845              ?
```

Figure 3-14 K-nearest Neighbors classifier output

1) From the classifier output in Figure 3-14, the price of the Audi car is predicted to be \$14997.5 and the price of the Honda car is predicted to be \$8845.
2) So, this model can be used to compare the market values of n number of brands using n number of features and fix the appropriate price for a car.
3) Since mean absolute error is very high (see Figure 3-15), we can reject this algorithm for our model, but a comprehensive analysis to select the algorithms for analyzing customer reviews is discussed in Chapter 4.

```
Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correlation coefficient                  0.9993
Mean absolute error                     80.5377
Root mean squared error                287.0704
Relative absolute error                  1.3712 %
Root relative squared error              3.6072 %
Total Number of Instances              199
```

Figure 3-15 Accuracy of K-nearest Neighbors classifier

3.5.5 Simple K-means Clustering

This algorithm assigns k random centroids initially. Each data point is taken and associated with the nearest centroid. In each iteration, a new set of k centroids are assigned calculated from the barycenter of the previous set. The barycenter is calculated by the Euclidean distance of each data point to every other point. When it reaches a stage in which no more change is required in centroids, the iteration is stopped and the points are allocated to the nearest centroid. This type of algorithm is used when the number of clusters is known in the first place. In case of a Network company wanting to install k towers across the country with each tower to be placed as far from each other as possible and also consider the density of users in each area, this algorithm is the best choice.

Case Study: Healthcare infrastructure planning

Suppose the US government wants to establish cardiovascular facilities; Simple K-means can be used for identifying the optimal locations. When the heart disease mortality data with the location data is given to Simple K-means, it assigns four centroids initially and iterates till four optimum locations are found. The dataset downloaded from Healthcare.gov [99] had six attributes as shown in Table 3-5.

Table 3-5 List of attributes and its type

Attribute	Description	Datatype
LocationAbbr	US state	Nominal
Data_Value	Mortality rate	Numeric
Data_Value_Type	Age group	Nominal
Stratification1	Gender	Nominal
Stratification2	Race	Nominal
Location 1	Latitude, Longitude	Numeric

```
Class attribute: LocationAbbr
Classes to Clusters:

   0    1    2    3  <-- assigned to cluster
 201  165   78   60 | AK
 275  332  285  170 | CA
 484  492  105  143 | AL
 533  574  101  160 | AR
  11   12    2   11 | AS
  85   78   83   42 | AZ
 323  278  349  220 | CO
  30   38   53   41 | CT
   8   13    8    7 | DC
  16   19   15   22 | DE
 153  399  198  474 | FL
   3    1    1    0 | GU
 299 1119  274 1157 | GA

Cluster 0 <-- AL
Cluster 1 <-- AR
Cluster 2 <-- CO
Cluster 3 <-- GA
```

Figure 3-16 Four centroids assigned by Simple K-means algorithm

Observations:

1) Simple K-means was used to form four clusters based on the heart disease mortality data. Four centroids are assigned to regions where the density of data points is more. AL, AR, CO, and GA are the centroids of the four clusters formed (see Figure 3-16). Visualization of the data points for each state is shown in Figure 3-17. The ability to form clusters and their respective centroids may vary depending on the quality of the data used. So proper pre-processing techniques must be used to clean the data. Pre-processing techniques used in the software model are described in Chapter 5.

2) On analyzing the cardiovascular mortality data from 2013 to 2016 using the K-means algorithm, we infer that preference should be given for establishing cardiovascular facilities in Alabama, Arkansas, Colorado, and Georgia.

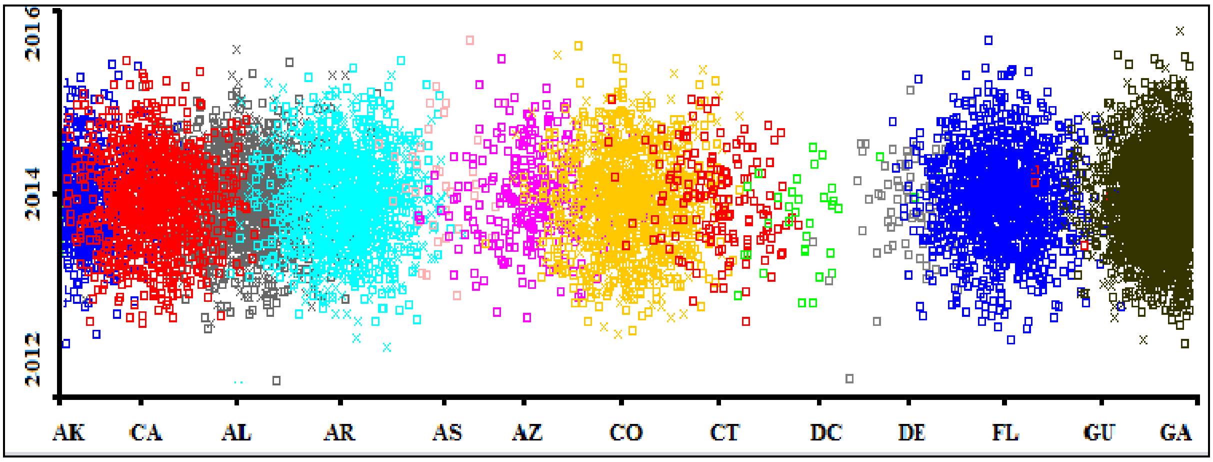

Figure 3-17 Visualization of data points for each State using Simple K-means

3.5.6 Hierarchical Clustering

On a set of N data items to be clustered, this algorithm assigns each data item to a cluster and finds the distance between each cluster. Then it merges the two closest clusters and repeats till a single cluster is formed. This type of algorithm can be used in the healthcare domain to find the dissimilarities between the similarities. One such application is to find the sub-types of breast tumor in independent genes.

There are different ways of calculating the distance between each cluster and they are listed below:

1) Single linkage: the distance between two clusters is the shortest distance between the points in the cluster.
2) Complete linkage: the distance between two clusters is the farthest distance between the points in the cluster.
3) Average linkage: the distance between two clusters is the average distance between the points in one cluster and the points in the other cluster [100].

Case Study: Emission control and management

In this case study, we have used the particulate emission data [101] to cluster all the provinces and territories in Canada using a *Hierarchical* clustering algorithm. The same approach can be used to cluster similar air zones and air sheds [102]. To efficiently manage the local air quality, Canadian Ambient Air Quality Standards (CAAQS) can be made specific to these air zones and air sheds. The dataset had five attributes as shown in Table 3-6.

Table 3-6 List of attributes and its type

Attribute	Description	Datatype
Province or territory	Canadian provinces	Nominal
1990 (emissions in tones per sq. km)	Emissions in tones per sq. km in 1990	Numeric
2015 (emissions in tones per sq. km)	Emissions in tones per sq. km in 2015	Numeric
1990 excluding open sources (emissions in tones per sq. km)	Emissions in tones excluding open sources per sq. km in 1990	Numeric
2015 excluding open sources (emissions in tones per sq. km)	Emissions in tones excluding open sources per sq. km in 2015	Numeric

```
Class attribute: Province or territory
Classes to Clusters:

 0 1  <-- assigned to cluster
 1 0 | Newfoundland and Labrador
 0 1 | Prince Edward Island
 1 0 | Nova Scotia
 1 0 | New Brunswick
 0 1 | Quebec
 0 1 | Ontario
 1 0 | Manitoba
 1 0 | Saskatchewan
 1 0 | Alberta
 1 0 | British Columbia
 1 0 | Yukon
 1 0 | Northwest Territories and Nunavut

Cluster 0 <-- Newfoundland and Labrador
Cluster 1 <-- Prince Edward Island
```

Figure 3-18 Two centroids assigned by Hierarchical clustering

Observations:

1) *Hierarchical* clustering algorithm was used with complete linkage configuration to form two clusters. In Figure 3-18, "1" means yes, the province is allocated to that cluster.
2) The provinces of Newfoundland and Labrador, Nova Scotia, New Brunswick, Manitoba, Saskatchewan, Alberta, British Columbia, Yukon, Northwest Territories, and Nunavut have very less pollution per square kilometer and fall under cluster 0.
3) The provinces of Prince Edward Island, Quebec, and Ontario have the highest air pollution rate per square kilometer and fall under cluster 1.

3.5.7 Farthest_First Clustering

The Farthest_First clustering algorithm places each cluster center at the point farthest from the existing cluster center in the next iteration by calculating the Euclidean distance. When all the clusters points are far from each other, the algorithm stops and returns the center. In this algorithm also the number of clusters can also be configured.

Case Study: Distribution channel optimization

Suppose if a Canadian e-commerce distribution company wants to build three warehouses for storing and distributing imported goods from China to various locations within Canada; the *Farthest_First* clustering algorithm can be used to identify the locations for the warehouses and optimize their distribution channels.

Input dataset: Major importers by country 2016, Canadian government open data portal [103].

Number of clusters: 3

Number of attributes: 5

Table 3-7 List of attributes and its type

Attribute	Description	Datatype
Country	China	Nominal
Company-Enterprise	Company address	Nominal
City-Ville	City	Nominal
Province_Eng	Province	Nominal
Postal_Code-Code_Postal	ZIP code	Nominal

Observations:

1) Centroids are assigned to regions where the density of data points are more. From Figure 3-20, we can infer that the cluster density (i.e. number of data points) are more in QC, ON, and BC. So, building the three warehouses at the centroids of each cluster formed (see Figure 3-19) i.e. British Columbia, Quebec and Ontario would be optimal
2) Note that the above analysis is for the understanding of readers; the locations may be changed based on the historical data used. So accurate data for at least the past ten years is recommended.

```
Class attribute: PROVINCE_ENG
Classes to Clusters:

   0   1   2  <-- assigned to cluster
   0 375   0 | Quebec
   0   0 753 | Ontario
  73   0   0 | Alberta
 174   0   0 | British Columbia
   3   0   0 | Newfoundland and Labrador
  21   0   0 | Manitoba
   6   0   0 | Saskatchewan
   9   0   0 | New Brunswick
  16   0   0 | Nova Scotia
   1   0   0 | Prince Edward Island

Cluster 0 <-- British Columbia
Cluster 1 <-- Quebec
Cluster 2 <-- Ontario
```

Figure 3-19 Three centroids assigned by Farthest_First algorithm

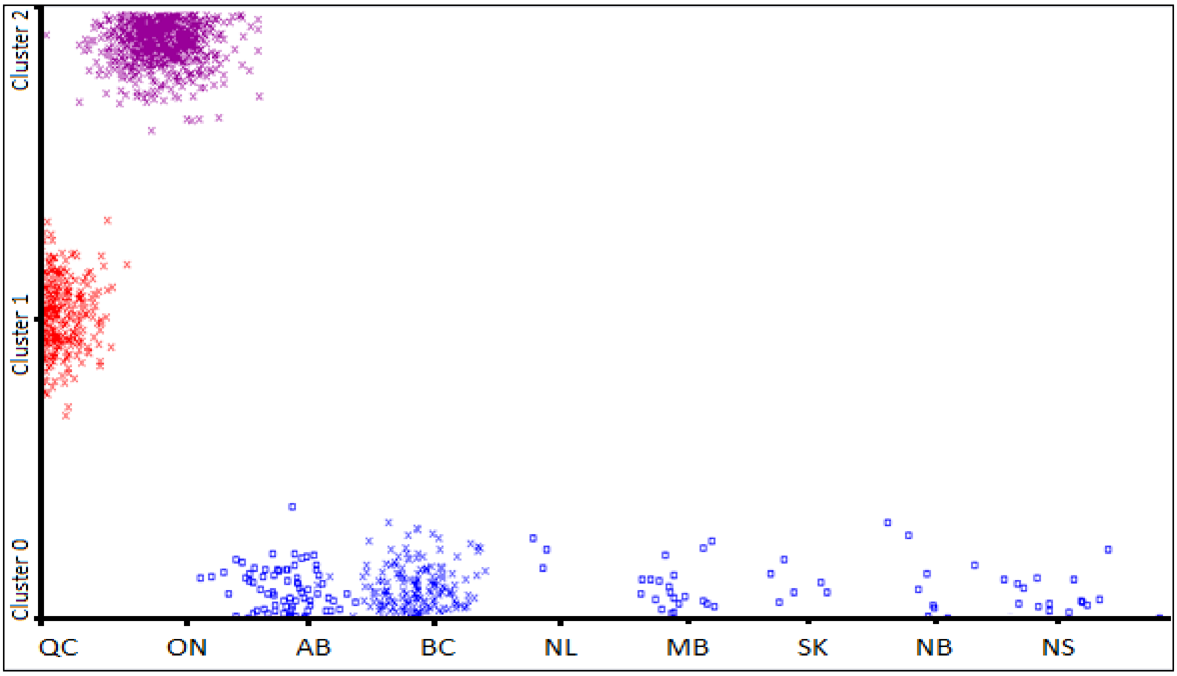

Figure 3-20 Farthest_First cluster assignments visualization

CHAPTER 4. DATA MINING ALGORITHMS REVIEW

The algorithms discussed in the previous chapter will result in different classification quality depending on the match to the dataset used. In the comparative analysis of classification algorithms for a diabetes dataset [104] and Swahili language Tweets [105] *Naive Bayes* algorithm was more accurate and suitable while the performance of J48 was better in an analysis done on a bank dataset [106] and credit card fraud detection [7]. In this chapter, the experimental procedures performed to select the classification [107] [108] [109] and clustering [110] algorithms are discussed. The results from this chapter were accepted for publication by the AIACT'19 conference (see Preface for list of proceedings).

The selected algorithms were used in the real-life implementation of the software model on Durabuilt Windows and Doors Inc., a leading windows and doors manufacturing company in Northern Alberta; and Kathir Food Experience Inc., a famous South Indian Restaurant in Edmonton.

4.1 Classification Algorithms

In this section, two classification algorithms namely *J48* and *Naive Bayes* which alternatively work better in different fields (as discussed in Chapter 3) are analyzed to demonstrate which one fits the best for customer reviews data from Facebook.

J48 Algorithm. The *J48* decision tree algorithm builds a decision tree after analyzing the training instances. In the decision tree, the topmost node is the root node and the leaf nodes are the class values. In general, the nodes are attributes and the branches are decisions. While classifying the test instances using the constructed decision tree, it starts from the root node and traverses to the leaf node based on the condition at each node and assigns a class value.

Naive Bayes Algorithm. The *Naive Bayes* algorithm is a simple probabilistic classifier that uses Bayes theorem to calculate the probability for each instance by counting the frequency and combinations of values in a given dataset. It assumes all attributes to be independent given the value of the class variable [111].

4.1.1 Performance Factors

This section includes the performance factors that were used in the experimental analysis of classification algorithms [112].

Confusion matrix

The Confusion matrix in Table 4-1 can be used to identify checking statuses of correlations between the predicted results and the actual judgments as shown below:

True Positive (TP): When the prediction is YES and the actual value is also YES

True Negative (TN): When the prediction is NO and the actual value is also NO

False Negative (FN): When the prediction is NO but the actual value is YES

False Positive (FP): When the prediction is YES but the actual value is NO

Table 4-1 Confusion matrix

	Predicted Yes	**Predicted No**
Actual Yes	TP	FN
Actual No	FP	TN

Correctly Classified Instances

Out of the total instances, correctly classified instances are the number of instances whose predicted and actual class values are equal, i.e., the sum of TN and TP.

$$TN + TP \tag{1}$$

Incorrectly Classified Instances

Out of the total instances, incorrectly classified instances are the number of instances whose predicted and actual class values are different, i.e., the sum of FN and FP.

$$FN + FP \tag{2}$$

Precision

$$\frac{TP}{TP+FP} \tag{3}$$

Precision gives what fraction of the predicted positive values are actually positive.

Recall

$$\frac{TP}{TP+FN} \tag{4}$$

Recall gives what fraction of actually positive values is predicted positive.

F-Measure

$$2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{5}$$

F-Measure is the weighted average of precision and recall.

MCC

$$\frac{TP * TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{6}$$

Matthews's Correlation Coefficient (MCC) is a factor to measure the quality of prediction. Its value ranges from -1 to 1.

Note that

-1 denotes total deviation from prediction and actual

1 denotes correct prediction.

ROC area

Receiver Operating Characteristic (ROC) curve is a graph plot of the true positive rate against the false positive rate for each class label. Accuracy is measured by the area under the curve. Note that the area range and its implication are given below:

.90 - 1 = excellent (A)

.80 - .90 = good (B)

.70 - .80 = fair (C)

.60 - .70 = poor (D)

.50 - .60 = fail (F)

4.1.2 Experimental Analysis

Experiments are performed on customer reviews data by using classification algorithms in the WEKA tool. The different performance factors for positive and negative customer reviews using *J48* and *Naive Bayes* algorithms are in Table 4-2. Total number of instances is 2000.

Table 4-2 Performance factors for positive and negative using J48 and Naive Bayes

	J48		**Naive Bayes**	
	Positive	Negative	Positive	Negative
TP Rate	0.962	0.959	0.730	0.705
FP Rate	0.041	0.038	0.295	0.270
Precision	0.959	0.962	0.712	0.723
Recall	0.962	0.959	0.730	0.705
F-Measure	0.961	0.960	0.721	0.714
MCC	0.921	0.921	0.435	0.435
ROC Area	0.990	0.990	0.808	0.807

Figure 4-1 and Figure 4-2 are graphical representations of the various performance factors of classification algorithms in this case.

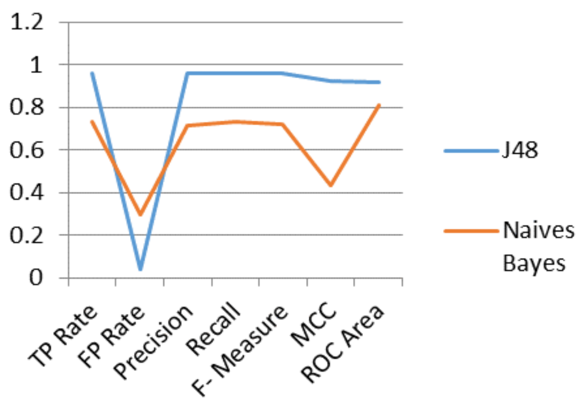

Figure 4-1. Performance factors for positive comments between J48 and Naive Bayes

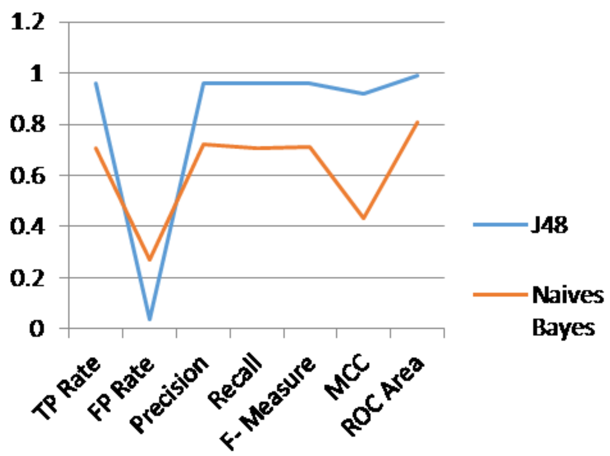

Figure 4-2. Performance factors for negative comments between J48 and Naive Bayes

After testing the classification algorithms over customer reviews data, the following were inferred:

1) From Table 4-2: *J48* performs better than the *Naive Bayes* algorithm.
2) From Table 4-3: though *Naive Bayes* equally takes less time to build a model, *J48* classifies most of the instances correctly (see Figure 4-3) and increases the accuracy. Hence, the *J48* algorithm is used in our model.

Table 4-3 Comparison based on accuracy and time

Algorithm	Correctly classified instances	Percentage of Correctly classified instances	Incorrectly classified instances	Percentage of incorrectly classified instances	Time taken to build a model (in seconds)
J48	1921	96.05	79	3.95	0.13
Naive Bayes	1475	71.75	565	28.25	0.85

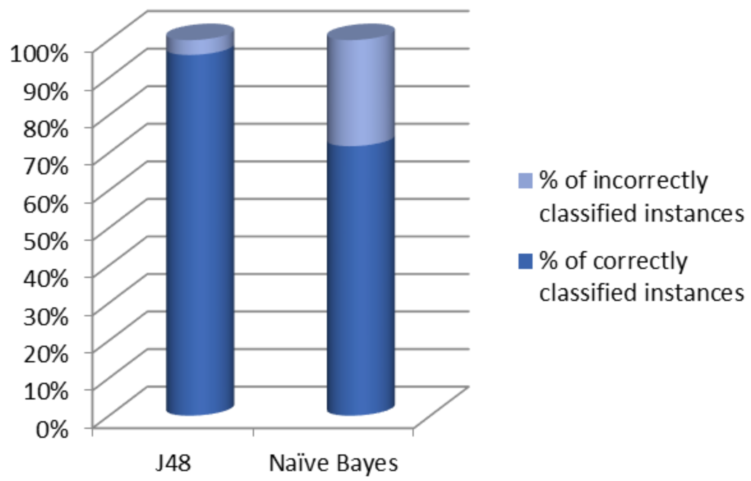

Figure 4-3 Graph of accuracy

4.2 Clustering Algorithms

In this section, *Simple K-means, Hierarchical, Farthest_First, and Make_Density_Based* clustering algorithms [113] are analyzed to demonstrate which one fits the best for customer reviews data.

Simple K-means Clustering Algorithm. It is an iterative clustering algorithm in which items are moved among clusters till all instances are covered and the desired cluster is reached.

Hierarchical Clustering Algorithm. Finds the similarity or dissimilarity between every pair of objects in the data set by calculating the distance between objects. It then groups the objects into a binary, hierarchical cluster tree. In the next step, it determines where to cut the hierarchical tree into clusters by pruning branches off the bottom of the hierarchical tree so as to partition data.

Farthest_First Clustering Algorithm. It is a variant of *Simple K-means*. This places the cluster center at the point further from the present cluster. The points that are farther are clustered together first. This feature of the *Farthest_First* clustering algorithm speeds up the clustering process in many situations such as when less reassignment and adjustment are needed.

Make_Density_Based Clustering Algorithm. A cluster is a dense region of points that separates low-density regions from tightly dense regions. This clustering algorithm can be used when the clusters are irregular. The *Make_Density_Based* Clustering algorithm can also be used in noise and when outliers are encountered. The points with the same density and present within the same area will be connected to form clusters.

4.2.1 Experimental Analysis

Experiments are performed on customer reviews data by using clustering algorithms in the WEKA tool. It has been constrained to test algorithm with a total number of 2000 review instances.

Table 4-4 Time taken (in seconds) to form Clusters

Number of Clusters	**Simple K-means**	**Hierarchical**	**Farthest First**	**Make Density Based**
10	.221	.363	.224	.23
20	.225	.356	.227	.237
50	.227	.344	.234	.244

From the experimental results of clustering algorithms over customer reviews data as shown in Table 4-4, we inferred the following:

1) As the cluster size increased, the time taken for clustering in *Make_Density_Based* and *Farthest_First* clustering algorithm increased.
2) The *Hierarchical* clustering algorithm took more time when the number of clusters was less.
3) In all cases, *Simple K-means* remains consistent. So, *Simple K-means* clustering algorithm is used in our software application.

CHAPTER 5. PROCESS MODEL AND IMPLEMENTATION

In this chapter, the framework is discussed in detail along with its real-life implementation at Kathir Food Experience Inc., a famous South Indian Restaurant in Edmonton. The results from this chapter were published in ICCBD'18 [114] and IC4E'19 [115] conferences.

The data collected from Facebook have been loaded into WEKA [116] for sentimental analysis [117] [118]. Sentiment analysis was performed since lots of data were available to learn the feel of customers review about the issues, activities, and features of the two different organizations under study. To do that, customer reviews were first analyzed through the steps of word parsing and tokenization (extract terms), stop-words removal (elimination of non-value adding words for sentiment analysis), lemmatization and stemming (convert all infections to their root word), and term selection/feature extraction (remove the terms that have poor prediction ability). Then the cleaned customer feedback texts were classified by using the *J48* algorithm [119] in two iterations (1) to determine which product or service category the review is talking about and (2) to find the sentiment of the customer review (positive or negative). The next step in data mining is to design the clustering model. The dataset was categorized into three by using *Simple K-means* clustering [120] based on the sentiment of the comments as satisfactory, not satisfactory, and neither satisfactory nor unsatisfactory. Features that are categorized under not satisfactory are inferred to be looked into by the management. The flowchart of the proposed feedback analysis model is shown in Figure 5-1.

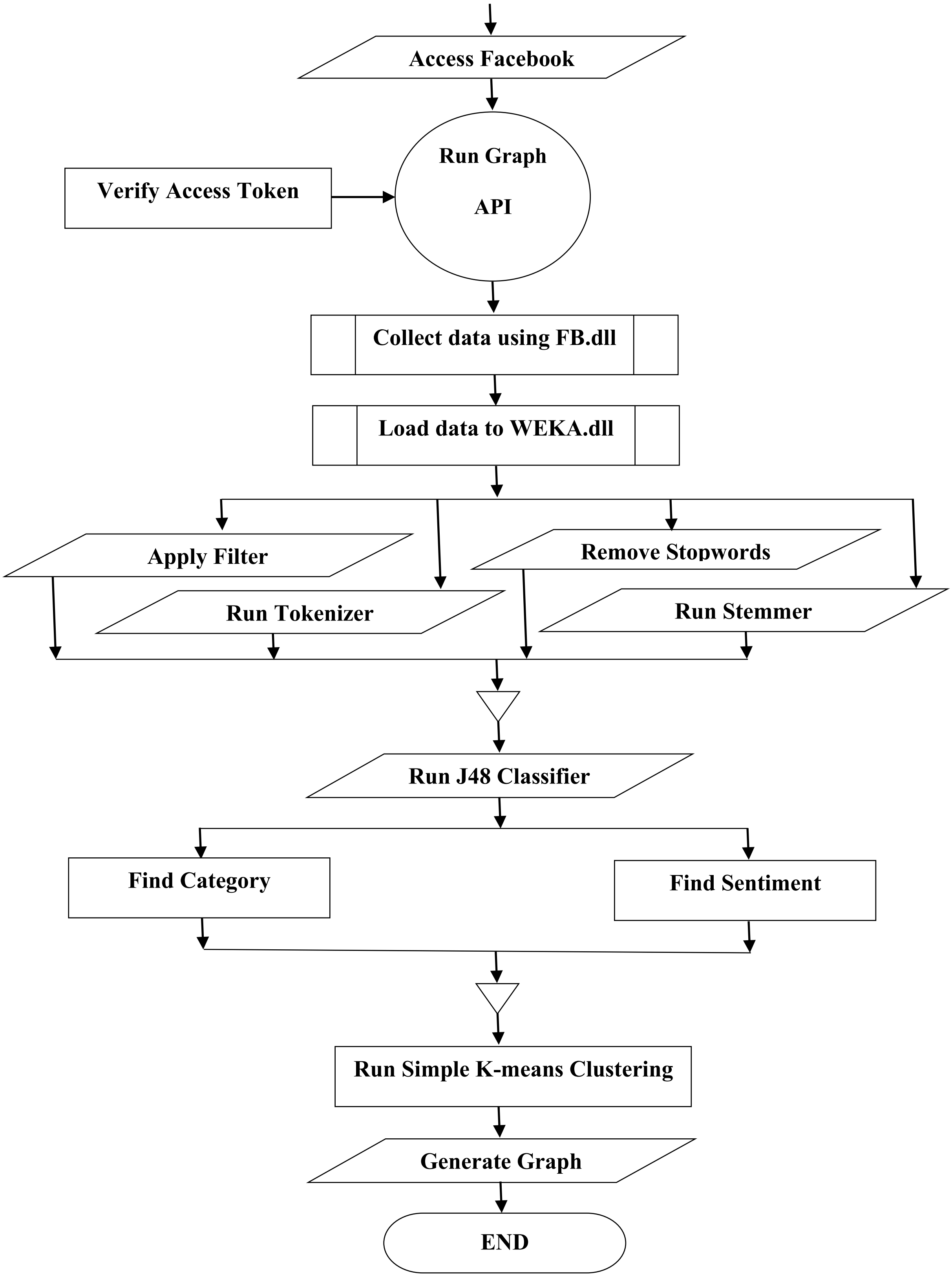

Figure 5-1 Flowchart of the proposed feedback analysis model

Our proposed model has three major modules;

1) A data collection module which collects data from Facebook;
2) A data preprocessing module which filters and formats the data;
3) A data classifying and clustering module which is the main module that classifies and clusters data using *J48* [121] and *Simple K-means* [122] algorithms, respectively.

5.1 Data Collection Module

Data is collected from Facebook using the Facebook Graph API [123]; the C#.Net code for this is shown in the Appendix-A Class-1. Modules within the data collection module are explained in brief below.

5.1.1 Facebook Graph API

Facebook provides an option for its authorized users to extract data from a Facebook account and from the pages they own using Facebook Graph API. There are many software packages [124] available for importing data from Facebook but in this paper, we have directly pulled data from the Facebook page through the Facebook Graph API using Access Token [125]. It allows querying data, posting new stories, managing ads, uploading photos, and performing a wide variety of other tasks. For the C#.NET application to communicate with Facebook Graph API, Facebook.dll has to be imported. Once it is imported, the application can programmatically query data from any Facebook account using an authorized access token taken from developers.Facebook.com. Steps are explained in detail in the upcoming section.

5.1.1.1 Facebook developers account

For any application to access Facebook data, it needs a developer app created using an authorized Facebook login.

The steps to create a developer app and get app id and secret are given below:

Step 1: Go to developers.Facebook.com and create an account.

Step 2: Go to developers.Facebook.com/tools/explorer.

Step 3: Go to "My apps," drop down in the top right corner and select "Add a new app." Provide a display name and category and then "Create App ID."

Step 4: Go to developers.Facebook.com/tools/explorer. Check for "Graph API Explorer" below "My Apps." From "Graph API Explorer" drop-down and select your app.

Step 5: Then, select "Get Token." From this drop-down, select "Get User Access Token." Select permissions from the menu that appears and then select "Get Access Token."

5.1.1.2 Access Token

The developer app has an app id, secret, and access token that provide temporary, secure access to Facebook APIs. An access token is an opaque string that identifies a user, app, or page and can be used by the app to make graph API calls. The access token provides permission to APIs that read, write, or modify the data belonging to a Facebook page. Sample data extracted from Facebook is shown in Figure 5-2.

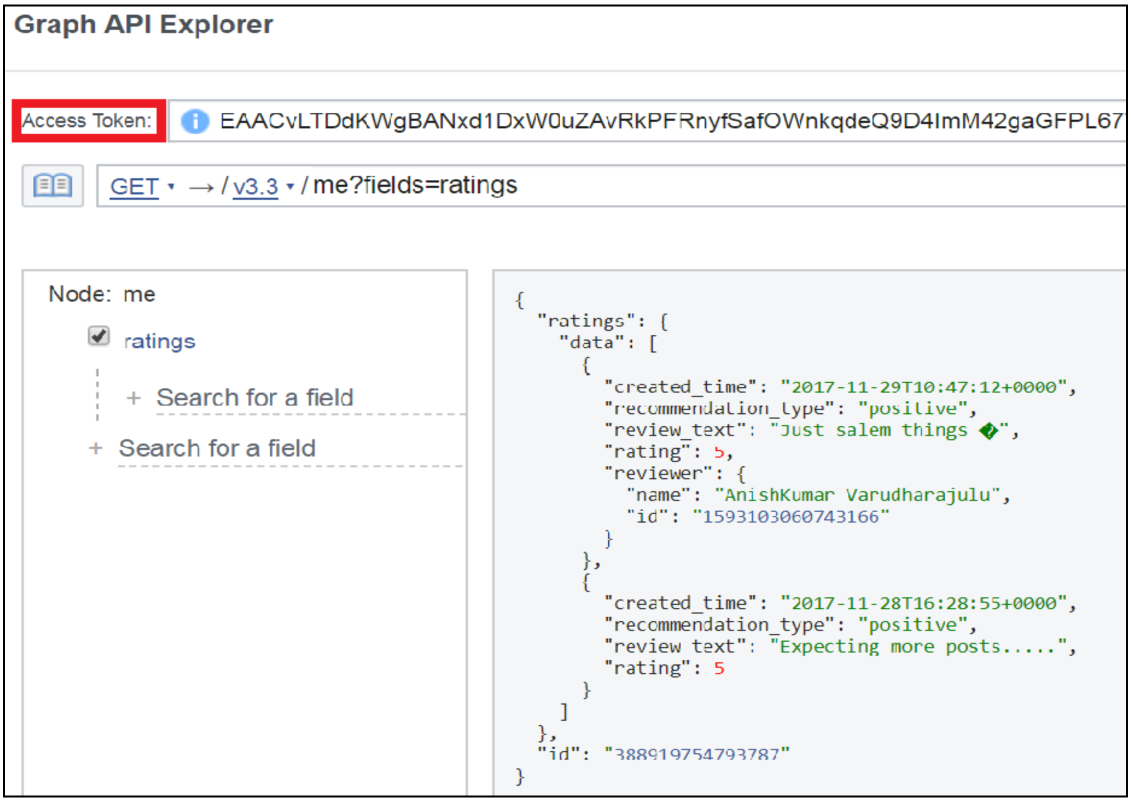

Figure 5-2 Graph API showing extracted data from the Facebook page

5.1.2 Codes and Datasets

The data to be analyzed is collected from the organization's Facebook page. Customers write reviews and rate the organization on their Facebook page. These reviews are collected by connecting to Facebook API using JSON (JavaScript Object Notation).

5.1.2.1 The Code of Extracting Data from Facebook Graph API

Node. A node represents an object. An object here can be a user or a page with a unique id. To access the details of the user node, the below JSON request URL can be used:

```
https://graph.Facebook.com/{your-user-id}?fields=id,name&access_token={your-user-accesstoken}
```

Figure 5-3 URL to access user detail

The JSON request URL in Figure 5-3 will give the output response as shown in Figure 5-4:

```
{
"name": "Your Name",
"id": "your-user-id"
}
```

Figure 5-4 User details extracted from Facebook

Edge. Each node has edges. Edges are also a type of node that are connected to a node. To retrieve an edge, the edge name has to be specified in the JSON request URL. For example, a "user" node has a "feed" edge which returns all the posts of that user. To access all posts of a user, the below JSON request URL can be used:

```
https://graph.Facebook.com/{your-user-id}/feed?access_token={your-user-access-token}
```

Figure 5-5 URL to access all posts of a user

The JSON request URL in Figure 5-5 will give the output response as shown in Figure 5-6.


```
{
 "data": [
  {
   "created_time": "2017-12-08T01:08:57+0000",
   "message": "Love this puzzle. One of my favorite puzzles",
   "id": "post-id"
  },
  {
   "created_time": "2017-12-07T20:06:14+0000",
   "message": "You need to add grape as a flavor.",
   "id": "post-id"
  }
 ]
}
```


Figure 5-6 User feed extracted from Facebook

Fields. Fields represent the properties of a node. When a node is queried using JSON, it responds with a set of fields by default. We can customize the JSON query to return only a specified list of fields, each with a unique id. The below JSON request URL can be used to access specific fields like birthday, email, and hometown.

```
https://graph.Facebook.com/{your-userid}?fields=birthday,email,hometown&access_token
={your-user-access-token}
```

Figure 5-7 URL to access fields

The JSON request URL in Figure 5-7 will give the output response as shown in Figure 5-8.


```
{
  "birthday": "01/01/1985",
  "email": "your-email@email.addresss.com",
  "hometown": "Your, Hometown",
  "id": "{your-user-id}"
}
```


Figure 5-8 Fields extracted from Facebook

As shown in Figure 5-9, the developed software prototype, uses the JSON request URL to collect data. The *fields* collected are 1) review_text; 2) reviewer; 3) rating; and 4) created time (see Figure 5-10). Here, ratings is the *edge* and the organization's Facebook page is the *node*.

```
https://graph.Facebook.com/{page-
id}/ratings?fields=review_text,reviewer,rating,created_time&access_token={access-token}
```

Figure 5-9 URL used to access review and ratings

```
{
    “rating”: {
        “data”:[
                {
                “created time”: “2018-01-25T00:11:02+0000”
                “reviewer”:
                        {
                        "name": "xxx",
                        "id": "123"
                        }
                “rating” = 5,
                “review-text”: “Great Service”
                 }
                ]
               }
}
```


Figure 5-10 Sample data extracted from the Facebook page

5.1.2.2 The Code of Exporting Data

The JSON data retrieved using the above URL is converted and stored as a “.arff” file using C#.NET as shown in Appendix-A Class-1.

5.1.3 Service Feature Selection

The two required inputs for the process model are reviews and rating which are collected from the previous step; they are framed into an “.arff” file (see Figure 5-11) that is intelligible by WEKA for further processing. The list of attributes used here are as follows:

1) Comment

2) Rating

3) Category (Class attribute in iteration 1)

4) Sentiment (Class attribute in iteration 2)

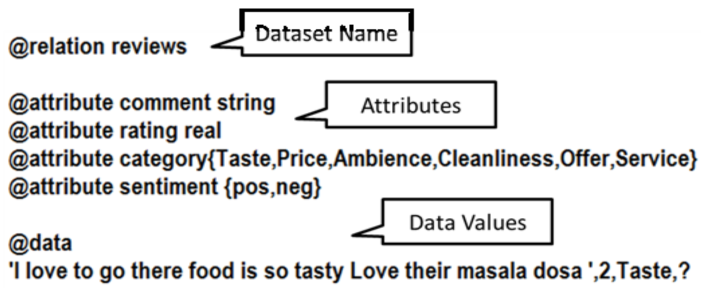

Figure 5-11 Sample snapshot from a ".arff" file

5.1.4 Class

Among the attribute list, one would be the target/class variable based on which the classifier classifies input in each iteration. Here the data is first classified based on category and then based on sentiment. So, *Category* is the class attribute in iteration 1 and *Sentiment* is the class attribute in iteration 2.

5.1.4.1 Class Diagram

The Class Diagram in Figure 5-12 is an illustration of all the classes used in the project. It describes the attributes and operations of each class and their visibility. It also shows the relationship and dependency among them.

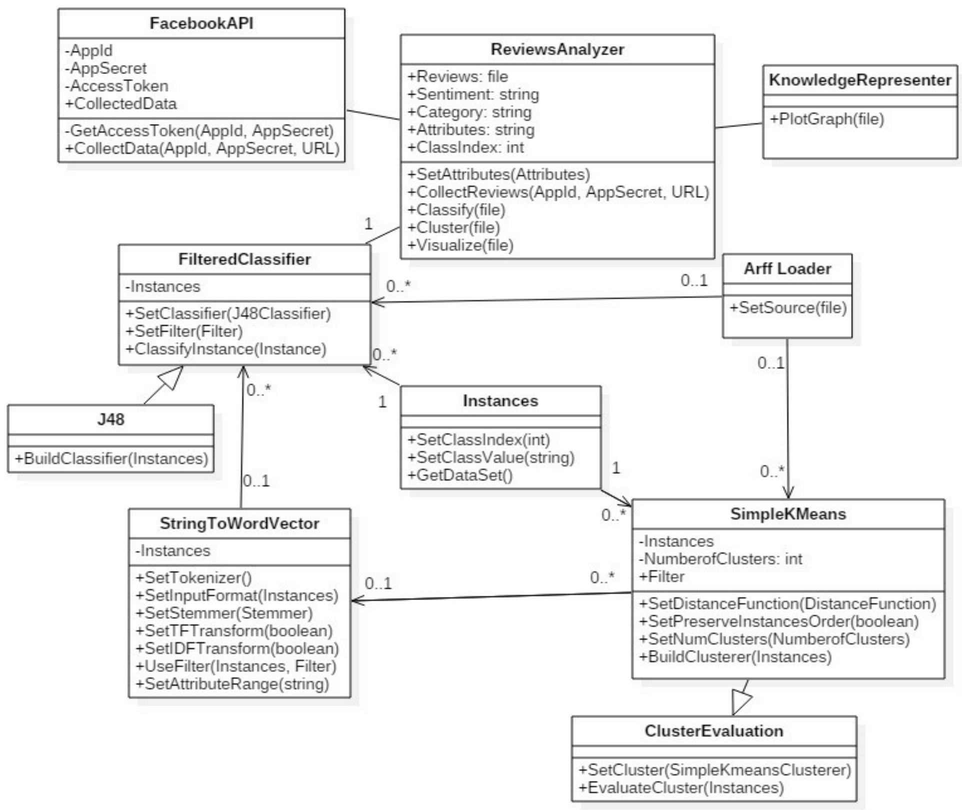

Figure 5-12 Class diagram for the process model

5.1.4.2 Visibility

Class visibility depicts the scope of the attribute/operation. Table 5-1 explains the various class accessibility levels.

Table 5-1 Class visibility

Public	Can be accessed by other classes	+
Protected	Can be accessed only by elements that have a generalization relationship	#
Private	Can be accessed only within that class	-

In Facebook API, AppId and AppSecret are declared private because those are confidential and irrelevant data to expose to other classes. Likewise, in classes *StringToWordVector*, *SimpleKMeans*, and *FilteredClassifier*, the attribute instances are declared private.

5.1.4.3 Generalization

This association an “is a” or inheritance relationship between classes. A subclass inherits all the properties of the base class. Here, *FilteredClassifier* and *J48* have generalization association. *FileteredClassifier* is the base class and *J48* is the subclass. *J48* inherits all the properties of *FilteredClassifier*. In other words, *J48* is a kind of *FilteredClassifier*. Likewise, in *SimpleKmeans* and *ClusterEvaluation*, *ClusterEvaluation* is the base class and *SimpleKMeans* is the subclass.

5.1.4.4 Cardinality

As shown in Table 5-2, Cardinality denotes the “has a” relationship among classes.

Table 5-2 Cardinality

0	No instances (rare)	-
0..1	No instances, or one instance	Customer reviews application can use 0 or 1 *FilteredClassifier*
1	Exactly one instance	A *FilteredClassifier* can have only one instance
0..*	Zero or more instances	*StringToWordVector* / *ArffLoader* can be associated with any no of *FilteredClassifiers* / *SimpleKMeans*
*****	Zero or more instances	-
1..*	One or more instances	*SimpleKMeans* can use one or more instances

5.2 Data Preprocessing Module

Social media reviews are very noisy and full of all kinds of spelling, grammatical, and punctuation errors. Most natural language processing (NLP) tools such as part-of-speech (POS) taggers and parsers need clean data to perform accurately. Thus, a significant amount of pre-processing is needed before any analysis. See [8] for some pre-processing tasks and methods. So, the data is processed before it is taken for actual classification and clustering.

The C#.Net code for this is shown in Appendix-A Class-1.

5.2.1 Cleaning

The reviews collected from the Facebook database cannot be expected to be formally written. It may contain some junk data like hashtags, emoticons, line breaks, and punctuations. All characters other than letters and numbers should be cleaned. The data cleaning module in the model deliberately removes such characters from the reviews leaving behind only the meaningful data. Regular expression used to remove special characters is shown in Figure 5-13. This expression is added to the C# .Net code developed. It will remove any character that is not a number or an alphabet from the customer reviews collected.

```
Regex.Replace(review_text,"[^a-zA-Z0-9]+", " ")
```

Figure 5-13 Regular expression to remove special characters

5.2.2 Filtering

StringtoWordVector. The *J48* classifying algorithm cannot handle attributes of type String. So, attribute *comment* has to be converted to a format that is intelligible to the *J48* algorithm. The StringtoWordVector filter class in WEKA transforms string attributes into word vectors, i.e. creating one attribute for each word which either encodes presence or word count within the string [126]. The number of words to be in the output vector can also be configured with a parameter *WordsToKeep*. If it is not specified, by default it takes 1000.

5.2.3 Tokenizer

Tokenization is the task of breaking a character sequence up into pieces (words/phrases) called tokens and at the same time throws away certain characters such as punctuation marks. The list of tokens is then used for further processing. In this model, the *WordTokenizer* algorithm is used to break down the character sequence into words.

5.2.4 Stop words

In the collected data there may be few commonly used words which will not add value to the output. Using the option *stopwordsHandler*, WEKA can be configured so as to ignore those commonly used words. Among the different types of stop word handlers, the *MultiStopwords* algorithm serves the purpose well. It applies the specified stop words algorithms one after other. As soon as a word has been identified as a stop word, the loop is exited.

5.2.5 Stemming

After the commonly used words are removed, the data may still contain repetitive words implying the same meaning and those have to be identified and addressed next. The process of finding the root of any particular word is called *Stemming*. There are several algorithms that excel in finding the root word. They are language dependent. In this model, the *IteratedLovinsStemmer* algorithm is used. It stems the word (in case it's longer than 2 characters) until no further changes are required.

5.3 Data Classifying and Clustering Module

5.3.1 J48 Classifier

The pre-processed data is classified using the *J48* classifier algorithm in two iterations. In iteration 1, the data is classified based on the class attribute *Category* and then in Iteration 2, it is classified based on the class attribute *Sentiment*. The C#.Net code for this is shown in Appendix-A Class-2.

Training

J48 builds decision trees from training data, using the concept of information entropy. The training data is a set S = {S1, S2… Sn} of already classified samples. At each node of the tree, *J48* chooses the attribute of the data that most effectively splits its set of samples into subsets. Each branch is split based on a condition on any attribute. All leaves represent a class attribute. So for each instance, the classification starts from the root and traverses along a branch based on the attribute values of that instance and will at last lead to a class attribute.

The steps for building decision trees are as follows:

Step 1: The leaf is labeled with the same class if the instances belong to the same class.

Step 2: For every attribute, the potential information will be calculated and the gain in information will be taken from the test on the attribute.

Step 3: Finally, the best attribute will be selected based on the current selection parameter.

Pseudocode for *J48* algorithm is as follows:

```
BEGIN
Create a root node N;
        IF (T belongs to the same class C)
        {
                Leaf node = N;
                Mark N as Class C;
                Return N;
        }
        For i = 1 to n
        {Calculate information gain, Ti};
        T = testing attribute;
        N.T = attribute having highest information gain;
        Return N;
END
```

Classification

From the reviews that are collected from the Facebook page, we have to infer what it is about and what its mood is. So, the data have to be classified in two iterations, one iteration to find the *Category* (Taste, Price, Service, etc.), and the second iteration to find the *Sentiment* (positive or negative). For both the iterations, the *J48* classifier is used. Since the review is of string type and the *J48* classifier cannot handle string attributes, *StringToWordVector* is used as a filter to convert strings to words.

The configurations are shown in Figure 5-14 and Figure 5-15, respectively.

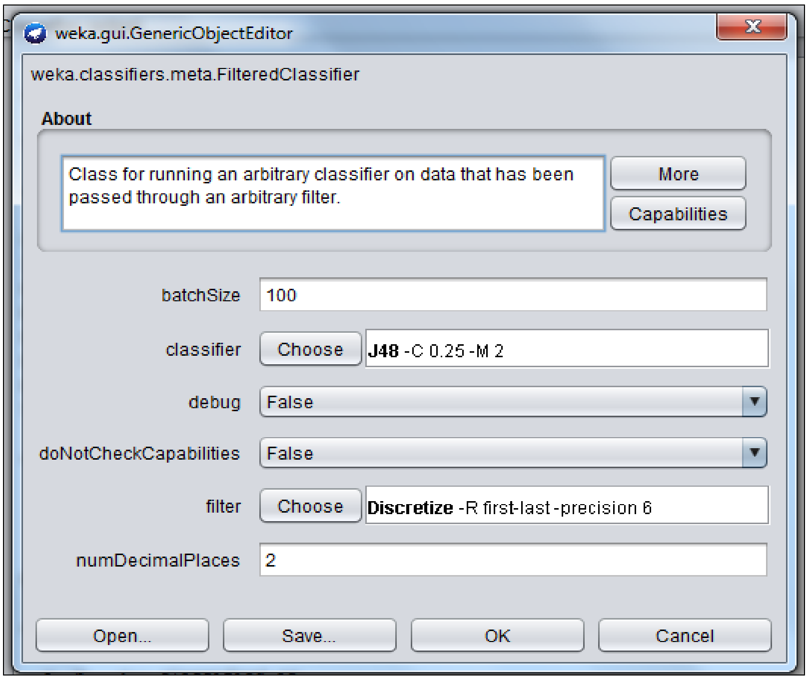

Figure 5-14 Classifier configuration

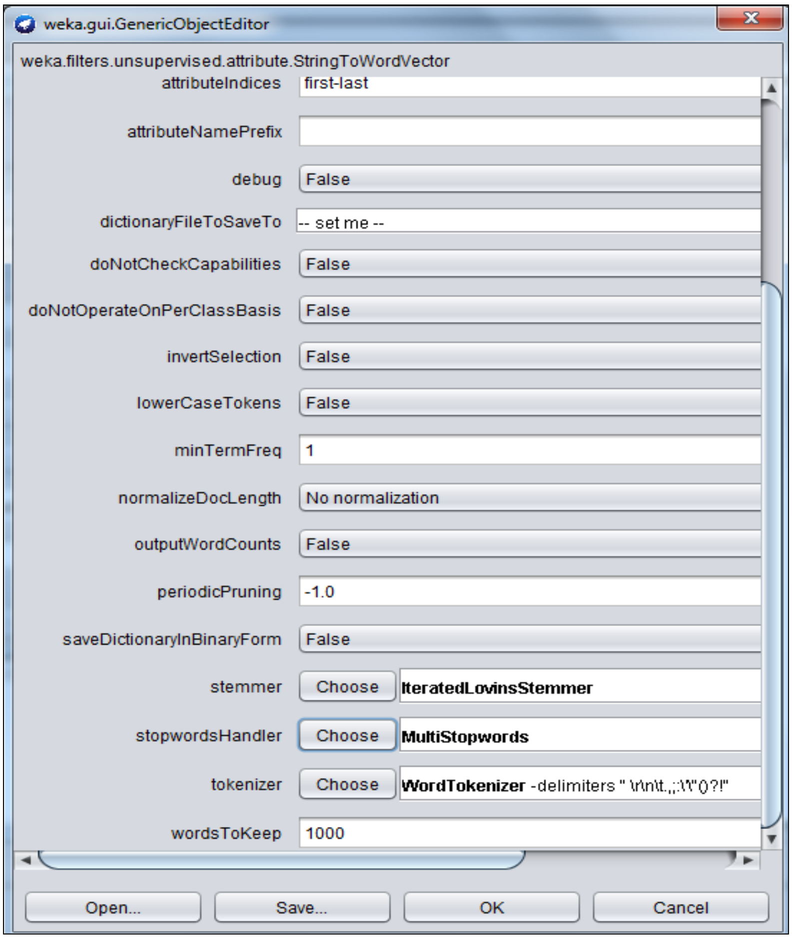

Figure 5-15 Filter configuration

Before starting the actual classification in each iteration, WEKA is trained. Training files are prepared which has most of the possible cases for each category and sentiment. From the training instances, the *J48* classifier builds a *J48* pruned tree. Keeping this as a model, it classifies the incoming data. Classifier models generated for category classification of restaurant customer reviews had 1246 instances with an accuracy of 94.2%. The ROC area for different features under consideration falls in the excellent region i.e. 0.9 to 1 as shown in Figure 5-16. Based on the performance factors discussed in Chapter 4, this model is considered optimal for categorizing the Facebook reviews to the different features.

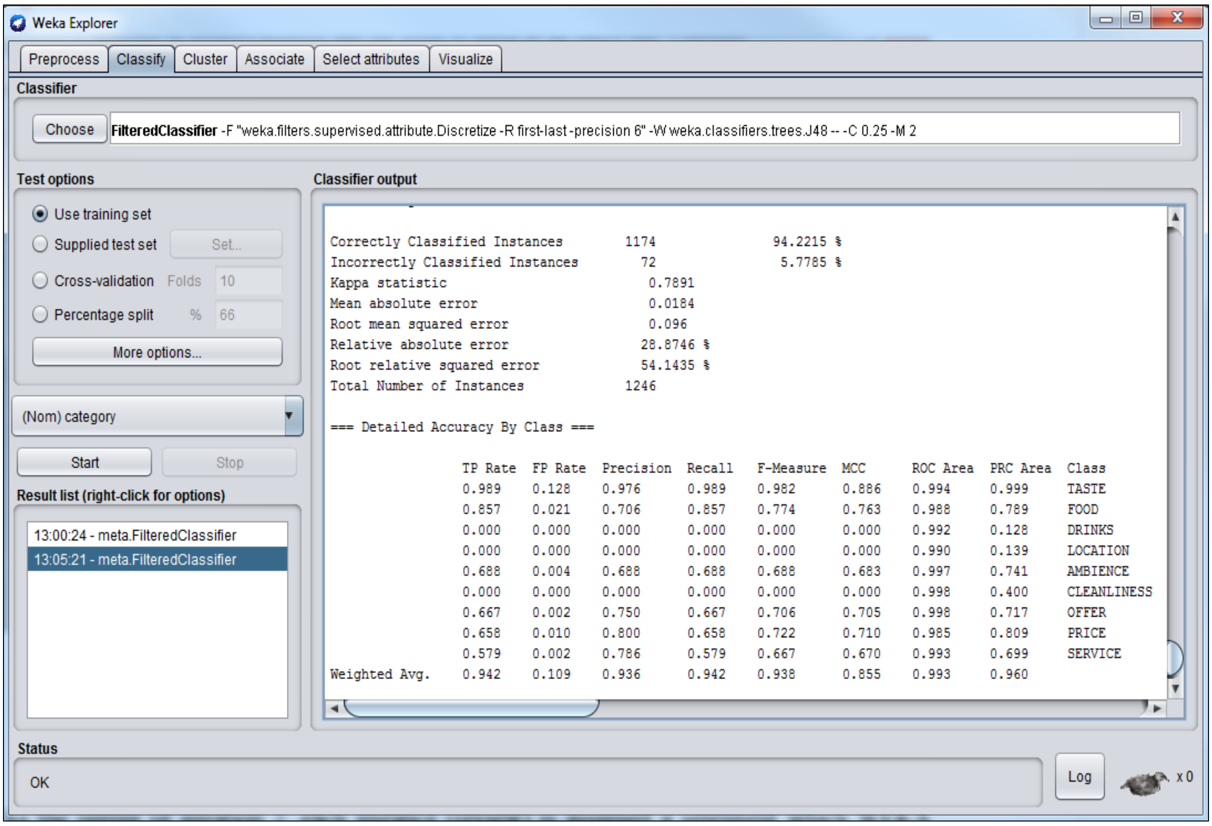

Figure 5-16 Category classification for restaurant reviews

As the output of iteration 1, each instance (review) is assigned a category, which WEKA finds it is best suitable for, and saved as an ".arff" file as shown in Appendix-B Categorized restaurant reviews. This file is processed, i.e. one more attribute sentiment is added and given as an input for the second iteration. In this iteration, data is classified for sentiment and each instance is assigned as positive/negative and the result is saved as an ".arff" file. Classifier models generated for sentiment classification of restaurant customer reviews had 1246 instances with an accuracy of 72.8%. The ROC area for *positive* and *neutral* sentiments falls in the fair region i.e. 0.7 to 1, and the ROC area for *negative* sentiment falls in the excellent region i.e. 0.9 to 1 as shown in Figure 5-17. Based on the performance factors discussed in Chapter 4, this model is considered optimal for finding the sentiment of the customer reviews collected from Facebook.

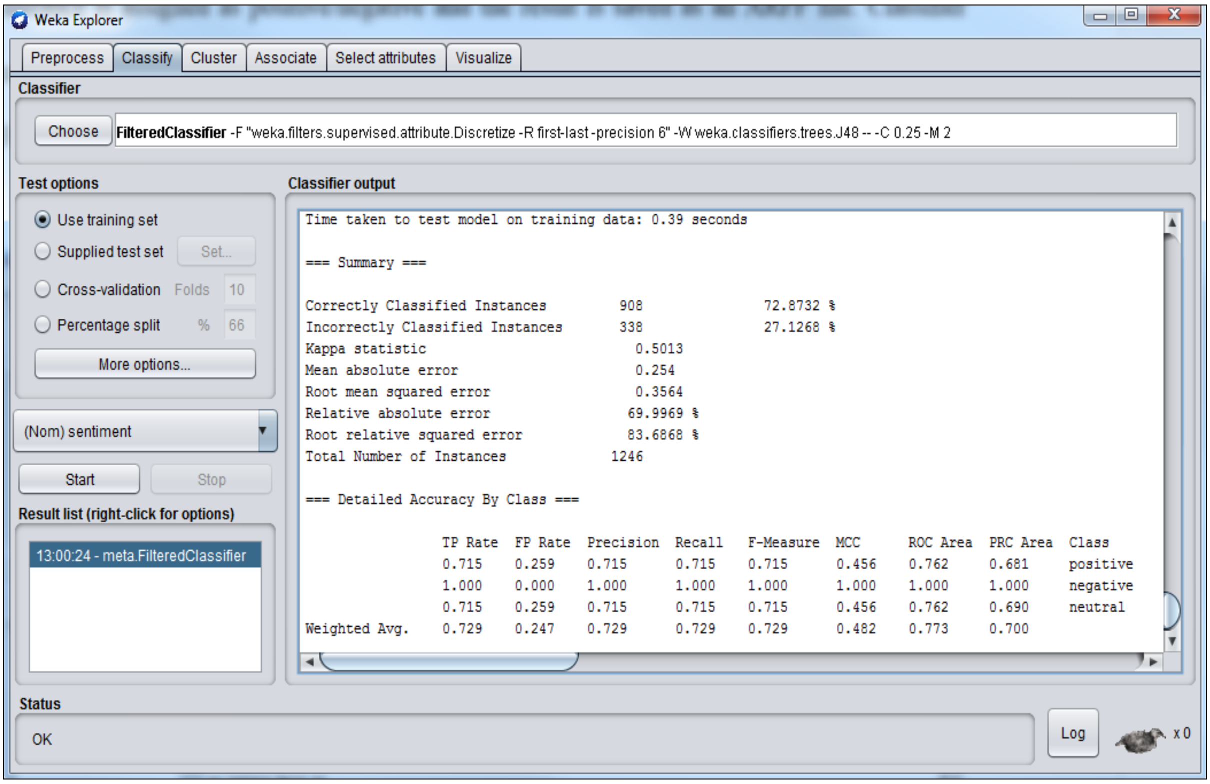

Figure 5-17 Sentiment classification for restaurant reviews

As the output of iteration 2, each instance (review) is assigned a sentiment, which WEKA finds it is best suitable for, and saved as an ".arff" file as shown in Appendix-B Sentimental analysis results for restaurant reviews.

5.3.2 Simple K-means clustering and Visualization

In the last step, the vast amount of social media data was classified using the *J48* classification algorithm to know which categories, i.e. features, the customers are concerned about, along with whether they are satisfied with the feature or dissatisfied about it. Even though the unused data was made useful, the outcome is an ".arff" file, and it requires manual effort to convert the ".arff" file into an Excel file and analyze it. So, this process is automated using the *Simple K-means* clustering algorithm. The C#.Net code for this is shown in Appendix-A Class-3.

}

The Simple K-means algorithm randomly chooses k centroids initially, and points that are closer to any centroid are clustered together.

Step 1: k initial means (here k=3) are randomly generated within the data domain.

Step 2: k clusters are created by associating every observation with the nearest mean.

Step 3: The centroid of each of the k clusters becomes the new mean.

Step 4: Steps 2 and 3 are repeated until convergence has been reached.

Pseudocode for *Simple k-means* algorithm is as follows:

```
BEGIN
        Select number of clusters,  K = 3;
        Assign three initial random centroids (means), C;
        Associate every observation to the nearest mean;
        WHILE (Centroid fail to converge)
        {
                {
                        {Calculate the distance d between the point and the centroid;
                        IF(d < C)
                                Assign the point to the cluster;
                        ELSE
                                Discard the point;
                }
        Re-calculate the new centroid of the cluster by averaging all the points in the cluster;
        }
END
```

The classified customer reviews are grouped into three clusters using WEKA. Each cluster holds instances of the same feature.

Each cluster contains the below data:

1) Instances that are satisfactory to the customers (Blue in Figure 5-18).

2) Instances that are not satisfactory (Red in Figure 5-18).

3) Instances that are neither satisfactory nor unsatisfactory (Green in Figure 5-18).

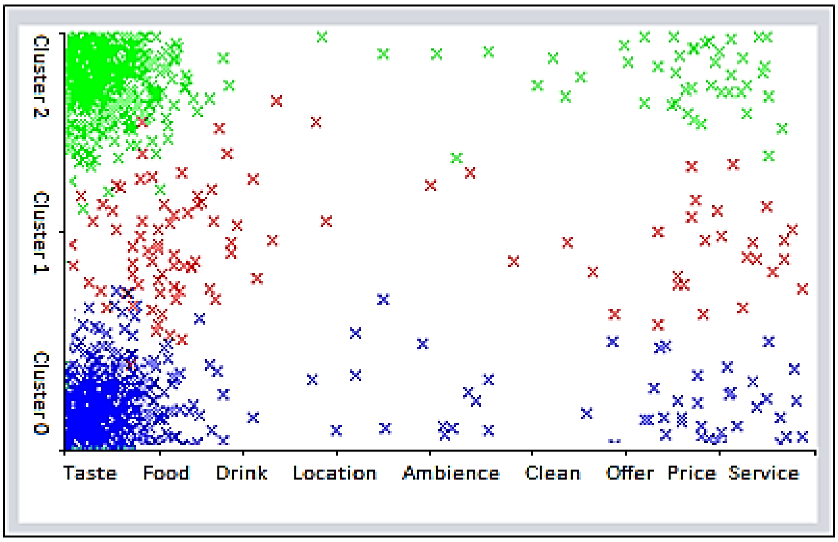

Figure 5-18 Restaurant review instances grouped into 3 Clusters

5.3.3 GUI Developed for Restaurant MITACS Project

In the case study, reviews of a popular restaurant are collected from its official Facebook page using Facebook Graph API and classified using the *J48* classifier in two iterations and then clustered using the *Simple K-Means* algorithm. From the outcomes of the feedback analysis model developed for the restaurant we infer that features like *Taste, Food, Drink Variety, Price,* and *Service* are given more importance by customers and they are least bothered about features like *Location, Clean, Offers*, and *Ambiance* of the restaurant.

Language Used: C#.

IDE: Visual Studio 2013.

Libraries Imported: Facebook.dll, WEKA.dll.

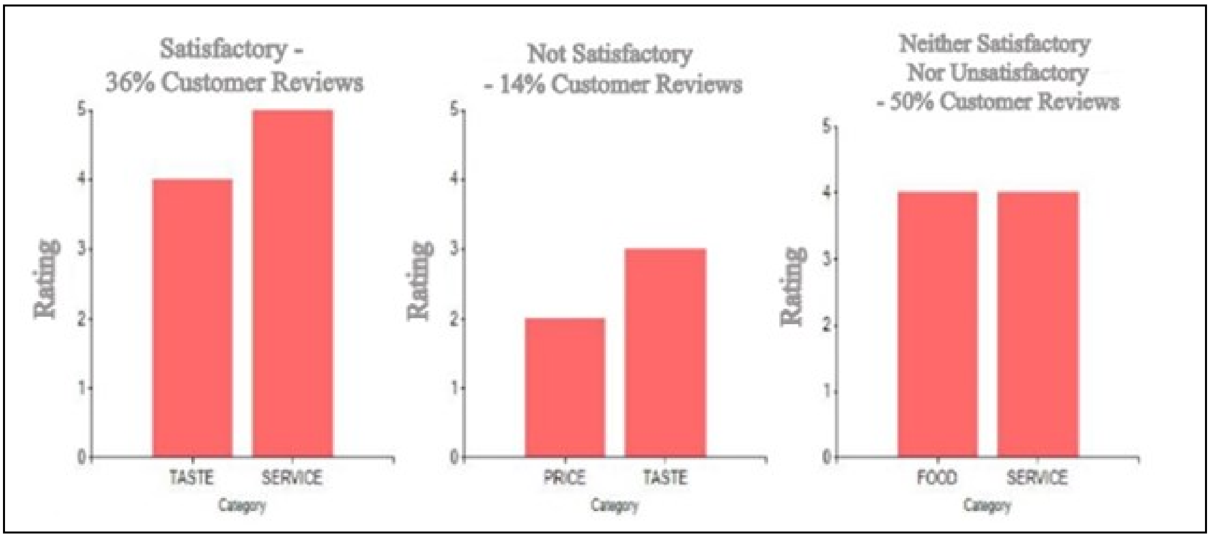

Figure 5-19 GUI of process model outcome

The designed software model was used to collect reviews from the Facebook page of a restaurant using Facebook.dll to find what is being talking about, whether it is positive or negative using the classification concept in WEKA.dll and finally cluster the results to visualize them in three graphs (see Figure 5-19): Categories that the customers are 1) satisfied with, 2) Not satisfied with, and 3) Neither satisfied nor unsatisfied with. The graphs also show the percentage of reviews from which the results are inferred. All the above actions are carried out with just a button click!

CHAPTER 6. CASE STUDY ON AN ENGINEERING FIRM'S PRODUCTS AND PROCESSES

In this section, the real-life implementation of the proposed framework on Durabuilt Windows and Doors Inc. (Durabuilt), a leading windows and doors manufacturing company in Northern Alberta, is discussed in detail. The framework is implemented in three steps as shown below in the tree diagram (see Figure 6-1). All the customer reviews from Facebook were first assigned to one of the three categories; 1) Windows, 2) Doors, or 3) Service. In the next step, the Windows and Doors reviews were analyzed separately to identify the performances of their respective features.

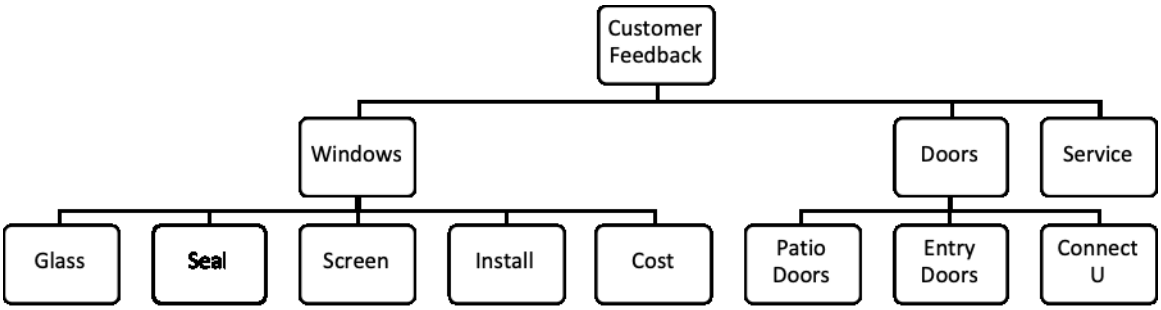

Figure 6-1 Hierarchy of Durabuilt reviews analysis

6.1 Iteration 1: Analyzing Overall Firm Performance

6.1.1 Iteration 1.1: Categorizing Product and Service Reviews

In Iteration 1.1, all the reviews from Durabuilt are divided into *product,* i.e. windows and doors, and *service* reviews as shown in the ".arff" file (see Appendix - C.1). Customer feedback reviews from Facebook on *Windows* and *Doors* are analyzed in iteration 2 and 3, respectively, to find out their respective feature performances. *Service* reviews are filtered in the first iteration and will not be considered while analyzing reviews related to product features. Performance of the classification model generated for product and service reviews categorization for Durabuilt is shown in Figure 6-2. The model had 187 instances with an accuracy of 100%. Based on the performance factors discussed in Chapter 4, since all the performance factors are 1, and ROC area is in the excellent region, i.e. 0.9 to 1, this model is considered optimal for categorizing the Facebook reviews to the different features under consideration.

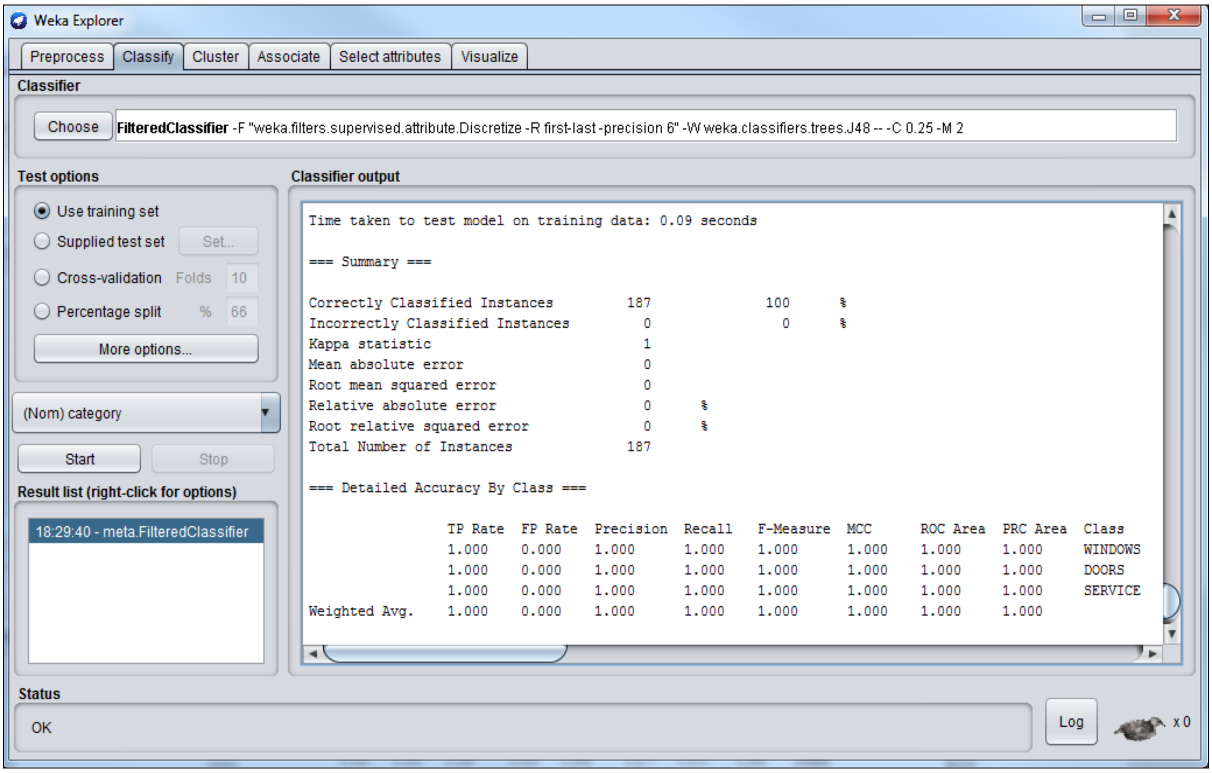

Figure 6-2 Model for categorizing product and service reviews

6.1.2 Iteration 1.2: Sentimental Analysis of the Categorized Products and Service

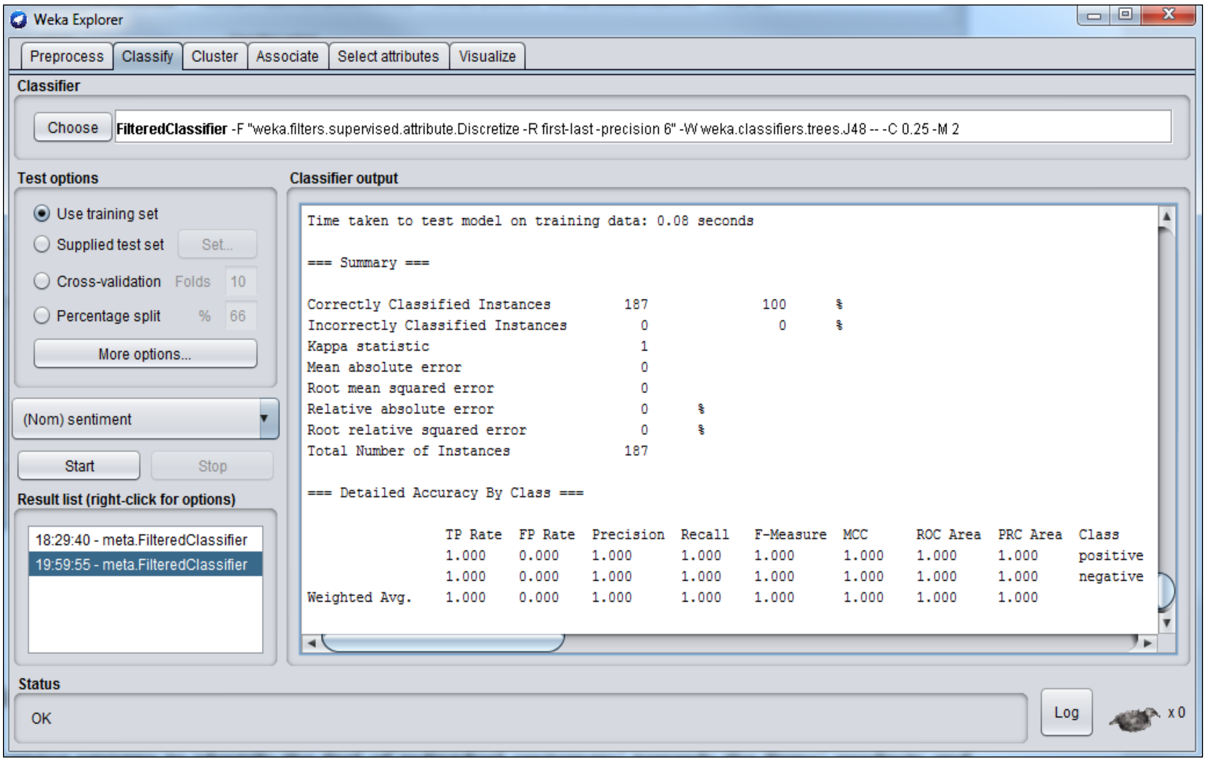

Figure 6-3 Sentimental analysis model for classifying overall firm performance

In iteration 1.2, Sentimental analysis was done on the categorized product and service reviews to identify how customers feel about the firm's products and processes. The Sentimental analysis model developed had 187 instances with an accuracy of 100%. Based on the performance factors discussed in Chapter 4, since all the performance factors are 1, and ROC area is in the excellent region, i.e. 0.9 to 1 (see Figure 6-3), this model is considered optimal for finding the sentiment of the customer reviews collected from Facebook.

The *Simple K-means* algorithm was then used to cluster the output (".arff") file generated in iteration 1.2 (shown in Appendix - C.2) to find the overall customer satisfaction on the firm's products and processes. In Figure 6-4, green indicates satisfied customer reviews, and red indicates reviews from unsatisfied customers.

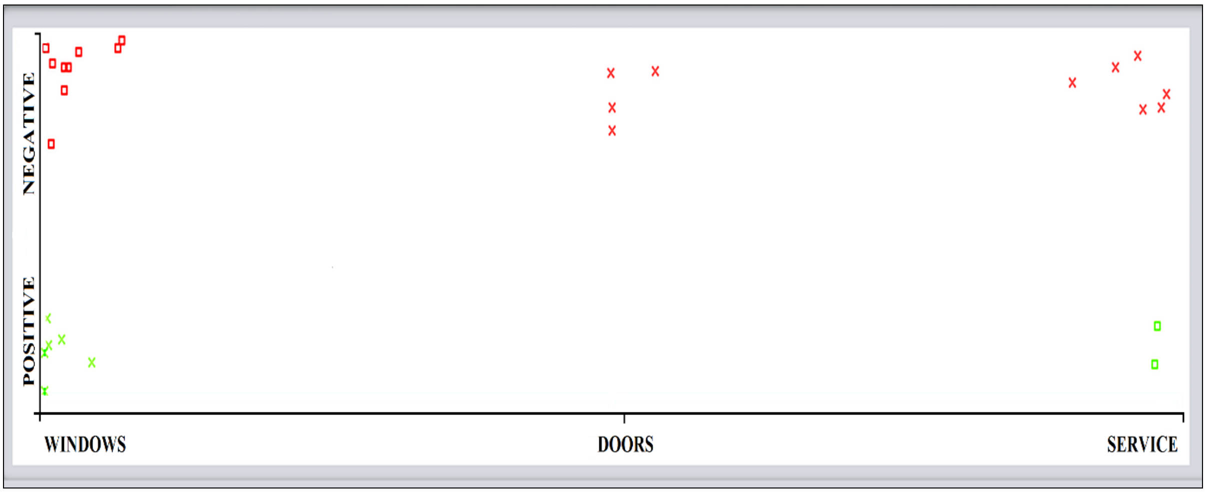

Figure 6-4 Overall customer satisfaction of Durabuilt

From the outcomes of the first iteration, we cannot clearly identify the reasons for customer dissatisfaction with the window and door products. So, there was a need to analyze the specific product features in a more comprehensive way as a continuous improvement process to identify the root causes for customer complaints and increase customer satisfaction.

6.2 Iteration 2: Analyzing Window Feature Performances

6.2.1 Iteration 2.1: Categorizing Window Features

Durabuilt follows the Bill of Materials (BoM) based pricing system. BoM contains various product features and their respective prices. Whenever a quote for a new order or service order is created, customers are charged based on the product features in BoM. In iteration 2.1, using the classification model developed and the customer reviews of Durabuilt, five such product features were categorized; 1) Glass, type of window (single, dual pane or triple pane); 2) *Seal*, sealant used for sealing the dual and triple pane window glasses; 3) *Screen*, window screens; 4) *Install*, reviews on installation process and performance of installers; and 5) *Cost*, labor and other charges related to the new order or service order (see Appendix - C.3). The model had 176 instances with an accuracy of 100%. Based on the performance factors discussed in Chapter 4, since all the performance factors are 1, and ROC area is in the excellent region i.e. 0.9 to 1 (see Figure 6-5), this model is considered optimal for categorizing the Facebook reviews to the different window features under consideration.

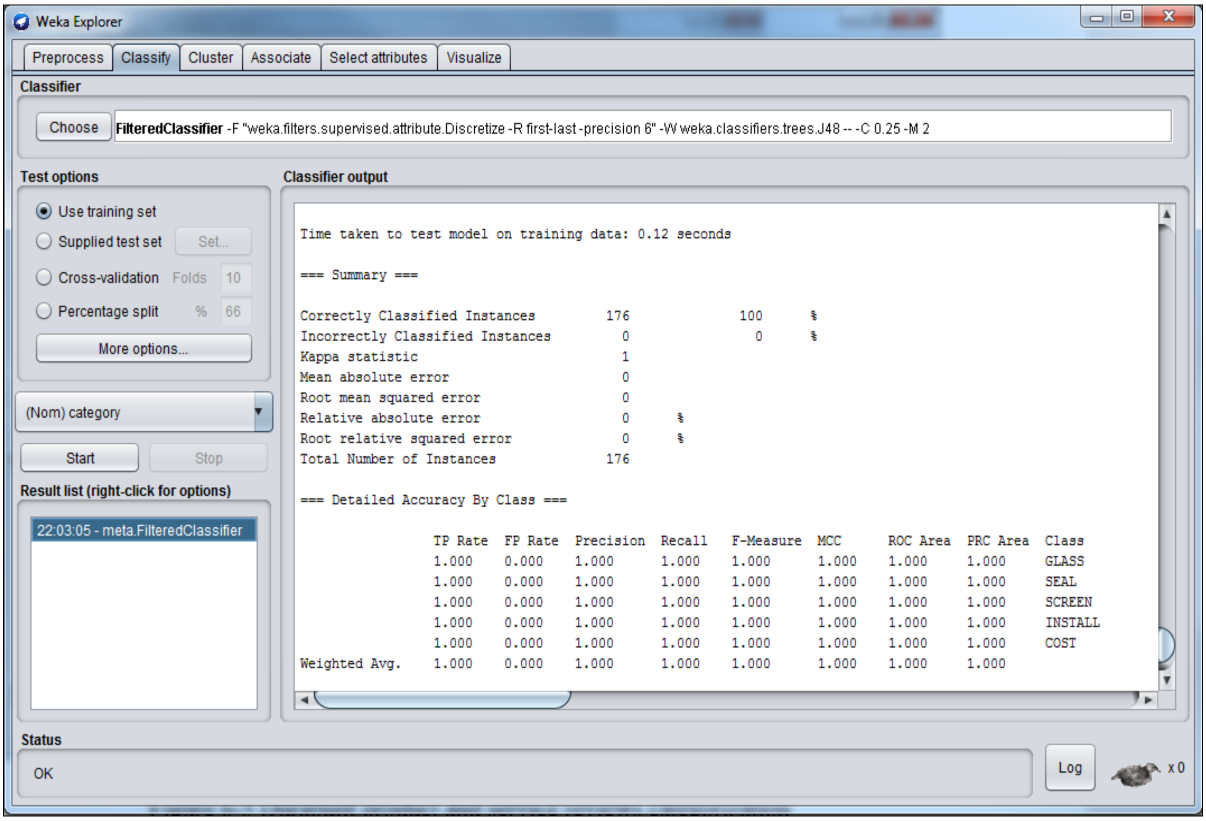

Figure 6-5 Model for categorizing window features

6.2.2 Iteration 2.2: Sentimental Analysis on Window Features

In this iteration, Sentimental analysis was done on the reviews related to window features categorized in the last iteration. The positive and negative reviews of customers about each window feature can be identified using the model shown in Figure 6-6. The model had 176 instances with an accuracy of 100%. Based on the performance factors discussed in Chapter 4, since all the performance factors are 1, and ROC area is in the excellent region, i.e. 0.9 to 1 (see Figure 6-6), this model is considered optimal for finding the sentiment of the customer reviews collected from Facebook. The results of the sentimental analysis are stored in an .”arff” file as shown in Appendix - C.6.

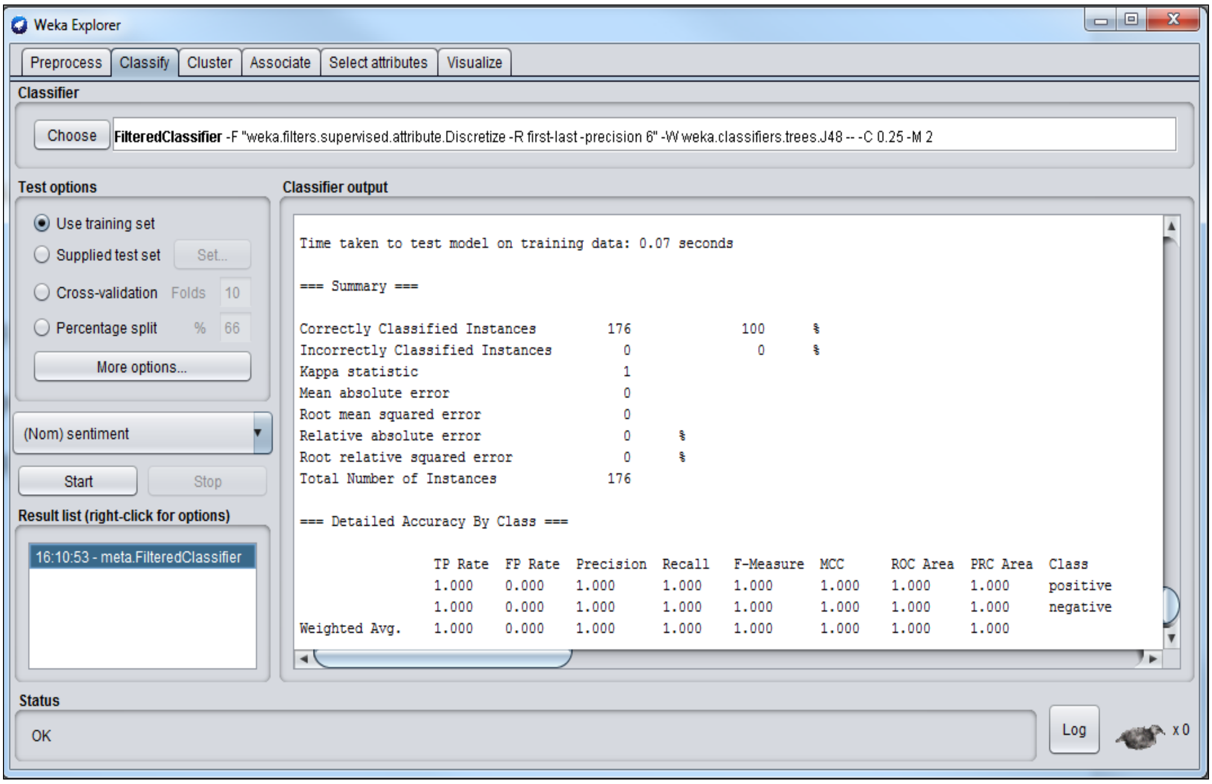

Figure 6-6 Sentimental analysis model for window features

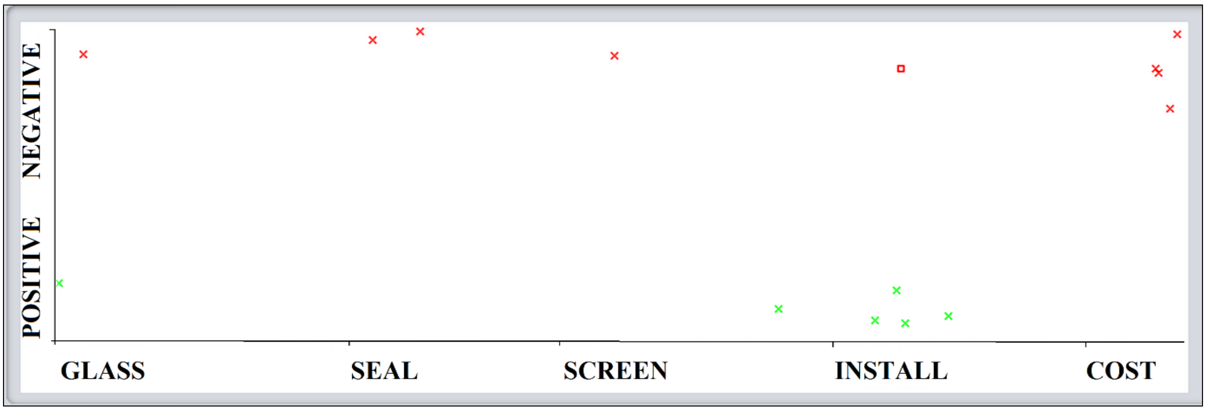

Figure 6-7 Customer satisfaction visualization for window features

The graph shown in Figure 6-7 was generated using the outcomes of sentimental analysis, which are shown in Appendix - C.4. From the analysis on window features, we infer that the customers are not satisfied with product features like *Glass, Seal, Screen,* and *Cost.* However, customers are satisfied with the *Install* feature.

6.3 Iteration 3: Analyzing Door Feature Performances

6.3.1 Iteration 3.1: Categorizing Door Features

Durabuilt offers two types of doors, 1) *Entryways*, entry doors and 2) *Patio doors*, available in a hinged or sliding design. In this iteration, reviews were categorized using the classification model (see Figure 6-8) into three features related to doors: 1) *PatioDoor*; 2) *ConnectYou*, support department related to door complaints; and 3) *EntryDoor*. The results were stored in an ".arff" file as shown in Appendix - C.5. The model had 26 instances with an accuracy of 100%. Based on the performance factors discussed in Chapter 4, since all the performance factors are 1, and ROC area is in the excellent region, i.e. 0.9 to 1 (see Figure 6-8), this model is considered optimal for categorizing the Facebook reviews to the different door features under consideration.

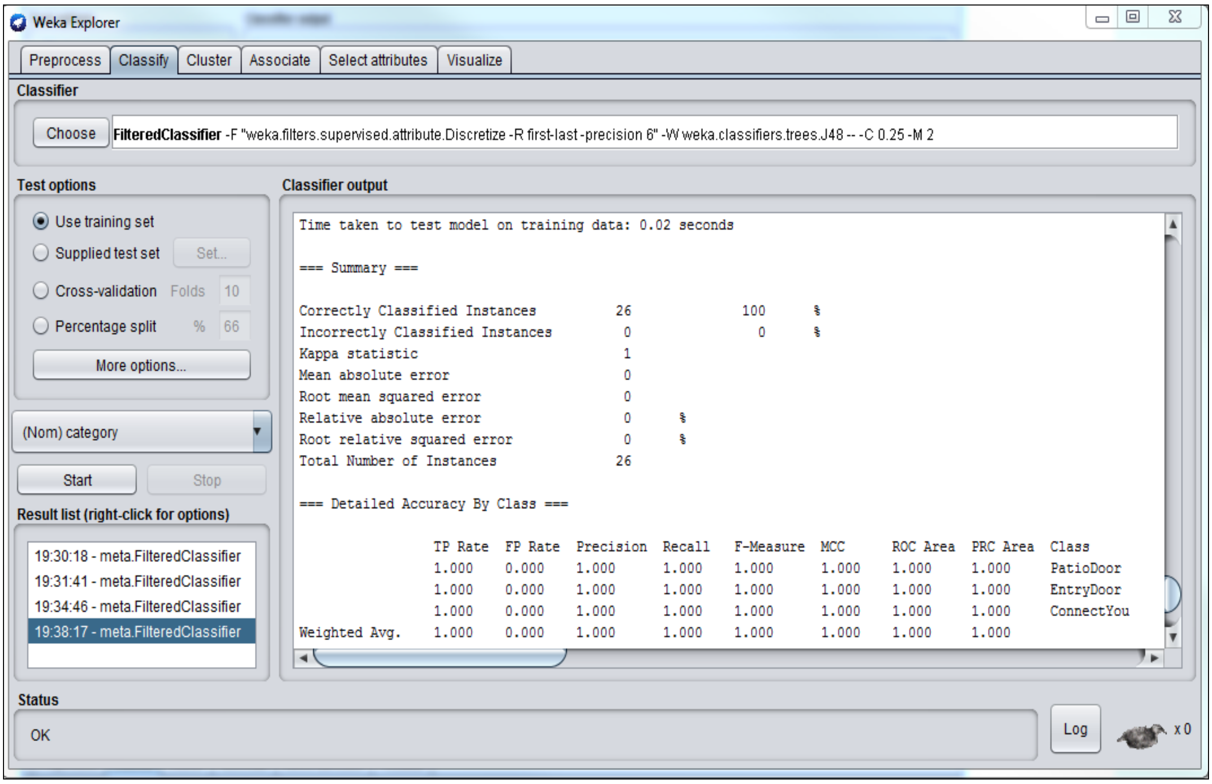

Figure 6-8 Model for categorizing door features

6.3.2 Iteration 2.2: Sentimental Analysis on Door Features

The reviews categorized in previous iterations are analyzed using the model shown in Figure 6-9 to find the feel of customers about each door feature. The model had 240 instances with an accuracy of 100%. Based on the performance factors discussed in Chapter 4, since all the performance factors are 1, and ROC area is in the excellent region, i.e. 0.9 to 1 (see Figure 6-9), this model is considered optimal for finding the sentiment of the customer reviews collected from Facebook. The results of the sentimental analysis are stored in an ".arff file" as shown in Appendix - C.6.

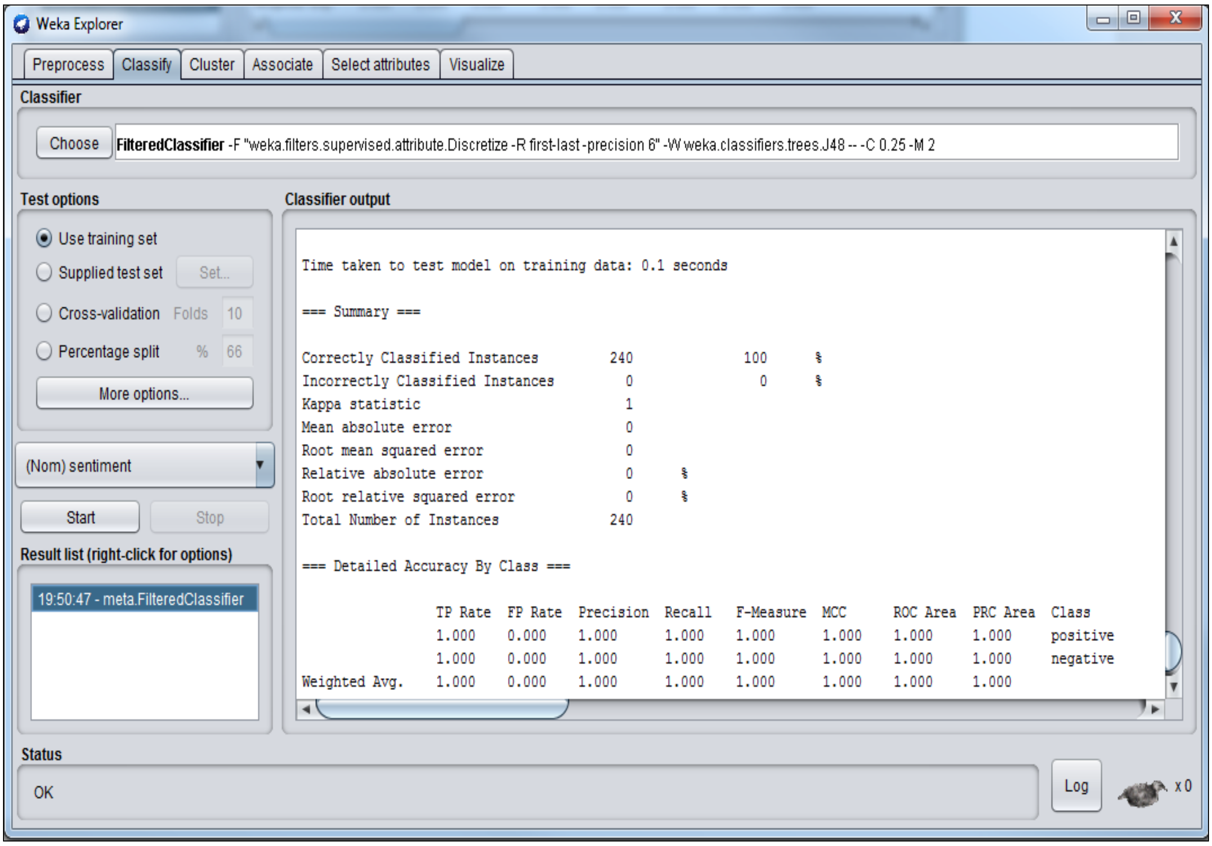

Figure 6-9 Sentimental analysis model for door features

The graph shown in Figure 6-10 was generated using the outcomes of the sentimental analysis. From the analysis, we infer that all the door features are not performing well. There is a need for improvement in all the areas analyzed with respect to doors.

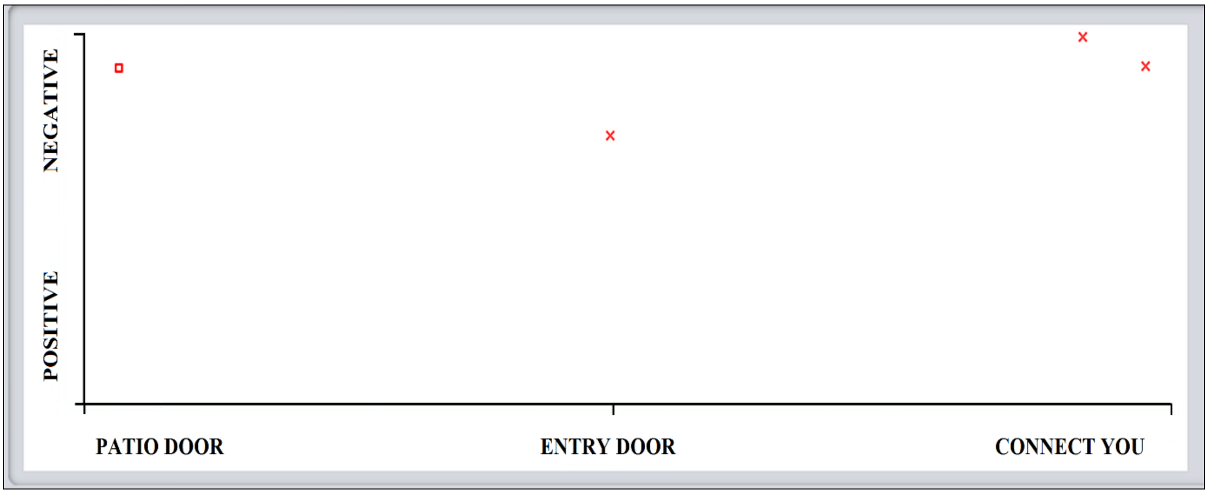

Figure 6-10 Customer satisfaction visualization for door features

CHAPTER 7. CONCLUSION AND FUTURE RESEARCH

In this chapter, conclusions, key contributions, and recommendations for future research are presented. First, the process of development of the collaborative feature-oriented software application is discussed briefly. Then, the contribution of this research is summarized; finally, recommendations for future study are proposed.

7.1 Summary

In summary, an attempt was made to gather all the surveys, proposed frameworks, and comparisons of the efficiency of various classification and clustering algorithms in Chapter 2. From the case studies in Chapter 3, it can be concluded that data mining has the capability to be applied in many different fields. This proves that data mining can be a good foundation for empirical research. In Chapter 4, a decision was made to use the *J48* classification algorithm and the *Simple K-means* clustering algorithm, which are the most suitable for analyzing customer reviews on Facebook.

The conceptual framework proposes a feedback analysis model that indicates the change in feature performance. The different product features can be monitored in real-time. The experiment in Chapter 5 has proved that the product feature performances can be monitored throughout their lifecycle and how real-world issues can be identified with the adoption of ICT, thus enabling effective risk management.

Lack of information processing is a significant challenge which hinders the process of feature innovation and loss of valuable data along the product lifecycle. Also, it can lead to inefficient utilization of the available big data as companies set up manual feedback systems for monitoring customer satisfaction. In addition, it has hindered the development of standardized frameworks and solutions to improve CCE practices. These are some of the challenges which the proposed framework has addressed.

7.2 Challenges Faced and Proposed Solutions in the Implementation

The research presented in this thesis offers the following contributions:

The data in social media acts as a potential gold mine for discovering product knowledge. A large amount of feedback data is available in the form of Facebook reviews. This research focused on extracting and consolidating feature opinions of customers from Facebook reviews to monitor and measure customer satisfaction on different product features. First, there was a need to automatically extract this data since a lot of man-hours should be invested in manually collecting the customer reviews. This process was automated using the Facebook Graph API to collect data in real-time.

There was a need for focusing on challenges like how to reduce the time, memory, i.e., size, and dimensionality of the data without affecting the characters of the social networks, i.e., only data that is required to represent the context must be extracted. In addition, the showstopper for designing an opinion mining system for analyzing reviews arises from the fact that customer reviews are often noisy. Any noisy or corrupted data might affect the result. Data collected from Facebook contained irrelevant data that are not required for further processing. The data was also not compatible with WEKA. All these challenges were overcome using the data pre-processing module in which the data was cleaned, filtered, tokenized, stemmed and stop words were removed.

Finally, given the considerable amount of data that is associated with Facebook, the creation of the relevant technology to facilitate ICT was a significant challenge in the development of this feedback analysis model. C#.Net coding language is used to facilitate the integration of the system, which aids in the process of knowledge assimilation for feature-based knowledge abstraction and also performs the analytics. The system identifies product features from customers' opinion and categorizes them as positive or negative. Opinions on Facebook can be congregated at any desired level of specificity i.e. feature level or product level, user level or service level, etc. We have developed a system based on this approach, which provides the user with a platform to analyze opinion expressions from Facebook.

The application of the proposed feedback analysis model on customer reviews of Durabuilt Windows and Doors Inc., a leading windows and doors manufacturing company in Northern Alberta; and Kathir Food Experience Inc., a famous South Indian Restaurant in Edmonton, illustrates how the existing techniques can be used as a real-time diagnostic tool to track product features and their performances. The proposed informatics framework can act as a link between the real world and the management. It can categorize social media reviews to provide better visualization for decision-making.

The proposed informatics framework can also be used after the product has been sold to the customers for effective management of product lifecycle activities, can aid in new product development, maintenance and service of the existing product, and to understand when the production of the existing product should be stopped. Thus integrating people, processes, and technology.

7.3 Limitations and Recommendations for Future Research

Based on the research presented here, the following are the recommendations for future work:

(1) Make the software application more user-friendly.

(2) Integrate conventional engineering tools and techniques with the developed informatics software applications.

(3) ICT models like the one developed here can aid in achieving the objectives of Industry 4.0 [127], where all data should be connected, and the decision should be taken with precise understanding automatically. With such models, more data can be made accessible and processed over the internet. So, there is a need for managing security and privacy issues.

(4) We can collect large customer-generated data including videos and images for analyzing the customer's experience, and graphs can be used to visualize both positive and negative experience results on a timely basis. To some degree, the final results from data visualization may be the things that decision makers care most about rather than the original data. Thus, proposing appropriate data visualization tools can add great value.

(5) We can develop a predictive model that can analyze multilingual texts, i.e., reviews that are given by customers in their native languages other than English.

(6) The design of better algorithms and knowledge representation schemes will be an important complement to the tremendous potential offered by emerging data mining technologies which can be used to develop fully functional big data processing software.

REFERENCES

[1] F. Remondino and T. Chen, "35 Years of Internet, 10 Years of Isprs Online," *Wired*, 1994.

[2] D. B. Malkoff, "A framework for real-time fault detection and diagnosis using temporal data," *Artif. Intell. Eng.*, vol. 2, no. 2, pp. 97–111, Apr. 1987.

[3] C. V. Ramamoorthy and B. W. Wah, "Knowledge and data engineering," *IEEE Trans. Knowl. Data Eng.*, vol. 1, no. March, pp. 9–16, Mar. 1989.

[4] J. Kaplan, "Artificial intelligence," *Commun. ACM*, vol. 60, no. 1, pp. 36–38, 2016.

[5] N. Hemageetha and G. M. Nasira, "Classification of Soil type in Salem district using J48 algorithm," *Int. J. Control Theory Appl.*, vol. 9, no. 40, pp. 33–41, 2016.

[6] A. AJINKYA, A. LONDHE, J. R.S, and S. B. R, "Analyzing Student's Learning Experiences Through," no. 12, 2016.

[7] F. Alam and S. Pachauri, "Comparative Study of J48 , Naive Bayes and One-R Classification Technique for Credit Card Fraud Detection using WEKA," *Adv. Comput. Sci. Technol.*, vol. 10, no. 6, pp. 1731–1743, 2017.

[8] L. Dey and S. M. Haque, "Opinion mining from noisy text data," *Int. J. Doc. Anal. Recognit.*, vol. 12, no. 3, pp. 205–226, 2009.

[9] L. Einav and J. Levin, "Economics in the age of big data," *Science*, vol. 346, no. 6210, 2014.

[10] D. Kajaree and R. Behera, "A Survey on Web Crawler Approaches," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 2, pp. 1302–1309, 2017.

[11] J. Backhouse, "What makes lecturers in higher education use emerging technologies in their teaching," *Knowl. Manag. E-Learning*, vol. 5, no. 3, pp. 345–358, 2013.

[12] E.Raju and K.Sravanthi, “Analysis of Social Networks Using the Techniques of Web Mining,” Elsevier Ltd, Edmonton, 2017.

[13] B. Marr, “Big Data Sources - https://www.smartdatacollective.com/big-data-20-free-big-data-sources-everyone-should-know/,” 2014.

[14] Facebook Graph API, “https://developers.facebook.com/docs/graph-api,” 2018.

[15] UK Government, “https://data.gov.uk/,” 2018.

[16] CIA, “https://www.cia.gov/library/publications/the-world-factbook/,” 2018.

[17] US Government Portal, “http://data.gov/,” 2018.

[18] Google Trends, “https://trends.google.com/trends/?geo=US,” 2018.

[19] National Climatic DataCenter, “https://www.ncdc.noaa.gov/Data-Access/Quick-Links#Loc-Clim,” 2018.

[20] Likebutton, “https://domainnamesales.com/domain/likebutton.com?landerid=likebutton5b1ae07b27daa2.01205237,” 2018.

[21] J. Constine, “Facebook now has 2 billion monthly users...and responsibility,” *TechCrunch*, 2017.

[22] D. Zeng, H. Chen, R. Lusch, and S. H. Li, “Social media analytics and intelligence,” *IEEE Intell. Syst.*, vol. 25, no. 6, pp. 13–16, 2010.

[23] M. Mele and G. Campana, “Prediction of Kansei engineering features for bottle design by a Knowledge Based System,” *Int. J. Interact. Des. Manuf.*, vol. 12, no. 4, pp. 1201–1210, 2018.

[24] X. G. Ming *et al.*, “Collaborative process planning and manufacturing in product lifecycle management,” *Comput. Ind.*, vol. 59, no. 2–3, pp. 154–166, Mar. 2008.

[25] R. Mejia, D. Romero, A. Molina, N. Galeano, and N. P. Verdeza, “Collaborative Engineering Environments for Virtual Organisations,” *Int. J. Inf. Technol. Manag.*,

vol. 8, no. 3, p. 298, 2009.

[26] G. Pezzotta, S. Cavalieri, and D. Romero, "Collaborative product-service systems engineering: Towards an active role of customers and stakeholders in value co-creation," *2017 Int. Conf. Eng. Technol. Innov. Eng. Technol. Innov. Manag. Beyond 2020 New Challenges, New Approaches, ICE/ITMC 2017 - Proc.*, vol. 2018-Janua, pp. 1247–1254, 2018.

[27] J. Li, F. Tao, Y. Cheng, and L. Zhao, "Big Data in product lifecycle management," *Int. J. Adv. Manuf. Technol.*, vol. 81, no. 1–4, pp. 667–684, 2015.

[28] V. Robin, B. Rose, and P. Girard, "Modelling collaborative knowledge to support engineering design project manager," *Comput. Ind.*, vol. 58, no. 2, pp. 188–198, 2007.

[29] A. Amrani, S. Zouggar, M. Zolghadri, and P. Girard, "Towards a collaborative approach to sustain engineer-to-order manufacturing," *2010 IEEE Int. Technol. Manag. Conf. ICE 2010*, pp. 1–8, 2010.

[30] K. Nagorny, S. Scholze, and A. W. Colombo, "A real-world application scenario for a novel collaborative ICT engineering platform," *Proc. - 2017 IEEE 15th Int. Conf. Ind. Informatics, INDIN 2017*, pp. 1175–1180, 2017.

[31] A. K. Kamrani and E. S. A. Nasr, "Product design and development framework in collaborative engineering environment," *Int. J. Comput. Appl. Technol.*, vol. 32, no. 2, p. 85, 2008.

[32] L. P. Saikia and S. Singh, "Feature extraction and performance measure of requirement engineering (RE) document using text classification technique," *Proc. 4th IEEE Int. Conf. Recent Adv. Inf. Technol. RAIT 2018*, pp. 1–6, 2018.

[33] R. Pinquié, P. Véron, F. Segonds, and N. Croué, "A collaborative requirement mining framework to support OEMs," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9320, pp. 105–114, 2015.

[34] Y. Ma and J. Y. H. Fuh, "Product lifecycle modelling, analysis and management," *Comput. Ind.*, vol. 59, no. 2–3, pp. 107–109, 2008.

[35] L. Montgomery, D. Damian, T. Bulmer, and S. Quader, "Author Correction: Customer support ticket escalation prediction using feature engineering," *Requir. Eng.*, vol. 23, no. 3, p. 1, 2018.

[36] J. Bragge, "Journal of the Association for Information Systems Engineering E-Collaboration Processes to Obtain Innovative End-User Feedback on Advanced Web-Based Information Systems * Engineering E-Collaboration Processes to Obtain Innovative End-User Feedback on Ad," *Jais*, vol. 10, no. March 2009, pp. 196–220, 2009.

[37] Restaurant Review _ Customer speaks _ Authenticate Indian Restaurant, "http://viceroyofindia.com/restaurant-reviews.htm" 2008.

[38] BBB Business Profile, DriveTime | Reviews and Complaints, "https://www.bbb.org/mbc/reviews/," 2019.

[39] E. Venkatesan and T. Velmurugan, "Performance analysis of decision tree algorithms for breast cancer classification.," *Indian J. Sci. Technol.*, vol. 8, no. 29, pp. 1–8, 2015.

[40] O. Abbas, "Comparisons Between Data Clustering Algorithms," *Int. Arab J. Inf. Technol.*, vol. 5, no. 3, pp. 320–325, 2008.

[41] ISO, "ISO 9000:2015, Quality management systems — Fundamentals and vocabulary," 2015.

[42] J. Kim, S. Cho, D. Park, K. Lee, C. Choi, and W. Cho, "Local Festival Marketing and Application Plan for Agricultural Products by Utilizing Big Data from Online Shopping Mall," *Proc. 2015 Int. Conf. Big Data Appl. Serv. - BigDAS '15*, pp. 233–236, 2015.

[43] M. Yang and S. Hu, "Research on the E-business platform of agricultural products and

rice marketing channel based on network big data," *Rev. Tec. la Fac. Ing. Univ. del Zulia*, vol. 39, no. 12, pp. 258–265, 2016.

[44] I. Kuzminov, P. Bakhtin, and A. Lavrynenko, "Big-data-augmented approach to emerging technologies identification: Case of agriculture and food sector," *CEUR Workshop Proc.*, vol. 1886, pp. 130–134, 2016.

[45] S. S. Gill, I. Chana, and R. Buyya, "IoT Based Agriculture as a Cloud and Big Data Service," *J. Organ. End User Comput.*, vol. 29, no. 4, pp. 1–23, 2017.

[46] S. K. Jones, A. K. Fremier, F. A. DeClerck, D. Smedley, A. O. Pieck, and M. Mulligan, "Big data and multiple methods for mapping small reservoirs: Comparing accuracies for applications in agricultural landscapes," *Remote Sens.*, vol. 9, no. 12, 2017.

[47] Y. Liu, "Innovation of E-commerce Fresh Agricultural Products Marketing Based on Big Internet Data Platform," *Agro Food Ind. Hi. Tech.*, vol. 28, no. 3, pp. 1739–1743, 2017.

[48] G. M. Alves and P. E. Cruvinel, "Big Data Environment for Agricultural Soil Analysis from CT Digital Images," *Proc. - 2016 IEEE 10th Int. Conf. Semant. Comput. ICSC 2016*, no. September, pp. 429–431, 2016.

[49] K. Sabarina and N. Priya, "Lowering data dimensionality in big data for the benefit of precision agriculture," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 548–554, 2015.

[50] F. Hao, X. Luo, and C. Mu, "Research on Key Technologies of Intelligent Agriculture Based on Agricultural Big Data," *Proc. - 2016 Int. Conf. Smart City Syst. Eng. ICSCSE 2016*, pp. 598–601, 2017.

[51] A. Abraham, K. Franke, and M. Köppen, *Intelligent systems design and applications*. 2003.

[52] M. R. Bendre, R. C. Thool, and V. R. Thool, "Big data in precision agriculture:

Weather forecasting for future farming," *Proc. 2015 1st Int. Conf. Next Gener. Comput. Technol. NGCT 2015*, no. September, pp. 744–750, 2016.

[53] N. Rao, "Big Data and Climate Smart Agriculture - Review of Current Status and Implications for Agricultural Research and Innovation in India," *SSRN Electron. J.*, 2017.

[54] Y. Fei, "Innovative Strategies of Agricultural Product Network Marketing under the Background of E-commerce," 2017.

[55] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, "A review on the practice of big data analysis in agriculture," *Comput. Electron. Agric.*, vol. 143, no. January, pp. 23–37, 2017.

[56] N. Xie, W. Wang, B. Ma, X. Zhang, W. Sun, and F. Guo, "Research on an Agricultural Knowledge Fusion Method for Big Data," *Data Sci. J.*, vol. 14, no. 7, pp. 1–9, 2015.

[57] P. Liu, G. Dai, and T. Fu, "A web services based email extension for remote monitoring of embedded systems," *Proc. - SNPD 2007 Eighth ACIS Int. Conf. Softw. Eng. Artif. Intell. Networking, Parallel/Distributed Comput.*, vol. 2, pp. 412–416, 2007.

[58] J. Zhan, C. L. Hsieh, I. C. Wang, T. S. Hsu, C. J. Liau, and D. W. Wang, "Privacy-preserving collaborative recommender systems," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 4, pp. 472–476, 2010.

[59] T. Bogers and A. Van Den Bosch, *Collaborative and content-based filtering for item recommendation on social bookmarking websites*, vol. 532. 2009.

[60] T. Bhuvaneswari, S. Prabaharan, and V. Subramaniyaswamy, "An effective prediction analysis using J48," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 8, pp. 3474–3480, 2015.

[61] W. Hengliang and Z. Weiwei, "A customer churn analysis model in e-business environment," *Int. J. Digit. Content Technol. its Appl.*, vol. 6, no. 9, pp. 296–302,

2012.

[62] M. S. Vyas and R. Gulwani, "Predictive analytics for E learning system," *2017 Int. Conf. Inven. Syst. Control*, pp. 1–4, 2017.

[63] US Healthcare data, "https://www.healthdata.gov/," 2018. .

[64] A. V. Solanki, "Data Mining Techniques Using WEKA classification for Sickle Cell Disease," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5857–5860, 2014.

[65] S. Y. Wani, M. M. Kirmani, and S. I. Ansarulla, "Prediction of Fake Profiles on Facebook using Supervised Machine Learning Techniques-A Theoretical Model," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7 (4), no. August, 2016.

[66] M. M. Rahman, "Mining Social Data to Extract Intellectual Knowledge," *Int. J. Intell. Syst. Appl.*, vol. 4, pp. 15–24, 2012.

[67] S. Pippal, L. Batra, A. Krishna, H. Gupta, and K. Arora, "Data mining in social networking sites: A social media mining approach to generate effective business strategies," *Int. J. Innov. Adv. Comput. Sci.*, vol. 3, no. 2, pp. 22–27, 2014.

[68] W. J. Xu, "Strategy study on China E-business service in big data era," *2013 Int. Forum Mater. Sci. Ind. Technol. IFMSIT 2013*, vol. 798, pp. 983–986, 2013.

[69] F. Gu, B. Ma, J. Guo, P. A. Summers, and P. Hall, "Internet of things and Big Data as potential solutions to the problems in waste electrical and electronic equipment management: An exploratory study," *Waste Manag.*, vol. 68, pp. 434–448, 2017.

[70] T. H. Clark and H. G. Lee, "Performance, interdependence and coordination in business-to-business electronic commerce and supply chain management," *Inf. Technol. Manag.*, vol. 1, pp. 85–105, 2000.

[71] J. T. Mentzer *et al.*, "Defining Supply Chain Management," *J. Bus. Logist.*, vol. 22, no. 2, pp. 1–25, 2001.

[72] D. J. Flint, E. Larsson, and B. Gammelgaard, "Exploring Processes for Customer

Value Insights, Supply Chain Learning and Innovation: an International Study,” *J. Bus. Logist.*, vol. 29, no. 1, pp. 257–281, 2008.

[73] M. Grean and M. J. Shaw, “Supply-Chain Partnership between P&G and Wal-Mart,” *E-bus. Manag.*, pp. 155–171, 2000.

[74] E. Hofmann, “Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect,” *Int. J. Prod. Res.*, vol. 55, no. 17, pp. 5108–5126, 2017.

[75] S. Tiwari, H. M. Wee, and Y. Daryanto, “Big data analytics in supply chain management between 2010 and 2016: Insights to industries,” *Comput. Ind. Eng.*, vol. 115, no. October 2017, pp. 319–330, 2018.

[76] D. frederick Ross, *Competing Through Supply Chain Management*, vol. New York. 1988.

[77] E. O. Makori, “Promoting innovation and application of internet of things in academic and research information organizations,” *Libr. Rev.*, vol. 66, no. 8–9, pp. 655–678, 2017.

[78] J. D. Frutos and D. Borenstein, “A framework to support customer-company interaction in mass customization environments,” *Comput. Ind.*, vol. 54, no. 2, pp. 115–135, 2004.

[79] ISO, “ISO 10002, Quality management — Customer satisfaction — Guidelines for complaints handling in organizations,” vol. 2014. Edmonton, 2014.

[80] ISO, “ISO/TS 10004, Quality management — Customer satisfaction — Guidelines for monitoring and measuring,” Elsevier Ltd, Edmonton, 2012.

[81] International Standard Organization, “ISO/IEC 17025 General requirements for the competence of testing and calibration laboratories,” *Int. Stand.*, vol. 2005, pp. 1–36, 2005.

[82] S. Q. L. T. Reports, S. Q. L. Support, and T. Information, *TECHNICAL REPORT ISO / IEC TR Information technology — Database languages — SQL Technical Reports — Information*, vol. 2015. 2015.

[83] Times of India, "Zettabyte pips petabyte as largest data storage unit," 2010.

[84] D. J. Flint and J. T. Mentzer, "LOGISTICS INNOVATION: A CUSTOMER VALUE-ORIENTED SOCIAL PROCESS," vol. 26, no. 1, 2005.

[85] S. K. Chandrasegaran *et al.*, "The evolution, challenges, and future of knowledge representation in product design systems," *CAD Comput. Aided Des.*, vol. 45, no. 2, pp. 204–228, 2013.

[86] Wikipedia, "Software Development Kit: https://en.wikipedia.org/wiki/Software_development_kit," 2018. .

[87] G. Kaur and A. Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes," *Int. J. Comput. Appl.*, vol. 98, no. 22, pp. 13–17, 2014.

[88] A.K. Varudharajulu and Y. Ma, "A Survey on Big Data Process Models for E-Business, E-Management, E-Learning, and E-Education," published in *International Journal of Innovative Research in Computer and Communication (IJIRCCE)*, pp. 9523–9553, 2018.

[89] R. T. Bayes, P. Laplace, and S. H. Jeffreys, "Bayes ' theorem." pp. 1–12, 2018.

[90] Ö. F. Arar and K. Ayan, "A feature dependent Naive Bayes approach and its application to the software defect prediction problem," *Appl. Soft Comput.*, vol. 59, pp. 197–209, 2017.

[91] K. M. Al-Aidaroos, A. A. Bakar, and Z. Othman, "Medical data classification with Naive Bayes approach," *Information Technology Journal*, vol. 11, no. 9. pp. 1166–1174, 2012.

[92] F. Song, S. Liu, J. Yang, and S. Xia, "Maximum scatter difference classifier and its

application of text categorization," *Jisuanji Gongcheng/Computer Eng.*, vol. 31, no. 5, 2005.

[93] Complaint category identification training dataset, "https://catalog.data.gov/dataset/consumer-complaint-database," 2018.

[94] J. Kevric, S. Jukic, and A. Subasi, "An effective combining classifier approach using tree algorithms for network intrusion detection," *Neural Comput. Appl.*, vol. 28, pp. 1051–1058, 2017.

[95] V. Chaurasia and S. Pal, "A Novel Approach for Breast Cancer Detection using Data Mining Techniques," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 2456–2465, 2014.

[96] CNC Mill machine completion detection training dataset, "https://www.kaggle.com/shasun/tool-wear-detection-in-cnc-mill," 2018.

[97] MechanicalFailureFittingTrainingData, "https://www.kaggle.com/binovi/mechanical-fitting-failure-data#mechanical_fitting_failures_2011_2018-03," 2018.

[98] Automobile price prediction training dataset, "https://www.kaggle.com/vipashakaul/predicting-the-price-of-an-automobile," 2018.

[99] US Healthcare data, "https://www.healthdata.gov/dataset/heart-disease-mortality-data-among-us-adults-35-stateterritory-and-county/resource/350fa186," 2014.

[100] T. Sorlie *et al.*, "Repeated observation of breast tumor subtypes in independent gene expression data sets," *Proc. Natl. Acad. Sci.*, vol. 100, no. 14, pp. 8418–8423, 2003.

[101] Canadian Environmental Sustainability Indicators (CESI), "https://open.canada.ca/data/en/dataset/bdee1482-8e13-4db6-adf4-1a2615254871," .

[102] T. Roles, "*The Air Quality Management System*," pp. 1–5, 2012.

[103] Canada Major Importers, "https://open.canada.ca/data/en/dataset/9d81bb46-de89-41be-b7a4-b76c08f96cff," 2016.

[104] S. Joshi, B. Pandey, and N. Joshi, “Comparative analysis of Naive Bayes and J48 Classification Algorithms,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 12, pp. 813–817, 2015.

[105] H. Seif, “Naïve Bayes and J48 Classification Algorithms on Swahili Tweets: Perfomance Evaluation,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 1, p. 1, 2016.

[106] T. R. Patil, “Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification,” *Int. J. Comput. Sci. Appl. ISSN 0974-1011*, vol. 6, no. 2, pp. 256–261, 2013.

[107] M. N. Amin and A. Habib, “Comparison of Different Classification Techniques Using WEKA for Hematological Data,” *Am. J. Eng. Res.*, no. 43, pp. 2320–847, 2015.

[108] T. C. Sharma and M. Jain, “WEKA Approach for Comparative Study of Classification Algorithm,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 4, pp. 1925–1931, 2013.

[109] A. Naik and L. Samant, “Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime,” *Procedia Comput. Sci.*, vol. 85, pp. 662–668, Jan. 2016.

[110] G. Li and F. Liu, “Sentiment analysis based on clustering: A framework in improving accuracy and recognizing neutral opinions,” *Appl. Intell.*, vol. 40, no. 3, pp. 441–452, 2014.

[111] A. Goyal and R. Mehta, “Performance Comparison of Naïve Bayes and J48 Classification Algorithms,” *Int. J. Appl. Eng. Res.*, vol. 7, no. 11 SUPPL., pp. 1389–1393, 2012.

[112] M. Meenakshi and G. Geetika, “Survey on Classification Methods using WEKA,” *Int. J. Comput. Appl.*, vol. 86, no. 18, pp. 16–19, 2014.

[113] N. Sharma, A. Bajpai, and R. Litoriya, “Comparison the various clustering algorithms

of weka tools," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 5, pp. 73–80, 2012.

[114] A. K. Varudharajulu and Y. Ma, "Feature-based Restaurant Customer Reviews Process Model using Data Mining," published in *Proceedings of the International Conference on Computing and Big Data (ICCBD)*, pp. 32–37, 2018.

[115] A. K. Varudharajulu and Y. Ma, "Feature-based Facebook Reviews Process Model for E-Management using Data Mining," published in *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E)*, 2019.

[116] V. S. Kumbhar, "Hotel Websites Facebook Data Analysis Using Weka," in *Second International Conference of Commerce and Management (ICCM)*, 2017.

[117] K. Dave, K. Dave, S. Lawrence, S. Lawrence, D. M. Pennock, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," *Proc. 12th Int. Conf. World Wide Web*, pp. 519–528, 2003.

[118] S. Wakade, C. Shekar, K. J. Liszka, and C.-C. Chan, "Text Mining for Sentiment Analysis of Twitter Data," *Ike*, p. 5, 2012.

[119] R. Duriqi, V. Raca, and B. Cico, "Comparative Analysis of Classification Algorithms on Three Different Datasets using WEKA," *Embed. Comput. (MECO), 2016 5th Mediterr. Conf. on. IEEE*, pp. 96–101, 2016.

[120] S. Revathi and T. Nalini, "Performance Comparison of Various Clustering Algorithm," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 2, pp. 2277–128, 2013.

[121] C. Science, "A Comparative Analysis of Classification Algorithms on Weather Dataset Using Data Mining Tool Article History," vol. 10, no. December, pp. 0–5, 2017.

[122] B. Chaudhari and M. Parikh, "A Comparative Study of clustering algorithms Using weka tools," *Inf. Technol. J.*, vol. 1, no. 2, pp. 154–158, 2012.

[123] M. Madan, "Using Mining Predict Relationships on the Social Media Network : Facebook (FB)," vol. 4, no. 4, pp. 60–63, 2015.

[124] B. R.-P. of the 5th annual A. web science and U. 2013, “Studying Facebook via data extraction: the Netvizz application,” *Dl.Acm.Org*, 2013.

[125] M. A. Russell, “Mining Facebook: Analyzing Fan Pages, Examining Friendships, and More,” *Min. Soc. Web Anal. Data from Facebook, Twitter, LinkedIn, Other Soc. Media Sites*, pp. 45–88, 2014.

[126] R. R. Bouckaert *et al.*, “WEKA Manual for Version 3-6-5,” *Univ. Waikato*, p. 327, 2010.

[127] P. Samaranayake, K. Ramanathan, and T. Laosirihongthong, “Implementing industry 4.0 - A technological readiness perspective,” in *IEEE International Conference on Industrial Engineering and Engineering Management*, 2018.

APPENDIX – A: C#.Net Developed for the Prototype

Class 1 - C#.Net code to collect Facebook reviews and pre-process them to make it compatible to WEKA.

```
public void GetJsonRequest(string reviewsFile)
{
        string requestUrl = "https://graph.facebook.com/195861177110141/ratings?fields=review_text,reviewer,created_time&access_token=EAAfVEgGTzg0BAMz9on5vKMLruwzV00dhgD8RBkRZBjB2oGzs0XZAeTF6pkidnHgD1MKMzgOfu3uVuJZAWKqlrX71zbizgyXzeRBFR8zr7XSE2sJdGI4oRELP3fNtGPVPPqJRVbXmDfb0y04fZCS8uuV5KDR0AMiav6VWHffHqwZDZD";
        string jsonOutput;

        WebRequest apiRequest = WebRequest.Create(requestUrl);
        HttpWebResponse apiResponse = (HttpWebResponse)apiRequest.GetResponse();

        if (apiResponse.StatusCode == HttpStatusCode.OK)
        {
                using (StreamReader sr = new StreamReader(apiResponse.GetResponseStream()))
                jsonOutput = sr.ReadToEnd();
                jsonOutput = "[" + jsonOutput + "]";
                jsonOutput.Replace(" ", String.Empty);

                var result = JsonConvert.DeserializeObject<List<Data>>(jsonOutput);

                using (FileStream fileW = new FileStream(reviewsFile, FileMode.Truncate, FileAccess.Write))
                {
                        // create a new stream to write to the file.
                        StreamWriter sw = new StreamWriter(fileW);
                        sw.Flush();
                        sw.Write("@relation reviews");
                        sw.Write(sw.NewLine + "@attribute comment string");
                        sw.Write(sw.NewLine + "@attribute rating real");
                        // The next two attribute declaration was used based on the application.
                        // Attribute declaration for restaurant feedback analysis.
                        sw.Write(sw.NewLine + "@attribute category{TASTE,FOOD,DRINKS,LOCATION,AMBIENCE,CLEANLINESS,OFFER,PRICE,SERVICE}");
                        // Attribute declaration for Durabuilt feedback analysis.
                        sw.Write(sw.NewLine + "@attribute category{SEAL,SCREEN,SLAB,LABOUR,SERVICE,INSTALL}");
                        sw.Write(sw.NewLine + "@data");
```

```
                    // Write a string to the file.
                    foreach (Reviews comments in result[0].data)
                    {
                        if (comments.review_text != null)
                        {


                            // Below code block is for data pre-processing.
                            comments.review_text =
                            Regex.Replace(comments.review_text, "[^a-zA-Z0-
                            9]+", " ");
                            comments.review_text =
                            comments.review_text.Replace("\n", String.Empty);
                            comments.review_text =
                            comments.review_text.Replace("\r", String.Empty);
                            comments.review_text =
                            comments.review_text.Replace("\t", String.Empty);
                            sw.Write(sw.NewLine + "'" + comments.review_text +
                            "',?");
                        }
                    }
                    sw.Flush();
                    sw.Close();
                    fileW.Close();
                }
            }
        }
```

Class 2 - C#.Net code to classify data using J48 data mining algorithm.

```
public void Classify(string trainingFile, string testFile, string resultFile)
    {
      // Training file is loaded using the below code.
      Instances trainingSplits = new Instances(
                                    new BufferedReader(
                                        new FileReader(trainingFile)));

      weka.filters.unsupervised.attribute.ClassAssigner classAssigner = new
      weka.filters.unsupervised.attribute.ClassAssigner();

      classAssigner.setClassIndex("last");

      weka.classifiers.meta.FilteredClassifier filteredClassifier = new
      weka.classifiers.meta.FilteredClassifier();

      weka.classifiers.trees.J48 j48tree = (weka.classifiers.trees.J48)new
      weka.classifiers.trees.J48();

      int lastIndex = trainingSplits.numAttributes() - 1;

      trainingSplits.setClassIndex(lastIndex);

      weka.filters.unsupervised.attribute.StringToWordVector stringToNominal = new
      weka.filters.unsupervised.attribute.StringToWordVector(50);

      for (int attIndex = 0; attIndex < trainingSplits.numAttributes(); attIndex++)
      {
        if (trainingSplits.attribute(attIndex).isString())
        {
            stringToNominal.setIDFTransform(true);
            stringToNominal.setTFTransform(true);
            stringToNominal.setStemmer(new
            weka.core.stemmers.IteratedLovinsStemmer());
            stringToNominal.setInputFormat(trainingSplits);
            stringToNominal.setTokenizer(new weka.core.tokenizers.WordTokenizer());
            trainingSplits = weka.filters.Filter.useFilter(trainingSplits, stringToNominal);
        }
      }

      j48tree.buildClassifier(trainingSplits);
      filteredClassifier.setClassifier(j48tree);
      filteredClassifier.setFilter(stringToNominal);
      filteredClassifier.toString();

      // Test file to be classified is loaded using the below code.
      Instances unlabeled = new Instances(
```

```
                                            new BufferedReader(
                                                new FileReader(testFile)));

  unlabeled.setClassIndex(unlabeled.numAttributes() - 1);

  Instances labeled = new Instances(unlabeled);

  for (int i = 0; i < unlabeled.numInstances(); i++)
  {
    double clsLabel = filteredClassifier.classifyInstance(unlabeled.instance(i));
    labeled.instance(i).setClassValue(unlabeled.classAttribute().value((int)clsLabel));
  }

  // Classification result is stored in the ".arff" file using the below code.
  BufferedWriter writer = new BufferedWriter(
                   new FileWriter(resultFile));
  writer.write(labeled.toString());
  writer.newLine();
  writer.flush();
  writer.close();
}
```

Class 3 - C#.Net code to cluster data using Simple K-means data mining algorithm.

```
public void Cluster(string classifiedFile, string bad, string neutral, string good)
    {
            ClusterEvaluation eval;
            Instances data;
            SimpleKMeans cl;
            // Test file to be clustered is loaded using the below code.
            ArffLoader loader = new ArffLoader();
            loader.setSource(new java.io.File(classifiedFile));
            data = loader.getDataSet();
            weka.filters.unsupervised.attribute.StringToNominal
            stringToNominalClustering;

        for (int attIndex = 0; attIndex < data.numAttributes(); attIndex++)
        {
          if (data.attribute(attIndex).isString())
          {
                stringToNominalClustering = new
                weka.filters.unsupervised.attribute.StringToNominal();
                stringToNominalClustering.setAttributeRange("first-last");
                stringToNominalClustering.setInputFormat(data);
                data = weka.filters.Filter.useFilter(data, stringToNominalClustering);
          }
        }
        cl = new SimpleKMeans();
        cl.setDistanceFunction(new weka.core.EuclideanDistance());
        cl.setPreserveInstancesOrder(true);
        cl.setNumClusters(3);
        cl.buildClusterer(data);
        eval = new ClusterEvaluation();
        eval.setClusterer(cl);
        eval.evaluateClusterer(new Instances(data));
        int[] assignments = cl.getAssignments();

        // Clustering result is stored in the ".arff" file using the below code.
        BufferedWriter badFile = new BufferedWriter(
                new FileWriter(bad));

        BufferedWriter neutralFile = new BufferedWriter(
                new FileWriter(neutral));

        BufferedWriter goodFile = new BufferedWriter(
                new FileWriter(good));


        for (int i = 0; i < assignments.Length; i++)
        {
```

```
            switch (assignments[i])
            {
               case 0:
                  {
                     badFile.write(data.instance(i).ToString());
                     badFile.newLine();
                     badInstancesCount++;
                     break;
                  }
               case 1:
                  {
                     neutralFile.write(data.instance(i).ToString());
                     neutralFile.newLine();
                     neutralInstancesCount++;
                     break;
                  }
               case 2:
                  {
                     goodFile.write(data.instance(i).ToString());
                     goodFile.newLine();
                     goodInstancesCount++;
                     break;
                  }
            }
         }
         badFile.newLine();
         badFile.flush();
         badFile.close();
         neutralFile.newLine();
         neutralFile.flush();
         neutralFile.close();
         goodFile.newLine();
         goodFile.flush();
         goodFile.close();
      }
   }
```

APPENDIX – B: Output (".arff") Files Generated From Restaurant Case Study

APPENDIX – B.1: Iteration 1 (Service Features Classification Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {TASTE,FOOD,DRINKS,LOCATION,AMBIENCE,CLEANLINESS,OFFER,PRICE,SERVICE}

@data

'Sri Lankan food......One of the happening place in Edmonton. Masala Dosa, Idli vada platter, Payasam, Filter coffee',4,TASTE

'The food was execellent, not over priced, huge portions. The wait time to get food was a bit.',5,FOOD

'Ghee Masala Dosa is excellent. Service was great and prices are good too',5,TASTE

'Delicious food, great prices, great service, a great place to get your fill of Srilankan&Indian cuisine. Spice level has always been perfect to order for both eat in or take out.',5,SERVICE

'It was amazing.. first ever time had Masala dosa outstanding. Very well served.. employees were friendly and they took time time explain us what things carry what... whenever I will go to Edmonton and I am surely stoping at this place.',5,TASTE

'Delicious food I am regular to this place and love the south indian food there',5,FOOD

'Authentic & delicious Sri lankan / South Indian food Extremely friendly and fast service

Would highly recommend this place',5,FOOD

'I love this place - regularly crave it Staff are always super helpful and food comes out fast. Tastes like my dads cooking',5,FOOD

'Quality of food is excellent. Prices are VERY reasonable. Ambience and the dishes they serve the food on resembles a cafeteria. Will definitely be making frequent visits.',5,SERVICE

'I love to go there..food is so tasty.. Love their masala dosa..',5,TASTE

'Amazing experience: Great food Great price and Great service',5,FOOD

'Last time I went with my friends service was very nice. And food was also good.',5,FOOD

'Awesome food.',5,FOOD

'Coconut Chutney its not good , makes me to feel like they are using shredded coconut but otherwise quality and customer service was good.',4,SERVICE

'Great food, good customer service and reasonable prices...',4,FOOD

'Good masala dosa. The sambar and rasam was also nice.',4,TASTE

'Delicious food. Decor could be improved',4,FOOD

'Very nice south indian food.',4,FOOD

'Please bring back the mint chutney and not serve the mango chutney with the dosa. We love the red, green, and white chutneys. Or at least give customers the option',3,TASTE

'Poor food quality,and service... But the guy who talk to me was so polite and humble.',2,FOOD

APPENDIX – B.2: Iteration 2 (Service Features Sentimental Analysis Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {TASTE,FOOD,DRINKS,LOCATION,AMBIENCE,CLEANLINESS,OFFER,PRICE,SERVICE}

@attribute sentiment {positive,negative}

@data

'Sri Lankan food......One of the happening place in Edmonton. Masala Dosa, Idli vada platter, Payasam, Filter coffee',4,TASTE,positive

'The food was execellent, not over priced, huge portions. The wait time to get food was a bit.',5,FOOD,positive

'Ghee Masala Dosa is excellent. Service was great and prices are good too',5,TASTE,positive

'Delicious food, great prices, great service, a great place to get your fill of Srilankan&Indian cuisine. Spice level has always been perfect to order for both eat in or take out.',5,SERVICE,positive

'It was amazing.. first ever time had Masala dosa outstanding. Very well served.. employees were friendly and they took time time explain us what things carry what... whenever I will go to Edmonton and I am surely stoping at this place.',5,TASTE,positive

'Delicious food I am regular to this place and love the south indian food there',5,FOOD,positive

'Authentic & delicious Sri lankan / South Indian food Extremely friendly and fast service Would highly recommend this place',5,FOOD,positive

'I love this place - regularly crave it Staff are always super helpful and food comes out fast. Tastes like my dads cooking',5,FOOD,positive

'Quality of food is excellent. Prices are VERY reasonable. Ambience and the dishes they serve the food on resembles a cafeteria. Will definitely be making frequent visits.',5,SERVICE,positive

'I love to go there..food is so tasty.. Love their masala dosa..',5,TASTE,positive

'Amazing experience: Great food Great price and Great service',5,FOOD,positive

'Last time I went with my friends service was very nice. And food was also good.',5,FOOD,positive

'Awesome food.',5,FOOD,positive

'Coconut Chutney its not good , makes me to feel like they are using shredded coconut but otherwise quality and customer service was good.',4,SERVICE,negative

'Great food, good customer service and reasonable prices...',4,FOOD,positive

'Good masala dosa. The sambar and rasam was also nice.',4,TASTE,positive

'Delicious food. Decor could be improved',4,FOOD,negative

'Very nice south indian food.',4,FOOD,positive

'Please bring back the mint chutney and not serve the mango chutney with the dosa. We love the red, green, and white chutneys. Or at least give customers the option',3,TASTE,negative

'Poor food quality,and service... But the guy who talk to me was so polite and humble.',2,FOOD,negative

APPENDIX – C: Output (".arff") Files Generated From Durabuilt Case Study

APPENDIX – C.1: Iteration 1.1 (Product and Service Reviews Classification Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {WINDOWS,DOORS,SERVICE}

@data

'I am a contractor and have been installing durabuilt windows. I even have them in my personal house. They are great windows. Yes we have had problems at times but they have always been very helpful and fix the problems that arise. Highly recommend them',5,WINDOWS

'Durabuilt Nope Nada Dont do it Customer service is a Zero Our windows were ordered through the Lethbridge store... our contractor wanted to try them... hasnt used them since. No one answers the phone during business hours. Have tried to have a window fixed (that the seal went on) since May . A gentleman did come and have a look at the window as well as the terrible caulking of the corners - he was shocked to see such shoddy work. We are in a new build community I will not be reccommending Durabuilt. (we received a phone call, early Winter a survey, to see how our Durabuilt experience was They had received paper work saying the warranty work was complete....)',1,WINDOWS

'Very happy with the service we are receiving from Durabuilt. They stand by their product

and do their best to accommodate. This company is all about customer service. No run arounds or hiding. Thank you to Ted and the rest of the team working on our file. You guys handle things like pros and we appreciate all the efforts',5,SERVICE

'I got windows installed in the late summer . In December one of the windows cracked from the cold weather. Durabuilt is now telling me that I have to pay for the window to be replaced. They are now telling me that there is only a month warranty on windows that develop a crack even though the package they gave me when I bought them said nothing about that. I have talked to other window companies that have a year warranty if their windows developed a pressure crack. It is pretty sad that one of the biggest window companies around cant guarantee their product for even one winter. How is it my fault that the weather in alberta gets cold in December. Very very disappointed and pissed off.',2,WINDOWS

'Did a new house build and windows had the seals fail the first year and the second year eighteen more failed. The first were covered material and labour but the eighteen now failed they want two thousand six hundred dollars for labour since the labour portion of the warranty expired only a few months ago. Who knows how long ago the seals actually failed since you cant really notice till it gets cold. They offered to cover half of the payment but even with that weve had twenty one windows fail in two years and am wondering how long will the rest last and how long will the replacements last',1,WINDOWS

'The new windows and doors from Durabuilt Windows&Doors are amazing. The entire process was so easy made possible by excellent trained staff and sales and installation crews. Gord was our project manager salesman. He was amazing and took the time to help us make the right decisions based on age of our home and what we wanted/needed. Gord clearly outlined the entire process and what would happen from start to finish during our project. He helped us make a few adjustments to windows and doors that would compliment the age of

the house as well as help with making it look spectacular. Gord took the persona of salesman and made it more like a friendship He was truly amazing and helpful in every way possible Thank you Gord After the windows and doors were decided upon and ordered we had a few weeks to wait for the install to be done. We received updates and confirmations about the process along the way, and the appointments were booked to install the new front door and five windows. The installers showed up on time and prepared to work They had everything with them that they needed and wasted no time in getting to work. The installers were professional, happy, funny and caring to our needs. They were a lot of fun to have around for the day doing this work. They did not take any shortcuts and made sure everything fit as perfectly as they could in a house that is over hundred years old The installers sealed the windows for airtight and watertight connections to allow no leaks. They fixed and matched the siding on the outside of the house to make it look like these were original windows installed. The installers also put in a new front door and worked relentlessly to ensure a good tight fit for security and again ensuring watertight and airtight seals. We were originally quoted two days of work to have five windows and one door replaced. The installers finished this all in eight hours and did an excellent job. Their camaraderie and pride in their work showed through in everything that they did for us and the house. They cleaned up all of their work as if they had never been there also Once completed, we were left in a total state of shock and awe over how amazing the new door and windows looked. From the inside it feels almost breathtaking. From the outside, it almost looks like the house got a new lease on its life and like a new house Very Happy and impressed I wouldnt think twice about contacting Durabuilt Windows and Doors for all my future needs. They seem to treat you more as family than as customers A very welcome surprise. Thank You Very MUCH Durabuilt',5,WINDOWS 'we are first time home buyers and are extremely happy and content with the window and door package selected with our builder. I love natural lighting at home as it uplifts your *spirits and my windows just do that great may it be summer or winter, our home feels cozy with the triple pane glass',5,WINDOWS*

'Ridiculous. Called MONTHS ago to get our broken window fixed, they had to come THREE times to measure, still ended up ordering the wrong window, STILL waiting even though they were so quick to take money from us before they even do any of the work for a new window they cannot seem to deliver. Avoid this mess and use someone else',1,WINDOWS

'We had a new bay window installed in our kitchen in . Over the winter and into the spring it leaked down into the drywall. We didnt really notice it until April when the drywall got soft and started to peel. We contacted Durabuilt immediately. We had three different service people over the next couple of months come out and look at it but the leak wasnt fixed until September. Each time someone came out we showed them the damage to the drywall and insulation and were assured Durabuilt would take care of it. It is now Dec 18. |We have been back and forth with Durabuilt over this. We sent them two quotes for the repair as requested. One was for seven eighty dollars and one for eleven hundred dollars. They offered us three hundred dollars. It wasnt until I threatened to go to court and to the media that offered to pay the seven eighty dollars. I have been waiting a week for the cheque. I was told it would be ready at two pm today by Amy Wong. I tried phoning Amy at pm to confirm the cheque was ready but she didnt answer. instead I got her answering machine. I left a message saying I was on my way there. I drove from one end of the city to the other through a snowstorm to pick up the cheque to find out the President had not signed it yet after a week of waiting. Amy promised me she would get it signed and delivered to me before five pm. She arrived at Three pm with the cheque and my wifes name was spelled wrong. Collin instead of Colleen. Doesnt anyone read paperwork. Now I have to wait until Monday to get the cheque. Christmas is coming and we are having family over and we have a damaged wall in the kitchen that we may not be able to get fixed before Christmas all because of the incompetence shown by

everyone involved in this. This should never have taken eight months to resolve. Social media is a wonderful tool to get a message out and I think I will be using it to let everyone I know and everyone they know about the incompetence of Durabuilt Windows. And by the way, it is not slander if it is true, it is a review',1,WINDOWS

'I am a contractor who recently directed one of my clients to the West End Showroom. This was my first mistake. The sales people were helpful although not as knowledgeable as you would expect. They sold my client on their top of the line Vivace windows and garden door. My client liked the product and was intending on changing out the windows in the entire house. The problems started when we made our order in September of , we were told it was a two month wait and were given a delivery date. Two days before delivery i called to see what time they were going to arrive on site. We were avoided for three days, no one would return my calls until eventually they admitted that they did not place the order at all and that the new delivery date was now December twenty four th. I refused to receive them until January as no one wants a contractor in their home on Christmas eve. Our order arrived in the first week in January and all looked OK. We installed the windows and the door but noticed that the handle for the door and the handle for the screen actually hit each other when in the closed position enough that it would damage the new screen. I called the salesman to bring the issue forward and was assured that it would be taken care of which of course hasnt happened. Trying to get in touch with anyone at Durabuilt is almost impossible. You need to call, email and text just to get them to respond and when they do it is always we will get back to you (which they never do). They have now said that is our problem and they will not do anything for us. I must say that I have not seen customer service this bad ever. I am not sure how they got on the list for the best managed companys but i am sure that customer service didnt count for anything. My next step is to lodge a complaint at the Better Business Burrow and hope that might help. I know that I will never refer any clients to Durabuilt in the future

and I will tell anyone who will listen how bad it really is at Durabuilt. It doesnt matter how good you think your product is, if you have this bad of service eventually you will run out of customers. I will never purchase from here again',1,SERVICE

'I live in a condo with Durabuilt windows. The seal on some of the windows has broken and the windows (glass panes) need replacing. The windows are still under warranty. The office in Winnipeg, Design Gallery, was contacted to do the warranty work. The people at this office dont seem to be interested in doing the work. Phone calls are not returned, no interest in setting up appointments for a quote and do the measurements. This has been going on for three months For one owner, last year, it took six months to receive her window installation because of incorrect measures and staff unable to do the installation. Totally unacceptable',1,WINDOWS

'We recently had a basement window enlarged and replaced. This was a small job so many of the competitors refused to provide me a quote. Durabuilt did a great job of arranging the concrete cut and install. The care in finishing was amazing and they did a fantastic job of clean up. Durabuilt had also replaced windows in a previous house we owned - fantastic job, loved the windows. I will call again for any future window needs',5,WINDOWS

'Patio Screen doors are crap. Had a new patio door installed in Oct 1five not even two months later the screen came apart from the frame. Piller to Post came out to look at the screen for Durabuilt and blame my dog. Not sure how my pup can reach the top of the screen door or the top sides of the screen for that matter (screen separated in seven spots most of which are above chest height). Door wont open sticks when the temp drops below zero degrees',1,DOORS

'We ordered our doors at the end of July. The first botched installation occurred on September. The temporary replacement doors were installed on October 31. Its now almost Christmas and we STILL dont have our doors. In addition to the time spent waiting for the

doors we ordered, weve spent countless hours chasing down customer service reps, installation managers and whoever else weve been referred to. And thats one of our biggest complaints about Durabuilt: there isnt a single, consistent point of contact to deal with our product/service issues. Instead we are referred to different individuals each time and we end up explaining all of the issues all over again. Its been a very frustrating and disappointing experience',1,DOORS

'I live in an older home and a couple of windiw companies said my new front door and windows were not possible, but Durabuilt figured out a solution and they look fabulous. I have since had a garden window replaced in my diningroom and once again was not disappointed. Great quality and great installation',5,WINDOWS

'So as a builder I have dealt with Durabuilt for a few years. The orders I have placed service has been good. Though my recent experience makes me question how they are one of the hundred best managed companies. So I needed a replacement slab for a project and had the order # and type of door. E-mailed salesman but it bounced back. So I phoned in. Sorry your salesman no longer with us. My response transfer me to another one. Sorry you need to talk to the order desk. Get a voice message saying he is on holidays. Back to receptionist and he says he is not but finds him. I speak to him Sorry cant help you I will have a salesman call you back. No one calls. Few days later driving by Southside stop in. Receptionist very helpful. But sorry we cant access West end records. I leave and next day stop at West End location. Salesman helps me named Donn. Told I will get an e-mail that day. E-mail next day as I have heard nothing. He calls me back and apologizes but until IT transfers the account o him he cant help me. Does this sound like a hundred best well manged companies. Problem is I need to match an existing door. Heck I even want them to quote another Duplex I am building',1,DOORS

'Bought a front entry door from Durabuilt in Edmonton. I picked the door up myself. When I

arrived to pick it up the door was not wrapped in any cardboard or other type of protector. I approached the warehouse logistics supervisor regarding this issue. This little man was rude and dismissive. He said I woyks of had to ask for protectant and pay for it. This was a three thousand five hundred dollars door The door was then carelessly wrapped in plastic wrap and thrown in the back of my truck. I was rushed away as others customers were waiting to pick up. The door was installed by my two professional carpenters who were not happy with the door as it was not square. Difficult to I talk. Then the lock system did not work properly. Durabuilt did fix that issue in a timely fashion. The installed door is off my 1/16" I notice this as I am aware. There is a small gapping hole at the bottom of the door between weather strip leaks cold air. I will not buy from Durabuilt again',2,DOORS

'Our small bungalow house had Durabuilt windows installed some ten years ago and ALL window seals failing...we were told it was a 'bad batch' and cost to replace would be eight thousand dollars plus labour at one hundred fifty dollars per hr',1,WINDOWS

'I have installed thousands of their products, inferior product with no customer service.Try plygem or fancy',1,SERVICE

'I am so upset and the service we received from durabuilt. It was recommended to us from my sister and brother in law, the installers were in and out in one day and everything was perfect. The complete opposite in our case. I wish for no one to have this happen I them. Please share and make everyone aware to absolutely under no circumstance go through with this company. Very lazy workers and no respect.',1,SERVICE

'I love my Windows The sales people were very knowledgeable. The installation was quick and professional. I would recommend them on the future for sure',5,WINDOWS

'Still waiting on a call back after they said they would over a week ago',1,SERVICE

'Most unprofessional company that I have ever come across. Would never recommend them to anyone. Poor service all around.',1,SERVICE

'I called Durabuilt for a window quote, when a rep came to do the measurements, refused to listen to what I wanted and treated me completely unprofessionally. I did not even recieve a quote after Will take my business elsewhere and will spread the word regarding my dissatisfaction.',1,SERVICE

'We are at our wits end with this company. One of our crankout screens broke as they strangely appear not to be fitting properly for the last year and a half. The screen broke at the end of June....here we are in OCTOBER with no screen. I call at least weekly and have been told yes the screen is finished....oh its actually in production... We finally had delivery, only to open the screen and laugh in frustration as it is COMICALLY small...clearly not for this window or ANY other window we have in our home. We are having more windows replaced in our home and because of all of this, durabuilt was not even on our list of companies to call for a quote. I just called today as I was told earlier in the week that our window was done and I would be called back with a delivery date. Shockingly...no call came. So I called today and was told it is not done but rather in production. I STILL had to pay for this new screen by the way and have not been offered a refund for all of this run around. I am beyond angry and would caution ANYONE thinking about going with this company. We have other issues with our sliders that I have not even gotten into with them yet as I want to get this sorted out first. RUN AS FAST AS YOU CAN AWAY FROM THIS COMPANY',1,WINDOWS

'Why would Big companies like to rip people off .. bought a house that have the windows installed by durabuilt .. reported one window through which we have wind coming iN causing entire house to be cold .. ok reported it and they took 4-six weeks to come fix the window .. back to -two weather and wind started to come again so they never fixed it properly causing the family to suffer cold .. including our babies .. any new house comes with warranties so why would you not do the proper repairs and yet I still paid for the cost of

repairs',1,WINDOWS

'The Manitoba Durabuilt team was absolutely fantastic to deal with. The customer service was outstanding. In fact it was so great that the sales rep personally came out to measure the old windows in my place, and provide tips and tricks for installing the new windows and doors as a part of my renovation. The finished product looks amazing and I could not be happier.',5,SERVICE

APPENDIX – C.2: Iteration 1.2 (Product and Service Reviews Sentimental Analysis Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {WINDOWS,DOORS,SERVICE}

@attribute sentiment {positive,negative}

@data

'I am a contractor and have been installing durabuilt windows. I even have them in my personal house. They are great windows. Yes we have had problems at times but they have always been very helpful and fix the problems that arise. Highly recommend them',5,WINDOWS,positive

'Durabuilt Nope Nada Dont do it Customer service is a Zero Our windows were ordered through the Lethbridge store... our contractor wanted to try them... hasnt used them since. No one answers the phone during business hours. Have tried to have a window fixed (that the seal went on) since May . A gentleman did come and have a look at the window as well as the terrible caulking of the corners - he was shocked to see such shoddy work. We are in a new build community I will not be reccommending Durabuilt. (we received a phone call, early Winter a survey, to see how our Durabuilt experience was They had received paper work saying the warranty work was complete....)',1,WINDOWS,negative

'Very happy with the service we are receiving from Durabuilt. They stand by their product and do their best to accommodate. This company is all about customer service. No run

arounds or hiding. Thank you to Ted and the rest of the team working on our file. You guys handle things like pros and we appreciate all the efforts',5,SERVICE,positive

'I got windows installed in the late summer . In December one of the windows cracked from the cold weather. Durabuilt is now telling me that I have to pay for the window to be replaced. They are now telling me that there is only a month warranty on windows that develop a crack even though the package they gave me when I bought them said nothing about that. I have talked to other window companies that have a year warranty if their windows developed a pressure crack. It is pretty sad that one of the biggest window companies around cant guarantee their product for even one winter. How is it my fault that the weather in alberta gets cold in December. Very very disappointed and pissed off.',2,WINDOWS,negative

'Did a new house build and windows had the seals fail the first year and the second year eighteen more failed. The first were covered material and labour but the eighteen now failed they want two thousand six hundred dollars for labour since the labour portion of the warranty expired only a few months ago. Who knows how long ago the seals actually failed since you cant really notice till it gets cold. They offered to cover half of the payment but even with that weve had twenty one windows fail in two years and am wondering how long will the rest last and how long will the replacements last',1,WINDOWS,negative

'The new windows and doors from Durabuilt Windows&Doors are amazing. The entire process was so easy made possible by excellent trained staff and sales and installation crews. Gord was our project manager salesman. He was amazing and took the time to help us make the right decisions based on age of our home and what we wanted/needed. Gord clearly outlined the entire process and what would happen from start to finish during our project. He helped us make a few adjustments to windows and doors that would compliment the age of the house as well as help with making it look spectacular. Gord took the persona of salesman

and made it more like a friendship He was truly amazing and helpful in every way possible Thank you Gord After the windows and doors were decided upon and ordered we had a few weeks to wait for the install to be done. We received updates and confirmations about the process along the way, and the appointments were booked to install the new front door and five windows. The installers showed up on time and prepared to work They had everything with them that they needed and wasted no time in getting to work. The installers were professional, happy, funny and caring to our needs. They were a lot of fun to have around for the day doing this work. They did not take any shortcuts and made sure everything fit as perfectly as they could in a house that is over hundred years old The installers sealed the windows for airtight and watertight connections to allow no leaks. They fixed and matched the siding on the outside of the house to make it look like these were original windows installed. The installers also put in a new front door and worked relentlessly to ensure a good tight fit for security and again ensuring watertight and airtight seals. We were originally quoted two days of work to have five windows and one door replaced. The installers finished this all in eight hours and did an excellent job. Their camaraderie and pride in their work showed through in everything that they did for us and the house. They cleaned up all of their work as if they had never been there also Once completed, we were left in a total state of shock and awe over how amazing the new door and windows looked. From the inside it feels almost breathtaking. From the outside, it almost looks like the house got a new lease on its life and like a new house Very Happy and impressed I wouldnt think twice about contacting Durabuilt Windows and Doors for all my future needs. They seem to treat you more as family than as customers A very welcome surprise. Thank You Very MUCH Durabuilt',5,WINDOWS,positive

'we are first time home buyers and are extremely happy and content with the window and door package selected with our builder. I love natural lighting at home as it uplifts your

spirits and my windows just do that great may it be summer or winter, our home feels cozy with the triple pane glass',5,WINDOWS,positive

'Ridiculous. Called MONTHS ago to get our broken window fixed, they had to come THREE times to measure, still ended up ordering the wrong window, STILL waiting even though they were so quick to take money from us before they even do any of the work for a new window they cannot seem to deliver. Avoid this mess and use someone else',1,WINDOWS,negative

'We had a new bay window installed in our kitchen in . Over the winter and into the spring it leaked down into the drywall. We didnt really notice it until April when the drywall got soft and started to peel. We contacted Durabuilt immediately. We had three different service people over the next couple of months come out and look at it but the leak wasnt fixed until September. Each time someone came out we showed them the damage to the drywall and insulation and were assured Durabuilt would take care of it. It is now Dec 18. |We have been back and forth with Durabuilt over this. We sent them two quotes for the repair as requested. One was for seven eighty dollars and one for eleven hundred dollars. They offered us three hundred dollars. It wasnt until I threatened to go to court and to the media that offered to pay the seven eighty dollars. I have been waiting a week for the cheque. I was told it would be ready at two pm today by Amy Wong. I tried phoning Amy at pm to confirm the cheque was ready but she didnt answer. instead I got her answering machine. I left a message saying I was on my way there. I drove from one end of the city to the other through a snowstorm to pick up the cheque to find out the President had not signed it yet after a week of waiting. Amy promised me she would get it signed and delivered to me before five pm. She arrived at Three pm with the cheque and my wifes name was spelled wrong. Collin instead of Colleen. Doesnt anyone read paperwork. Now I have to wait until Monday to get the cheque. Christmas is coming and we are having family over and we have a damaged wall in the kitchen that we may not be able to get fixed before Christmas all because of the incompetence shown by

everyone involved in this. This should never have taken eight months to resolve. Social media is a wonderful tool to get a message out and I think I will be using it to let everyone I know and everyone they know about the incompetence of Durabuilt Windows. And by the way, it is not slander if it is true, it is a review',1,WINDOWS,negative

'I am a contractor who recently directed one of my clients to the West End Showroom. This was my first mistake. The sales people were helpful although not as knowledgeable as you would expect. They sold my client on their top of the line Vivace windows and garden door. My client liked the product and was intending on changing out the windows in the entire house. The problems started when we made our order in September of , we were told it was a two month wait and were given a delivery date. Two days before delivery i called to see what time they were going to arrive on site. We were avoided for three days, no one would return my calls until eventually they admitted that they did not place the order at all and that the new delivery date was now December twenty four th. I refused to receive them until January as no one wants a contractor in their home on Christmas eve. Our order arrived in the first week in January and all looked OK. We installed the windows and the door but noticed that the handle for the door and the handle for the screen actually hit each other when in the closed position enough that it would damage the new screen. I called the salesman to bring the issue forward and was assured that it would be taken care of which of course hasnt happened. Trying to get in touch with anyone at Durabuilt is almost impossible. You need to call, email and text just to get them to respond and when they do it is always we will get back to you (which they never do). They have now said that is our problem and they will not do anything for us. I must say that I have not seen customer service this bad ever. I am not sure how they got on the list for the best managed companys but i am sure that customer service didnt count for anything. My next step is to lodge a complaint at the Better Business Burrow and hope that might help. I know that I will never refer any clients to Durabuilt in the future

and I will tell anyone who will listen how bad it really is at Durabuilt. It doesnt matter how good you think your product is, if you have this bad of service eventually you will run out of customers. I will never purchase from here again',1,SERVICE,negative

'I live in a condo with Durabuilt windows. The seal on some of the windows has broken and the windows (glass panes) need replacing. The windows are still under warranty. The office in Winnipeg, Design Gallery, was contacted to do the warranty work. The people at this office dont seem to be interested in doing the work. Phone calls are not returned, no interest in setting up appointments for a quote and do the measurements. This has been going on for three months For one owner, last year, it took six months to receive her window installation because of incorrect measures and staff unable to do the installation. Totally unacceptable',1,WINDOWS,negative

'We recently had a basement window enlarged and replaced. This was a small job so many of the competitors refused to provide me a quote. Durabuilt did a great job of arranging the concrete cut and install. The care in finishing was amazing and they did a fantastic job of clean up. Durabuilt had also replaced windows in a previous house we owned - fantastic job, loved the windows. I will call again for any future window needs',5,WINDOWS,positive

'Patio Screen doors are crap. Had a new patio door installed in Oct 1five not even two months later the screen came apart from the frame. Piller to Post came out to look at the screen for Durabuilt and blame my dog. Not sure how my pup can reach the top of the screen door or the top sides of the screen for that matter (screen separated in seven spots most of which are above chest height). Door wont open sticks when the temp drops below zero degrees',1,DOORS,negative

'We ordered our doors at the end of July. The first botched installation occurred on September. The temporary replacement doors were installed on October 31. Its now almost Christmas and we STILL dont have our doors. In addition to the time spent waiting for the

doors we ordered, weve spent countless hours chasing down customer service reps, installation managers and whoever else weve been referred to. And thats one of our biggest complaints about Durabuilt: there isnt a single, consistent point of contact to deal with our product/service issues. Instead we are referred to different individuals each time and we end up explaining all of the issues all over again. Its been a very frustrating and disappointing experience',1,DOORS,negative

'I live in an older home and a couple of windiw companies said my new front door and windows were not possible, but Durabuilt figured out a solution and they look fabulous. I have since had a garden window replaced in my diningroom and once again was not disappointed. Great quality and great installation',5,WINDOWS,positive

'So as a builder I have dealt with Durabuilt for a few years. The orders I have placed service has been good. Though my recent experience makes me question how they are one of the hundred best managed companies. So I needed a replacement slab for a project and had the order # and type of door. E-mailed salesman but it bounced back. So I phoned in. Sorry your salesman no longer with us. My response transfer me to another one. Sorry you need to talk to the order desk. Get a voice message saying he is on holidays. Back to receptionist and he says he is not but finds him. I speak to him Sorry cant help you I will have a salesman call you back. No one calls. Few days later driving by Southside stop in. Receptionist very helpful. But sorry we cant access West end records. I leave and next day stop at West End location. Salesman helps me named Donn. Told I will get an e-mail that day. E-mail next day as I have heard nothing. He calls me back and apologizes but until IT transfers the account o him he cant help me. Does this sound like a hundred best well manged companies. Problem is I need to match an existing door. Heck I even want them to quote another Duplex I am building',1,DOORS,negative

'Bought a front entry door from Durabuilt in Edmonton. I picked the door up myself. When I

arrived to pick it up the door was not wrapped in any cardboard or other type of protector. I approached the warehouse logistics supervisor regarding this issue. This little man was rude and dismissive. He said I woyks of had to ask for protectant and pay for it. This was a three thousand five hundred dollars door The door was then carelessly wrapped in plastic wrap and thrown in the back of my truck. I was rushed away as others customers were waiting to pick up. The door was installed by my two professional carpenters who were not happy with the door as it was not square. Difficult to I talk. Then the lock system did not work properly. Durabuilt did fix that issue in a timely fashion. The installed door is off my 1/16" I notice this as I am aware. There is a small gapping hole at the bottom of the door between weather strip leaks cold air. I will not buy from Durabuilt again',2,DOORS,negative

'Our small bungalow house had Durabuilt windows installed some ten years ago and ALL window seals failing...we were told it was a 'bad batch' and cost to replace would be eight thousand dollars plus labour at one hundred fifty dollars per hr',1,WINDOWS,negative

'I have installed thousands of their products, inferior product with no customer service.Try plygem or fancy',1,SERVICE,negative

'I am so upset and the service we received from durabuilt. It was recommended to us from my sister and brother in law, the installers were in and out in one day and everything was perfect. The complete opposite in our case. I wish for no one to have this happen I them. Please share and make everyone aware to absolutely under no circumstance go through with this company. Very lazy workers and no respect.',1,SERVICE,negative

'I love my Windows The sales people were very knowledgeable. The installation was quick and professional. I would recommend them on the future for sure',5,WINDOWS,positive

'Still waiting on a call back after they said they would over a week ago',1,SERVICE,negative

'Most unprofessional company that I have ever come across. Would never recommend them to anyone. Poor service all around.',1,SERVICE,negative

'I called Durabuilt for a window quote, when a rep came to do the measurements, refused to listen to what I wanted and treated me completely unprofessionally. I did not even recieve a quote after Will take my business elsewhere and will spread the word regarding my dissatisfaction.',1,SERVICE,negative

'We are at our wits end with this company. One of our crankout screens broke as they strangely appear not to be fitting properly for the last year and a half. The screen broke at the end of June....here we are in OCTOBER with no screen. I call at least weekly and have been told yes the screen is finished....oh its actually in production... We finally had delivery, only to open the screen and laugh in frustration as it is COMICALLY small...clearly not for this window or ANY other window we have in our home. We are having more windows replaced in our home and because of all of this, durabuilt was not even on our list of companies to call for a quote. I just called today as I was told earlier in the week that our window was done and I would be called back with a delivery date. Shockingly...no call came. So I called today and was told it is not done but rather in production. I STILL had to pay for this new screen by the way and have not been offered a refund for all of this run around. I am beyond angry and would caution ANYONE thinking about going with this company. We have other issues with our sliders that I have not even gotten into with them yet as I want to get this sorted out first. RUN AS FAST AS YOU CAN AWAY FROM THIS COMPANY',1,WINDOWS,negative

'Why would Big companies like to rip people off .. bought a house that have the windows installed by durabuilt .. reported one window through which we have wind coming iN causing entire house to be cold .. ok reported it and they took 4-six weeks to come fix the window .. back to -two weather and wind started to come again so they never fixed it properly causing the family to suffer cold .. including our babies .. any new house comes with warranties so why would you not do the proper repairs and yet I still paid for the cost of

repairs',1,WINDOWS,negative

'The Manitoba Durabuilt team was absolutely fantastic to deal with. The customer service was outstanding. In fact it was so great that the sales rep personally came out to measure the old windows in my place, and provide tips and tricks for installing the new windows and doors as a part of my renovation. The finished product looks amazing and I could not be happier.',5,SERVICE,positive

APPENDIX – C.3: Iteration 2.1 (Window Features Categorization Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {GLASS,SEAL,SCREEN,INSTALL,COST}

@data

'I am a contractor and have been installing durabuilt windows. I even have them in my personal house. They are great windows. Yes we have had problems at times but they have always been very helpful and fix the problems that arise. Highly recommend them',5,INSTALL

'Durabuilt Nope Nada Dont do it Customer service is a Zero Our windows were ordered through the Lethbridge store... our contractor wanted to try them... hasnt used them since. No one answers the phone during business hours. Have tried to have a window fixed (that the seal went on) since May . A gentleman did come and have a look at the window as well as the terrible caulking of the corners - he was shocked to see such shoddy work. We are in a new build community I will not be reccommending Durabuilt. (we received a phone call, early Winter a survey, to see how our Durabuilt experience was They had received paper work

saying the warranty work was complete....)',1,SEAL

'I got windows installed in the late summer . In December one of the windows cracked from the cold weather. Durabuilt is now telling me that I have to pay for the window to be replaced. They are now telling me that there is only a month warranty on windows that develop a crack even though the package they gave me when I bought them said nothing about that. I have talked to other window companies that have a year warranty if their windows developed a pressure crack. It is pretty sad that one of the biggest window companies around cant guarantee their product for even one winter. How is it my fault that the weather in alberta gets cold in December. Very very disappointed and pissed off.',2,GLASS

'Did a new house build and windows had the seals fail the first year and the second year eighteen more failed. The first were covered material and labour but the eighteen now failed they want two thousand six hundred dollars for labour since the labour portion of the warranty expired only a few months ago. Who knows how long ago the seals actually failed since you cant really notice till it gets cold. They offered to cover half of the payment but even with that weve had twenty one windows fail in two years and am wondering how long will the rest last and how long will the replacements last',1,COST

'The new windows and doors from Durabuilt Windows&Doors are amazing. The entire process was so easy made possible by excellent trained staff and sales and installation crews. Gord was our project manager salesman. He was amazing and took the time to help us make the right decisions based on age of our home and what we wanted/needed. Gord clearly outlined the entire process and what would happen from start to finish during our project. He helped us make a few adjustments to windows and doors that would compliment the age of the house as well as help with making it look spectacular. Gord took the persona of salesman and made it more like a friendship He was truly amazing and helpful in every way possible

Thank you Gord After the windows and doors were decided upon and ordered we had a few weeks to wait for the install to be done. We received updates and confirmations about the process along the way, and the appointments were booked to install the new front door and five windows. The installers showed up on time and prepared to work They had everything with them that they needed and wasted no time in getting to work. The installers were professional, happy, funny and caring to our needs. They were a lot of fun to have around for the day doing this work. They did not take any shortcuts and made sure everything fit as perfectly as they could in a house that is over hundred years old The installers sealed the windows for airtight and watertight connections to allow no leaks. They fixed and matched the siding on the outside of the house to make it look like these were original windows installed. The installers also put in a new front door and worked relentlessly to ensure a good tight fit for security and again ensuring watertight and airtight seals. We were originally quoted two days of work to have five windows and one door replaced. The installers finished this all in eight hours and did an excellent job. Their camaraderie and pride in their work showed through in everything that they did for us and the house. They cleaned up all of their work as if they had never been there also Once completed, we were left in a total state of shock and awe over how amazing the new door and windows looked. From the inside it feels almost breathtaking. From the outside, it almost looks like the house got a new lease on its life and like a new house Very Happy and impressed I wouldnt think twice about contacting Durabuilt Windows and Doors for all my future needs. They seem to treat you more as family than as customers A very welcome surprise. Thank You Very MUCH Durabuilt',5,INSTALL

'we are first time home buyers and are extremely happy and content with the window and door package selected with our builder. I love natural lighting at home as it uplifts your spirits and my windows just do that great may it be summer or winter, our home feels cozy with the triple pane glass',5,GLASS

'Ridiculous. Called MONTHS ago to get our broken window fixed, they had to come THREE times to measure, still ended up ordering the wrong window, STILL waiting even though they were so quick to take money from us before they even do any of the work for a new window they cannot seem to deliver. Avoid this mess and use someone else',1,COST

'We had a new bay window installed in our kitchen in . Over the winter and into the spring it leaked down into the drywall. We didnt really notice it until April when the drywall got soft and started to peel. We contacted Durabuilt immediately. We had three different service people over the next couple of months come out and look at it but the leak wasnt fixed until September. Each time someone came out we showed them the damage to the drywall and insulation and were assured Durabuilt would take care of it. It is now Dec 18. |We have been back and forth with Durabuilt over this. We sent them two quotes for the repair as requested. One was for seven eighty dollars and one for eleven hundred dollars. They offered us three hundred dollars. It wasnt until I threatened to go to court and to the media that offered to pay the seven eighty dollars. I have been waiting a week for the cheque. I was told it would be ready at two pm today by Amy Wong. I tried phoning Amy at pm to confirm the cheque was ready but she didnt answer. instead I got her answering machine. I left a message saying I was on my way there. I drove from one end of the city to the other through a snowstorm to pick up the cheque to find out the President had not signed it yet after a week of waiting. Amy promised me she would get it signed and delivered to me before five pm. She arrived at Three pm with the cheque and my wifes name was spelled wrong. Collin instead of Colleen. Doesnt anyone read paperwork. Now I have to wait until Monday to get the cheque. Christmas is coming and we are having family over and we have a damaged wall in the kitchen that we may not be able to get fixed before Christmas all because of the incompetence shown by everyone involved in this. This should never have taken eight months to resolve. Social media is a wonderful tool to get a message out and I think I will be using it to let everyone I know

and everyone they know about the incompetence of Durabuilt Windows. And by the way, it is not slander if it is true, it is a review',1,COST

'I live in a condo with Durabuilt windows. The seal on some of the windows has broken and the windows (glass panes) need replacing. The windows are still under warranty. The office in Winnipeg, Design Gallery, was contacted to do the warranty work. The people at this office dont seem to be interested in doing the work. Phone calls are not returned, no interest in setting up appointments for a quote and do the measurements. This has been going on for three months For one owner, last year, it took six months to receive her window installation because of incorrect measures and staff unable to do the installation. Totally unacceptable',1,SEAL

'We recently had a basement window enlarged and replaced. This was a small job so many of the competitors refused to provide me a quote. Durabuilt did a great job of arranging the concrete cut and install. The care in finishing was amazing and they did a fantastic job of clean up. Durabuilt had also replaced windows in a previous house we owned - fantastic job, loved the windows. I will call again for any future window needs',5,INSTALL

'I live in an older home and a couple of windiw companies said my new front door and windows were not possible, but Durabuilt figured out a solution and they look fabulous. I have since had a garden window replaced in my diningroom and once again was not disappointed. Great quality and great installation',5,INSTALL

'Our small bungalow house had Durabuilt windows installed some ten years ago and ALL window seals failing...we were told it was a 'bad batch' and cost to replace would be eight thousand dollars plus labour at one hundred fifty dollars per hr',1,COST

'I love my Windows The sales people were very knowledgeable. The installation was quick and professional. I would recommend them on the future for sure',5,INSTALL

'We are at our wits end with this company. One of our crankout screens broke as they

strangely appear not to be fitting properly for the last year and a half. The screen broke at the end of June....here we are in OCTOBER with no screen. I call at least weekly and have been told yes the screen is finished....oh its actually in production... We finally had delivery, only to open the screen and laugh in frustration as it is COMICALLY small...clearly not for this window or ANY other window we have in our home. We are having more windows replaced in our home and because of all of this, durabuilt was not even on our list of companies to call for a quote. I just called today as I was told earlier in the week that our window was done and I would be called back with a delivery date. Shockingly...no call came. So I called today and was told it is not done but rather in production. I STILL had to pay for this new screen by the way and have not been offered a refund for all of this run around. I am beyond angry and would caution ANYONE thinking about going with this company. We have other issues with our sliders that I have not even gotten into with them yet as I want to get this sorted out first. RUN AS FAST AS YOU CAN AWAY FROM THIS COMPANY',1,SCREEN

'Why would Big companies like to rip people off .. bought a house that have the windows installed by durabuilt .. reported one window through which we have wind coming iN causing entire house to be cold .. ok reported it and they took 4-six weeks to come fix the window .. back to -two weather and wind started to come again so they never fixed it properly causing the family to suffer cold .. including our babies .. any new house comes with warranties so why would you not do the proper repairs and yet I still paid for the cost of repairs',1,INSTALL

APPENDIX – C.4: Iteration 2.2 (Windows Features Sentimental Analysis Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {GLASS,SEAL,SCREEN,INSTALL,COST}

@attribute sentiment {positive,negative}

@data

'I am a contractor and have been installing durabuilt windows. I even have them in my personal house. They are great windows. Yes we have had problems at times but they have always been very helpful and fix the problems that arise. Highly recommend them',5,INSTALL,positive

'Durabuilt Nope Nada Dont do it Customer service is a Zero Our windows were ordered through the Lethbridge store... our contractor wanted to try them... hasnt used them since. No one answers the phone during business hours. Have tried to have a window fixed (that the seal went on) since May . A gentleman did come and have a look at the window as well as the terrible caulking of the corners - he was shocked to see such shoddy work. We are in a new build community I will not be reccommending Durabuilt. (we received a phone call, early Winter a survey, to see how our Durabuilt experience was They had received paper work saying the warranty work was complete....)',1,SEAL,negative

'I got windows installed in the late summer . In December one of the windows cracked from the cold weather. Durabuilt is now telling me that I have to pay for the window to be replaced. They are now telling me that there is only a month warranty on windows that

develop a crack even though the package they gave me when I bought them said nothing about that. I have talked to other window companies that have a year warranty if their windows developed a pressure crack. It is pretty sad that one of the biggest window companies around cant guarantee their product for even one winter. How is it my fault that the weather in alberta gets cold in December. Very very disappointed and pissed off.',2,GLASS,negative

'Did a new house build and windows had the seals fail the first year and the second year eighteen more failed. The first were covered material and labour but the eighteen now failed they want two thousand six hundred dollars for labour since the labour portion of the warranty expired only a few months ago. Who knows how long ago the seals actually failed since you cant really notice till it gets cold. They offered to cover half of the payment but even with that weve had twenty one windows fail in two years and am wondering how long will the rest last and how long will the replacements last',1,COST,negative

'The new windows and doors from Durabuilt Windows&Doors are amazing. The entire process was so easy made possible by excellent trained staff and sales and installation crews. Gord was our project manager salesman. He was amazing and took the time to help us make the right decisions based on age of our home and what we wanted/needed. Gord clearly outlined the entire process and what would happen from start to finish during our project. He helped us make a few adjustments to windows and doors that would compliment the age of the house as well as help with making it look spectacular. Gord took the persona of salesman and made it more like a friendship He was truly amazing and helpful in every way possible Thank you Gord After the windows and doors were decided upon and ordered we had a few weeks to wait for the install to be done. We received updates and confirmations about the process along the way, and the appointments were booked to install the new front door and five windows. The installers showed up on time and prepared to work They had everything

with them that they needed and wasted no time in getting to work. The installers were professional, happy, funny and caring to our needs. They were a lot of fun to have around for the day doing this work. They did not take any shortcuts and made sure everything fit as perfectly as they could in a house that is over hundred years old The installers sealed the windows for airtight and watertight connections to allow no leaks. They fixed and matched the siding on the outside of the house to make it look like these were original windows installed. The installers also put in a new front door and worked relentlessly to ensure a good tight fit for security and again ensuring watertight and airtight seals. We were originally quoted two days of work to have five windows and one door replaced. The installers finished this all in eight hours and did an excellent job. Their camaraderie and pride in their work showed through in everything that they did for us and the house. They cleaned up all of their work as if they had never been there also Once completed, we were left in a total state of shock and awe over how amazing the new door and windows looked. From the inside it feels almost breathtaking. From the outside, it almost looks like the house got a new lease on its life and like a new house Very Happy and impressed I wouldnt think twice about contacting Durabuilt Windows and Doors for all my future needs. They seem to treat you more as family than as customers A very welcome surprise. Thank You Very MUCH Durabuilt',5,INSTALL,positive

'we are first time home buyers and are extremely happy and content with the window and door package selected with our builder. I love natural lighting at home as it uplifts your spirits and my windows just do that great may it be summer or winter, our home feels cozy with the triple pane glass',5,GLASS,positive

'Ridiculous. Called MONTHS ago to get our broken window fixed, they had to come THREE times to measure, still ended up ordering the wrong window, STILL waiting even though they were so quick to take money from us before they even do any of the work for a new window

they cannot seem to deliver. Avoid this mess and use someone else',1,COST,negative

'We had a new bay window installed in our kitchen in . Over the winter and into the spring it leaked down into the drywall. We didnt really notice it until April when the drywall got soft and started to peel. We contacted Durabuilt immediately. We had three different service people over the next couple of months come out and look at it but the leak wasnt fixed until September. Each time someone came out we showed them the damage to the drywall and insulation and were assured Durabuilt would take care of it. It is now Dec 18. |We have been back and forth with Durabuilt over this. We sent them two quotes for the repair as requested. One was for seven eighty dollars and one for eleven hundred dollars. They offered us three hundred dollars. It wasnt until I threatened to go to court and to the media that offered to pay the seven eighty dollars. I have been waiting a week for the cheque. I was told it would be ready at two pm today by Amy Wong. I tried phoning Amy at pm to confirm the cheque was ready but she didnt answer. instead I got her answering machine. I left a message saying I was on my way there. I drove from one end of the city to the other through a snowstorm to pick up the cheque to find out the President had not signed it yet after a week of waiting. Amy promised me she would get it signed and delivered to me before five pm. She arrived at Three pm with the cheque and my wifes name was spelled wrong. Collin instead of Colleen. Doesnt anyone read paperwork. Now I have to wait until Monday to get the cheque. Christmas is coming and we are having family over and we have a damaged wall in the kitchen that we may not be able to get fixed before Christmas all because of the incompetence shown by everyone involved in this. This should never have taken eight months to resolve. Social media is a wonderful tool to get a message out and I think I will be using it to let everyone I know and everyone they know about the incompetence of Durabuilt Windows. And by the way, it is not slander if it is true, it is a review',1,COST,negative

'I live in a condo with Durabuilt windows. The seal on some of the windows has broken and

the windows (glass panes) need replacing. The windows are still under warranty. The office in Winnipeg, Design Gallery, was contacted to do the warranty work. The people at this office dont seem to be interested in doing the work. Phone calls are not returned, no interest in setting up appointments for a quote and do the measurements. This has been going on for three months For one owner, last year, it took six months to receive her window installation because of incorrect measures and staff unable to do the installation. Totally unacceptable',1,SEAL,negative

'We recently had a basement window enlarged and replaced. This was a small job so many of the competitors refused to provide me a quote. Durabuilt did a great job of arranging the concrete cut and install. The care in finishing was amazing and they did a fantastic job of clean up. Durabuilt had also replaced windows in a previous house we owned - fantastic job, loved the windows. I will call again for any future window needs',5,INSTALL,positive

'I live in an older home and a couple of windiw companies said my new front door and windows were not possible, but Durabuilt figured out a solution and they look fabulous. I have since had a garden window replaced in my diningroom and once again was not disappointed. Great quality and great installation',5,INSTALL,positive

'Our small bungalow house had Durabuilt windows installed some ten years ago and ALL window seals failing...we were told it was a 'bad batch' and cost to replace would be eight thousand dollars plus labour at one hundred fifty dollars per hr',1,COST,negative

'I love my Windows The sales people were very knowledgeable. The installation was quick and professional. I would recommend them on the future for sure',5,INSTALL,positive

'We are at our wits end with this company. One of our crankout screens broke as they strangely appear not to be fitting properly for the last year and a half. The screen broke at the end of June....here we are in OCTOBER with no screen. I call at least weekly and have been told yes the screen is finished....oh its actually in production... We finally had delivery,

only to open the screen and laugh in frustration as it is COMICALLY small...clearly not for this window or ANY other window we have in our home. We are having more windows replaced in our home and because of all of this, durabuilt was not even on our list of companies to call for a quote. I just called today as I was told earlier in the week that our window was done and I would be called back with a delivery date. Shockingly...no call came. So I called today and was told it is not done but rather in production. I STILL had to pay for this new screen by the way and have not been offered a refund for all of this run around. I am beyond angry and would caution ANYONE thinking about going with this company. We have other issues with our sliders that I have not even gotten into with them yet as I want to get this sorted out first. RUN AS FAST AS YOU CAN AWAY FROM THIS COMPANY',1,SCREEN,negative

'Why would Big companies like to rip people off .. bought a house that have the windows installed by durabuilt .. reported one window through which we have wind coming iN causing entire house to be cold .. ok reported it and they took 4-six weeks to come fix the window .. back to -two weather and wind started to come again so they never fixed it properly causing the family to suffer cold .. including our babies .. any new house comes with warranties so why would you not do the proper repairs and yet I still paid for the cost of repairs',1,INSTALL,negative

APPENDIX – C.5: Iteration 3.1 (Door Features Categorization Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {PatioDoor,EntryDoor,ConnectYou}

@data

'Patio Screen doors are crap. Had a new patio door installed in Oct 1five not even two months later the screen came apart from the frame. Piller to Post came out to look at the screen for Durabuilt and blame my dog. Not sure how my pup can reach the top of the screen door or the top sides of the screen for that matter (screen separated in seven spots most of which are above chest height). Door wont open sticks when the temp drops below zero degrees',1,PatioDoor

'We ordered our doors at the end of July. The first botched installation occurred on September. The temporary replacement doors were installed on October 31. Its now almost Christmas and we STILL dont have our doors. In addition to the time spent waiting for the doors we ordered, weve spent countless hours chasing down customer service reps, installation managers and whoever else weve been referred to. And thats one of our biggest complaints about Durabuilt: there isnt a single, consistent point of contact to deal with our product/service issues. Instead we are referred to different individuals each time and we end up explaining all of the issues all over again. Its been a very frustrating and disappointing experience',1,ConnectYou

'So as a builder I have dealt with Durabuilt for a few years. The orders I have placed service

has been good. Though my recent experience makes me question how they are one of the hundred best managed companies. So I needed a replacement slab for a project and had the order # and type of door. E-mailed salesman but it bounced back. So I phoned in. Sorry your salesman no longer with us. My response transfer me to another one. Sorry you need to talk to the order desk. Get a voice message saying he is on holidays. Back to receptionist and he says he is not but finds him. I speak to him Sorry cant help you I will have a salesman call you back. No one calls. Few days later driving by Southside stop in. Receptionist very helpful. But sorry we cant access West end records. I leave and next day stop at West End location. Salesman helps me named Donn. Told I will get an e-mail that day. E-mail next day as I have heard nothing. He calls me back and apologizes but until IT transfers the account o him he cant help me. Does this sound like a hundred best well manged companies. Problem is I need to match an existing door. Heck I even want them to quote another Duplex I am building',1,ConnectYou

'Bought a front entry door from Durabuilt in Edmonton. I picked the door up myself. When I arrived to pick it up the door was not wrapped in any cardboard or other type of protector. I approached the warehouse logistics supervisor regarding this issue. This little man was rude and dismissive. He said I woyks of had to ask for protectant and pay for it. This was a three thousand five hundred dollars door The door was then carelessly wrapped in plastic wrap and thrown in the back of my truck. I was rushed away as others customers were waiting to pick up. The door was installed by my two professional carpenters who were not happy with the door as it was not square. Difficult to I talk. Then the lock system did not work properly. Durabuilt did fix that issue in a timely fashion. The installed door is off my 1/16" I notice this as I am aware. There is a small gapping hole at the bottom of the door between weather strip leaks cold air. I will not buy from Durabuilt again',2,EntryDoor

APPENDIX – C.6: Iteration 3.2 (Doors Features Sentimental Analysis Output)

@relation reviews

@attribute comment string

@attribute rating numeric

@attribute category {PatioDoor,EntryDoor,ConnectYou}

@attribute sentiment {positive,negative}

@data

'Patio Screen doors are crap. Had a new patio door installed in Oct 1five not even two months later the screen came apart from the frame. Piller to Post came out to look at the screen for Durabuilt and blame my dog. Not sure how my pup can reach the top of the screen door or the top sides of the screen for that matter (screen separated in seven spots most of which are above chest height). Door wont open sticks when the temp drops below zero degrees',1,PatioDoor,negative

'We ordered our doors at the end of July. The first botched installation occurred on September. The temporary replacement doors were installed on October 31. Its now almost Christmas and we STILL dont have our doors. In addition to the time spent waiting for the doors we ordered, weve spent countless hours chasing down customer service reps, installation managers and whoever else weve been referred to. And thats one of our biggest complaints about Durabuilt: there isnt a single, consistent point of contact to deal with our product/service issues. Instead we are referred to different individuals each time and we end up explaining all of the issues all over again. Its been a very frustrating and disappointing experience',1,ConnectYou,negative

'So as a builder I have dealt with Durabuilt for a few years. The orders I have placed service has been good. Though my recent experience makes me question how they are one of the hundred best managed companies. So I needed a replacement slab for a project and had the order # and type of door. E-mailed salesman but it bounced back. So I phoned in. Sorry your salesman no longer with us. My response transfer me to another one. Sorry you need to talk to the order desk. Get a voice message saying he is on holidays. Back to receptionist and he says he is not but finds him. I speak to him Sorry cant help you I will have a salesman call you back. No one calls. Few days later driving by Southside stop in. Receptionist very helpful. But sorry we cant access West end records. I leave and next day stop at West End location. Salesman helps me named Donn. Told I will get an e-mail that day. E-mail next day as I have heard nothing. He calls me back and apologizes but until IT transfers the account o him he cant help me. Does this sound like a hundred best well manged companies. Problem is I need to match an existing door. Heck I even want them to quote another Duplex I am building',1,ConnectYou,negative

'Bought a front entry door from Durabuilt in Edmonton. I picked the door up myself. When I arrived to pick it up the door was not wrapped in any cardboard or other type of protector. I approached the warehouse logistics supervisor regarding this issue. This little man was rude and dismissive. He said I woyks of had to ask for protectant and pay for it. This was a three thousand five hundred dollars door The door was then carelessly wrapped in plastic wrap and thrown in the back of my truck. I was rushed away as others customers were waiting to pick up. The door was installed by my two professional carpenters who were not happy with the door as it was not square. Difficult to I talk. Then the lock system did not work properly. Durabuilt did fix that issue in a timely fashion. The installed door is off my 1/16" I notice this as I am aware. There is a small gapping hole at the bottom of the door between weather strip leaks cold air. I will not buy from Durabuilt again',2,EntryDoor,negative