إقــــرار

أنا الموقع أدناه مقدم الرسالة التي تحمل العنوان:

استخدام طريقة التنقيب عن الآراء لتحسين المنتج

Using Opinion mining method for product improvement

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

DECLARATION

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification

Student's name: اسم الطالب: إيمان سمير سليمان ياسين

Signature: التوقيع:

Date: التاريخ: 2014 / 11 / 17

Islamic University of Gaza
Deanery of Graduate Studies
Faculty of Information Technology
Information Technology Program

Using Opinion Mining Method for Product Improvement

By:
EmanYassin

Supervised By:
Dr. Alaa El Halees

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master in Information Technology

September 2014

بســم الله الرحمن الرحيم

الجامعة الإسلامية – غزة
The Islamic University - Gaza

مكتب نائب الرئيس للبحث العلمي والدراسات العليا **هاتف داخلي:** 1150

ج س غ/35/

الرقم ....2014/10/01م Ref

التاريخ .... Date

نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحثة/ **إيمان سمير سليمان ياسين** لنيل درجة الماجستير في **كلية *تكنولوجيا المعلومات*** **برنامج تكنولوجيا المعلومات** وموضوعها:

استخدام طريقة التنقيب عن الآراء لتحسين المنتج

Using Opinion mining method for product improvement

وبعد المناقشة التي تمت اليوم الأربعاء 07 ذو الحجّة 1435هـ، الموافق 2014/10/01م الساعة العاشرة والنصف صباحاً، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

د. علاء مصطفى الهليس	**مشرفاً ورئيساً**	....
د. أشرف محمد العطّار	**مناقشاً داخلياً**	....
د. ناجي شكري الظاظا	**مناقشاً خارجيًا**	....

وبعد المداولة أوصت اللجنة بمنح الباحثة درجة الماجستير في **كلية *تكنولوجيا المعلومات*/ برنامج تكنولوجيا المعلومات.**

واللجنة إذ تمنحها هذه الدرجة فإنها توصيها بتقوى الله ولزوم طاعته وأن تسخر علمها في خدمة دينها ووطنها.

والله ولي التوفيق،،،

مساعد نائب الرئيس للبحث العلمي والدراسات العليا

١٢/١٠/٢٠١٤

أ.د. فؤاد علي العاجز

شئون البحث العلمي والدراسات العليا
Research & Graduate Affairs

قالوا سبحانك لا علم لنا إلا ما علمتنا إنك أنت العليم الحكيم

وفوق كل ذي علم عليم

قالوا سبحانك لا علم لنا إلا ما علمتنا إنك أنت العليم الحكيم

Dedication

To my beloved Parents …

To my beloved husband and his family…

To my sons Mustafa, Nada, Mohammed …

To my sisters and brothers…

To my best friends…

Acknowledgment

Praise is to ***Allah****, the Almighty for having guided me at every stage of my life.*

It gives me pleasure to thank my supervisor ***Prof. Alaa El-Halees****, without his help, guidance, and continuous follow-up ; this research would never have been.*

In addition, I would like to extend my thanks to the ***academic staff of the Faculty of Information Technology*** *and* ***colleagues*** *who helped me during my master's study and taught me different courses.*

I would also like to thank ***Dr. Abed Aschokry****,* ***Dr. Rawia Radi*** *for their valuable scientific and technical notes also I extend my special thanks to* ***Mr. Mohammed El-Khoudary****.*

Finally, Thanks due to my ***Parents****,* ***Husband*** *and my sisters* ***Arwa*** *,* ***Ansam*** *for their support and kept me productive all time.*

Eman S. S. Yassin
2014

Abstract

The internet contains an increasing number of online forums where customers exchange products opinions. Online customer reviews is considered as a significant informative resource which is useful for both potential customers and product manufactures. In this research, we have studied the opinion mining model to improve the product quality through customer opinions and obtain customer satisfaction improvement. Opinions and reviews can be easily expressed on the web such as in merchant sites, review portals, blogs, Internet forums and much more rather than a traditional market survey that seeks out customer opinions with questioners or interviews. This data reveals that both product manufactures and potential customers are very interested in this customer feedback or customer opinions. This would provide a knowledge of customers likes and dislikes on the products. This gives a better knowledge of the product quality , limitation and advantages over competitors.

The present work is carried out by using opinion mining in the proposed methodology to improve the product quality through customer opinions.

The proposed methodology is based on six sigma methodology. It consists of many steps which are: Define the problem , Measure the customers opinion by classify the opinions into positive and negative classes, Analyze the negative to extract the product feature, finally improve the feature by matching the feature from the product and next version for the same product. To evaluate our work, we chose Samsung Galaxy S2 as case study. We used it as the current product which will be improved through the customer opinion and regard that Samsung Galaxy S3 is the output product and the improved version As a result of our research, we got a set of features in Galaxy S2 which considered as negative by customer. Then we match with the positive for Galaxy S3. It is noticed that the customers complain and product defects are reduced. This is due to negative feature has significantly improved on the quality product, whereas the positive features are excluded.

Keywords: opinion mining, product quality improvement, negative reviews, Six sigma DMAIC methodology .

عنوان البحث

استخدام طرق التنقيب عن الآراء في تحسين المنتج

الملخص

يحتوي الإنترنت على عدد متزايد من المنتديات التي تشمل آراء الزبائن للمنتجات المصنعة. تعتبر هذه الآراء كمورد كبير للمعلومات التي تفيد كلا من الزبائن المحتملين وصانعي المنتجات. في هذا البحث قمنا بدراسة نموذج التنقيب عن الآراء لتحسين جودة المنتج والحصول على رضا الزبائن. يمكن التعبير عن الآراء بسهولة على شبكة الإنترنت مثل المواقع التجارية، المدونات، ومنتديات الإنترنت وأكثر من ذلك بدلا من العمل التقليدي الذي يسعى إلى آراء الزبائن بالاستبيانات أو المقابلات. هذه البيانات و الملاحظات مهمة لكل من صانعي المنتجات و الزبائن وهذا من شأنه توفير معرفة ماذا يحب او يكره الزبائن، اضافة الى التغذية الراجعة من هذه الآراء. كما انه يعطي معرفة أفضل لتحسين جودة المنتج للحصول على التميز أكثر من المنافسين.

في هذا العمل قمنا بتنفيذ طرق التنقيب عن الآراء في هذه المنهجية المقترحة لتحسين جودة المنتج بناء على اراء الزبائن .وتستند المنهجية المقترحة على منهجية ستة سيجما . وتتكون من عدة خطوات وهي :تحديد المشكلة، قياس رأي الزبائن من خلال تصنيفها إلى فئات الآراء الإيجابية والسلبية، تحليل الآراء السلبية لاستخراج ميزة المنتج، وأخيرا تحسين الميزة عن طريق مطابقة ميزة المنتج المقترح مع الإصدار القادم لنفس المنتج . لتقييم عملنا، اخترنا سامسونج جالاكسي S2 كدراسة حالة .نحن استخدمناه كالمنتج الحالي الذي سيتم تحسينه من خلال اراء الزبائن و باعتبار أن سامسونج جالاكسي S3هو المنتج المخرج ونسخة محسنة . ونتيجة لأبحاثنا، حصلنا على مجموعة من الميزات في جالاكسي S2 التي تعتبر سلبية عن طريق الزبائن .ثم تمت مطابقة الميزات مع إيجابية جالاكسي S3 و لقد لوحظ انه انخفضت شكاوي الزبائن و عيوب المنتج. هذا يرجع إلى تحسن السمات السلبية على جودة المنتج بشكل كبير، في حين استبعدت الميزات الإيجابية في هذا البحث.

الكلمات المفتاحية: التنقيب عن الآراء، تحسين جودة المنتج، الآراء السلبية، ومنهجية DMAIC ستة سيغما.

Table of contents

List of Figures

List of Tables

List of Abbreviations

DMAIC	Define, Measure, Analysis, Improve, Control
K-NN	K-Nearest Neighbor
NB	Naïve Bayes
SVM	Support Vector Machine
DT	Decision Tree
IE	Information Extraction
VOC	Voice Of Customer
DCP	Data Collection Plan
QFD	Quality Function Deployment
GE	General Electric
POS	Part Of Speech
VE	Value Engineering
CTQ	Critical To Quality
PDCA	Plan, Do, Check, Act
SIPOC	Supplier, Input ,Process, Output, Customers
CEO	Chief Executive Officer
SPC	Statistical Process Control
FMEA	Failure Mode and Effects Analysis

CHAPTER ONE

Chapter One
Introduction

In this chapter we will give an introduction to our thesis. It includes a historical background, a brief description of opinion mining, product quality improvement and six sigma DMIAC methodology which is the methodology we will depend on making our proposed methodology. In addition, it states the thesis problem, the research objectives , the significance of the thesis and the scope and limitation of the thesis work. Furthermore, research methodology and research format are discussed .

1.1 Background

Recently, people are becoming increasingly interested in sharing their personal complaints that could quickly proliferate through the Internet. Knowing products limitations and defects can be helpful in risk management, reducing future liabilities, as well as to make sound marketing strategies. These online reviews are also interesting to existing and potential customers too. It could significantly influence their purchase decisions of a product or in hiring a service [1].

Opinions and reviews can be easily posted on the web, where this leads to a rapid increasing scale of User Generated Contents (UGC), such as user feedback reviews, blogs, online forums, discussion groups, social media etc. Different from other kinds of online textual information, user generated contents are people-centric and contain a lot of subjective information [2]. Extracting these subjective texts and mining user opinions in the text is valuable to both customers and manufacturer. For customer, one could search on these UGC to find the opinion of existing users. They do not rely solely on opinions from friends and family when the search for a product review is simple and convenient. Customers are able to gather information anytime and anywhere. For manufacturer, apart from using traditional way such as customer surveys, they can gather customer reviews from these social media to make product improvements by understanding the likes and dislikes of customers. It provides them some information about any product defects at an early stage from customers [3][4][5].

1.1.1 Opinion mining

As electronic commerce (e-commerce) is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of customers reviews can be in hundreds. While traditional tool of collection opinion and customer review cannot be more efficiency nowadays, we need better alternative this is when opinion mining comes in. Opinion mining (also known as “sentiment classification”) is a subtask of text mining that automatically extracts knowledge from the various user-generated contents [6]. Basic task in opinion mining is about determining the subjectivity, polarity (positive or negative) and polarity strength (weakly positive, mildly positive, strongly positive, etc.) of a piece of text. Opinion mining has several important applications, such as determining critics’ opinions about a given product by classifying online product reviews, or tracking the shifting attitudes of the general public towards a political candidate by mining online forums or blogs [3][7][8].

The present work mainly focuses on opinion mining of online customer reviews and its investment to improve the product quality.

1.1.2 Product Quality Improvement

Product Quality Improvement concerns with improving the product quality that led to quality improvement approaches in manufacturing industry. Product quality is to give customer satisfaction by improving products and meet customer needs [9]. Product quality improvement is to improve company performance, product quality, worker productivity and labor management relations [10]. It also provides reduction in cost of quality, increased customer satisfaction, reduced customer complaints, reduced field returns, improved employee satisfaction level, improved supplier performance, and all the above in combination leading to improved market share [11]. Product quality is very important for any company to make better quality products. This is because, bad quality products will affect the customers confidence, image and sales of the company. In addition, the product quality is very important for customers. They are ready to pay high prices, but in return, they expect best-quality products. If they are not satisfied with the quality of product of company, they will purchase from the competitors. Nowadays, very good quality international products are available in the local market. So, if the domestic companies do not improve their products' quality, they will struggle to survive in the market [10].

1.1.3 Six sigma DMAIC methodology

Six Sigma was launched by Motorola in the late 70's and, in 1985, a quality engineer, Bill Smith, proposed the use of the sigma capacity as a common metric of quality measurement performance. Therefore, the name Six Sigma, based on the sigma capacity of a process, became a standard used in the organization as a goal for quality and as a guide for all improvement processes [12]. Six sigma is an organized methodology that guides to improve the organization products, services and process by continually reducing defects in the organization. It is a business strategy that focuses on improving customer requirements, business system, productivity and financial performance. The applications of six sigma principles to the organizations will succeed through senior management involvement, organizational commitment, cultural change and effective project management [13]. Six sigma allows organization to improve their product and services by designing and monitoring everyday business activities in ways that minimize waste and resources while increasing customer satisfaction. Six sigma guides organization into making fewer mistakes in everything they do from filling out a purchase order to manufacturing good products and eliminating poor quality at the earliest possible stage.

Six sigma DMAIC methodology include five phases of the process improvement methods which refer to the Definition, Measurement, Analysis, Improvement and Control. A detailed description of those phases will be given in chapter two. DMAIC has been widely used as the method for six sigma implementation projects in manufacturing, once its procedures are based on the well-known PDCA (Plan-Do-Check-Act) principles [14].

Six Sigma has a strategical approach for business improvement that differs completely from other quality like ISO 9001, Deming statistical quality control and Deming Plan, Do, Check, Act (PDCA). Also, Six Sigma emphasizes the achievement of financial returns and six sigma provides an organizational infrastructure consisting of key trained personnel for an effective implementation of this approach. In addition, six sigma emphasizes the data driven decision making approach instead of hypothesis [15].

1.2 Problem statement

In last two decades, the world has become totally dependent on the technology in everything of the life. Mobile phones companies seek to meet the needs of the target audience improve their product quality. So they conduct studies and devise the latest services to make its product easier for their customers in additional to keep them with attracting new ones. Customers services are constitute important part of the services mobile and condition of widening competition in the market rival up to crowding. Furthermore, the product lifecycles continue to shorten as customer quality expectations continue to increase. In order to remain competitive, business must continuously improve product quality and performance. To do those methods it is needed to extract customer opinion on a product. This can be done using opinion mining.

1.3 Objectives

1.3.1 Main objective

The main objective of this research is to use Opinion Mining methods for product quality improvement through customer opinion to obtain customer satisfaction.

1.3.2 Specific objectives

In relationship to problem statement, this research aims to:

- To utilize the opinion mining methods for product quality improvement.
- To verify the product quality improvement on the next version of the product.

1.4 Significance of the research

The significance of this research are:

- Opinion mining is used to maintain improvement of a product. This would produce long term stability and the defects of product reduced.
- The set of features extracted from the negative feedback can be utilized for e-commerce and many businesses' benefit. It can be taken into account in product quality improvement by understand what customer like and dislike in the product, product development plans and customer satisfaction improvement plans and developing marketing strategies, developing an ongoing marketing relationship with the customer.
- Saving efforts and time by helping the manufactures to find which features will be improved in the product that customer dislike it or they want to be more functional in this feature.

- Most customers express their opinions on various kinds of entities, such as products and services. These reviews not only provide customers with useful information for reference, but also are valuable for merchants to get the feedback from customers.
- Mining opinions from these vast amounts of reviews becomes urgent, and has attracted a lot of attentions from many researchers.

1.5 Scope and limitations of the research

This research propose a model to improve product quality through customer opinion. The work is also applied with some limitations and assumptions such as: This research focuses on mobile phone product, in specific (Samsung Galaxy S2) as a current product . The next version of the product is taken into consideration to use (Samsung Galaxy S3) . The classification methods is chosen as an opinion mining methods. English language was chosen for the data set. Amazon e-commerce website was the source of data set that used throughout this research .

1.6 Research Methodology

In order to achieve the objectives of the present work, we proposed a methodology that uses opinion mining method. The proposed methodology is based on Six Sigma DMAIC methodology (see Figure 1.1), which include the phases as follow:

- **Define the problem:** is to describe the problem statement, research goals, define of the customer as well as the voice of the customer (VOC).
- **Measure the opinions:** it will classify customer opinions into positive and negative classes using classification algorithm in Rapid Miner environment.
- **Analyze the negative opinions:** the negative opinions about a product feature from customer opinions will be extracted. This would be improved and given the most frequent features of customers' opinion using Stanford parser tool.
- **Improve the feature:** we try to match the negative opinion for the product in our case study with the positive opinion of the output product.

By applying the above steps hopefully a good results will be obtained to our case study. These steps will be describe in more detail in chapter four.

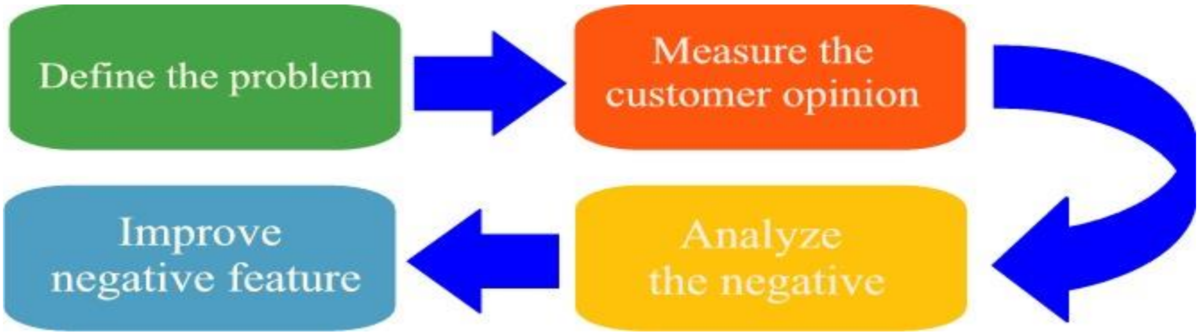


Figure (1.1) Overall The Methodology

1.7 Research Format

The research in general is organized as follows. Chapter one will present the introduction, research problem, objectives, scope and significant. Chapter two describes the theoretical model, product quality improvement, six sigma DMAIC methodology, opinion mining and part of speech tagging that required to the present work. Chapter three reviews the previous related work as literature survey. In chapter four, we proposed the methodology and the model that applied for the present work. Result and dissection as well as the evaluation of the model are included in this chapter. Conclusions and future work will be presented in chapter five.

CHAPTER TWO

Chapter two
Theoretical Foundation

This chapter will provide some knowledge relevant to the present work. We will introduce definition of product quality improvement, Six Sigma (DMAIC) methodology, opinion mining and Part of Speech.

2.1 Product quality improvement

An increasing concern with improving the product quality has led to the adoption of quality improvement approaches in manufacturing industry. These include product quality and quality improvement.

- **Product quality** means to incorporate features that have a capacity to meet consumer needs and gives customer satisfaction by improving products and making them free from any deficiencies or defects [9][16].
- **Product improvement** is to improve company performance, product quality, worker productivity and labor management relations. It is also necessary for any company for satisfaction of customers. Different process and product improvement methods are developed, the best product improvement approach is to interact with customers' needs and satisfaction by analyzing and understanding their likes and dislike. Different process and product improvement methods are developed and designed for measure the customer satisfaction [10]. The customer's satisfaction must therefore be measured in many different dimensions and to form the basis of quality improvements. Overall Customer Satisfaction Index (CSI) method is normally used to measure customer's satisfaction [17].
- **Quality improvement** a method for ensuring that all the activities necessary to design, develop and implement a product or service are effective and efficient with respect to the system and its performance. It is also a formal approach to the analysis of performance and systematic efforts to improve it. There are numerous models used for such as Plan, Do, Check, Act (PDCA), Six Sigma (DMAIC), Continuous Quality Improvement (CQI) and Total Quality Management (TQM) [18][11].
- **Product quality improvement** can be judged by indicators, such as, reduction in cost of quality, increased customer satisfaction, reduced customer complaints, reduced field returns, improved employee satisfaction level, improved supplier performance, and all the above in combination leading to improved market share [11].

Regarding to the above mentioned, we have focused on the interaction with customers feedback of the products. Thus, a few bad reviews give customers a reason to believe all good reviews. Fearing negative opinions or reviews is a mistake, however there are many case studies that show negative opinions, at least when mixed with positive ones, are a clear driver of sales. If all product opinions are glowing, people will be suspicious of their authenticity. Any customer-generated information about a product, whether it's entirely positive or slightly negative, helps increase sales, even products with average ratings convert better than products with no reviews. If the e-commerce web site provides a mix of positive and negative reviews, that shows two things : which are willing to give customers space to share their authentic opinions and provide feedback value. The little presence of negative feedback on web site about the product would reflect transparency in the product brand.

Negative feedback helps to improve everything from marketing to customer service to product design and also giving the opportunity to demonstrate the responsibility of the company [19][20][21]. Therefore, the present work focuses on the negative of customers feedback for product quality improvement.

Many product improvement methodologies are available and they are known by many names. Most use a structured approach to understanding the existing conditions, generate improvement ideas, then implement the changes. These are also designed in the most economical way to satisfy the needs of the customer. They also assume the current product or service fulfills the functional requirements of the market and the customer. Each of these product improvement methodologies look at the product through their own respective theory and tools. However, their particular perspective may or may not satisfy the customer needs [22][23].

2.2 Six Sigma DMAIC Methodology

Recently, many organizations have attempted to achieve customer satisfaction. One of the most important aspects of customer satisfaction is achieved through a high quality product, which also means a low defect product. Defects are usually included products containing a flaw in the manufacturing process, customer dissatisfaction in the service department. Six sigma is a methodology that relies on the scientific method to make significant reductions in customer defined defect rates. This effort would eliminate defects from every product, process, and transaction [1]. Six sigma success has also attributed to embracing it as an improvement strategy, philosophy and a way of doing business [24][25]. This section will discuss six sigma in terms of history, definition, applications, objectives and benefits

2.2.1 History of six sigma

In 1980s, Bill Smith, a senior engineer and scientist within Motorola's Communications Division, introduced the concept of six sigma in response to increasing complaints from the field sales force about warranty claims. It was a new method for standardizing the way defects are counted, with is sigma being near perfection. Smith crafted the original statistics and formulas that were the beginning of Motorola's six sigma methodology. He took his idea to CEO Bob Galvin ,who was struck by Smith's passion and came to recognize the approach as key to addressing quality concerns. six sigma became central to Motorola's strategy of delivering products that were fit for by customers [2]. So, Bill Smith considered as the father of six sigma [26]. Although six sigma began in "Motorola", its greatest successes have been in Allied Steel and General Electric GE. Following the recent merger of these two general organizations, General Electric has become the worldwide leader for six sigma all over the world [2][24][26].

2.2.2 Six Sigma Definition

Six Sigma is a business improvement approach that seeks to find and eliminate causes of mistakes or defects in business processes by focusing on process outputs that are of critical importance to customers [27]. The main feature of Six Sigma is to take the existing product, process and improves them in a better way. It is a powerful approach to achieve the financial goals for the organization and improving the company’s value [3]. As a metric, when a process is operating at Six Sigma level, it will produce defects at a rate of not more than 3.4 defects per one million opportunities. As a methodology, Six Sigma leads to business process improvement by focusing on understanding and managing customer expectations and requirements [28]. As a management system, Six Sigma is also used to ensure that critical improvement opportunity efforts developed through the metrics and methodology levels are aligned with the firm’s business strategy [29][30].

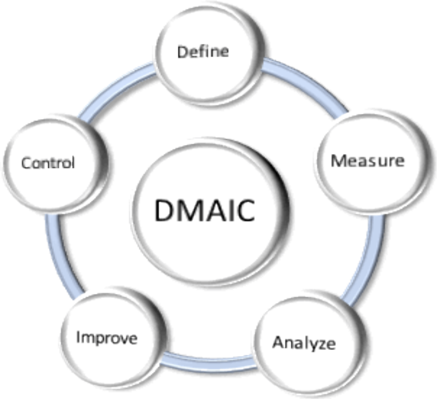


Figure (2.1) Six Sigma DMAIC Methodology from [23]

2.2.3 Six Sigma DMAIC Methodology Components

The DMAIC is a basic component of Six Sigma methodology. The DMAIC methodology has five phases Define, Measure, Analysis, Improvement and Control.

2.2.3.1 Define Phase

In this phase, we define the purpose of project, scope and process background for both internal and external customers. There are a different tools which is used in define phase like Supplier, Input, Process, Output, Customer (SIPOC), Voice of Customer (VOC) and Quality Function Deployment (QFD) [31] [32]. As a result, clear understanding of process improvement and how is it measure by the implementation of different tools. High level of process is also achieved. In addition, a lot of successful factors list show the customer requirement [31][33].

2.2.3.1.1 Definition of Customers

The customer of our product is an individual or business that purchases the goods or services produced by an institution. The customer is the end goal of business, where the customer who pays for supply and creates demand. Businesses will often compete through advertisements or sales in order to attract a larger customer base [34]. In our case study, the customers are stated and described in the next section.

2.2.3.1.2 Customer segmentation

Customer segmentation is the process of classifying people into groups that have some set of similar characteristics, resulting in the ability to be studied and targeted. The most basic method is to segment by simple demographics such as age, income, or marital status. The goal is to identify relatively homogeneous groups with similar behavior that will assist in customizing the message and/or offer for each segment [35].

- **In this study, we have classified the customers as follow:**
  - Geographical presence.
  - Demographic customer segmentation.
  - Psychographic segmentation.
  - Behavioral segmentation.
  - Psychological clause.
  - Socio – economic.

As an example of customer segmentation for Galaxy S2 was given by [36][37], as follows:

Geographical segmentation

As Geographical segmentation, they divide users to Urban and Rural as follows:

Urban users: Samsung targeting urban youth with many mobiles, Build in many mobile features like 3G, Wi-Fi, GPS and different OS.

Rural users: Rural users look for better brand image in the market like Samsung Brad Image. Samsung target rural users with special application for essential phones like flashlight, FM Radio, English messaging [37].

Demographic customer segmentation

Demographic customer segmentation involves study of some customer characteristics such as age, sex, profession, education.

(a) Demographic By Age:

Samsung Galaxy S2 target a few different audiences. The first audience is composed of young and social users. This audience is between the ages of 18-30 and work hard so that they have the budget for a smartphone. These users like to stay connected with their friends, family, and look at phones as a medium to do so. They like to text, use social network such as Facebook, Twitter, and LinkedIn. They also like to stay connected with friends by playing games with them such as Words with Friends or Draw Something. The Health Emphasis users can be a little older than the young/social users. They are typically 21-40 and their main goal is getting fit or maintaining their healthy lifestyle. They use their phone at the gym to be a music player or to watch videos to distract them from their intense workout. They also use their phone to find healthy recipes, work out tips, or even “virtual personal trainers”.

(b) Demographic By Gender:

Each gender interested to have Samsung Galaxy S2, because this mobile coming in different colors, Android OS more flexibility for each users. They, also, able to install and use many different applications for each gender.

(c) Demographic By Profession:

Samsung Galaxy S2 targeted professionals and people who work and stay for long time hours in their offices. The device is designed to enhance mobility and help them keep in touch with family and friends by making direct calls, sending messaging, chatting, and sharing pictures. Students are also included because the product helps them search for information quickly through easy access to websites. Businesspersons are included because the device can help them manage their schedule and receive updates and latest information about the market or banks.

- Psychographic segmentation

Samsung Galaxy S2 mainly focuses on lifestyles, hobbies, and interests of customers. Therefore, people are subdivided according to their values or lifestyles. Many research indicated that people within the same geographic or demographic group could exhibit extremely different psychographic cadres. The product was mainly targeted to people who spend most of their time on the internet, on social media platforms, or doing research. As a result, the product is developed with a 1.5 GHz processor, which extremely suitable for surfing. Besides that, psychographic segmentation was done to incorporate interests and hobbies of people who enjoy playing video games, listening to HD music, downloading HD movies, and doing business research.

- Behavioral segmentation

This would describe the loyalties of customers, Samsung Galaxy S2, behavioral segmentation is carried out the basis of usage rate and product features that useful to a customer. Good Brand Image from other electronic products. Samsung Galaxy S2 very fast, most importantly, a user has access to over 250 thousand apps via Google play [36]. Samsung Galaxy S2 will also be aimed universally at existing Samsung Galaxy customers and non-brand loyal consumers of competitors.

- Socio – economic

Most of the customers targeted to have Samsung Galaxy S2 are people who have middle and higher income, so that they might get smartphones.

2.2.3.2 Measure Phase

It is the second step of the six sigma methodology where a baseline measure is taken using actual data. The measure then becomes the origin from which the process can be improved. This develops measures or utilizes existing ones, such as SPC data or database information and pairs them accordingly with critical customer criteria, Data Collection Plan (DCP), Pareto diagrams and controls charts methods [38]. A Data Collection Plan (DCP) is then selected in the present work , where it includes the measures whose data needs to be collected , how much data to collect, data source, how will measure be implemented [39][40].

2.2.3.3 Analyze Phase

This phase identifies the root causes of problems and confirm them from data analysis [82]. There are different methods used for this phase are Regression Analysis, Design of Experiment and Process analysis. One of the most frequently used tools in the analyze step is the cause and effect diagram. It also determines how process inputs affect outputs, and give the best improvement [38][40].

2.2.3.4 Improve Phase

This phase develops the solution of problems in the process by implementing the different tools like FMEA and Pilot Plan. It can also eliminate the defects and able to produce the customer requirements [32]. As a result, identification of planned process that are implemented for improvement should reduce the impact and proposed solution for the problems. Improvement can be achieved and verified through many iterative and give facts with data and performance. Improvements should be selected based on probability of success, time to execute, impact on resources, and cost [32][33][40].

2.2.3.5 Control phase

This phase is used for data evolution of solution of problems and future plan and maintain the standing operating procedure. It also aims to reduce variation by controlling the inputs and monitoring the outputs of process. This would give analysis of data before and after, Well monitored system and Completed documentation of process results [31][40]. The main goal of this phase is to prevent – reoccurrence the defects and a similar problem a rises again [32].

2.2.4 Applications of six sigma

A number of manufacturing and service companies have recently realized that Six Sigma Methodology is flexible enough to be applied throughout all business functions. Examples of Six Sigma applications in different functional areas other than manufacturing operations will be discussed.

- **Sales and Marketing:**
  Many companies have considered using Six Sigma to improve marketing processes. For example, the marketing and sales organizations at GE and Dow have been using Six Sigma for new product development and customer support to reduce costs, improve performance and increase profitability [41]. Other companies use Six Sigma in marketing and sales as a road map to capture market data and competitive intelligence that will enable them to create products and services that meet customers' needs [42]. Companies combine Six Sigma methodology and online market research for better customer service could be improved their sales and marketing processes in order to get benefit [38].
- **Accounting and Finance:**
  The Six Sigma Methodology has contributed to reduce errors in invoice processing, reduction in cycle time, and optimized cash flow [28]. A number of companies have applied Six Sigma to the finance process to reduce variability in cycle times, error rates, costs, "days to pay" of accounts payable, and improve employees' productivity ratios [28][43]. Most studies have attempted to assess the impact of Six Sigma on financial performance that occurred at the aggregate industry level of analyses. However, few studies have reported of the impact of Six Sigma on the finance process itself [44].

2.2.5 Six sigma objective and benefits

The objective of Six Sigma is to enhance the performance measures that reflects the needs of the customer. In addition, the Six Sigma performance means a product defect rate of 3.4 per million opportunities for error. With Six Sigma methodology, the benefits of an organization include not only higher levels of quality but also lower levels of costs, higher customer loyalty, better financial performance and profitability of business [13].

2.3 Opinion Mining

Textual information includes two kinds: facts information and opinion information. Facts information is objective statements and opinion information is subjective statement that expresses persons’ opinion about objects. Most of researches on text information processing focus on mining and retrieval of facts information. But more and more researchers begin to become interested on mining of opinion information [5][45].

2.3.1 Opinion Mining Definition

It is the discipline of study that analyzes people's opinion, sentiment, evaluations, attitudes and emotions towards entities such a product, services, organization, individuals, issues, events topics and their attributes [46].

2.3.2 Document level, Sentence-level and Word level

A general method of opinion mining include document-level, sentence level and word level.

- **Document Level:** document opinion analysis is about classifying the overall sentiments expressed by the authors in the entire document text. The task is to determine whether a document is positive, negative or neutral about a certain object [47]. Document level polarity categorization attempts to classify sentiments in movie reviews, news articles, or web forum postings and blogs [48].

- **Sentence Level:** opinion mining at sentence level classifies each sentence as expressing a positive or a negative opinion. This is a more involved task than document-level classification, where less text means an error. It is an action that can be associated with two tasks. Initial work is to identify whether the sentence is subjective (opinionated) or objective. The second task is to classify a subjective sentence and determine if it is positive, negative or neutral [48].

- **word Level :** is a task of determining positive or negative sentiment of certain word in certain contexts or domain. Words that encode a desirable state (e.g. excellent) have a positive orientation, while words that represent an undesirable state have a negative orientation (e.g. bad). To apply opinion mining, researchers have compiled sets of words and phrases for adjectives, adverbs, verbs, and nouns. Such lists are collectively called the opinion lexicon [49].

  In our work we will concentrate on document-level where we can find negative opinion of a customer.

2.3.3 Classifying opinions into positive and negative

Semantic orientation determination is a task of determining whether a sentence or document has either positive or negative orientation. The approaches for this task consist of : unsupervised approach and supervised approach.

2.3.3.1 Unsupervised approach to sentiment classification

In an unsupervised approach, it will predict the semantic orientation of the documents based on the average semantic orientation of the adjective phrases and adverb phrases appearing in the documents. It also shows that the classifier is fed with some form of pre-annotated corpus, the training set, from which it can then train itself to classify unknown documents [49][50][51].

2.3.3.2 Supervised approach to sentiment classification

Another approach to sentiment classification is based on the supervised machine learning-based method. Using supervised approach, an annotator helps the classifier to learn how to classify document in a way answering questions from the classifier on how to annotate sentences. The task of sentiment classification can be considered as text classification task in which texts include several predefined categories using information from training texts. In the text classification task various machine learning methods have been applied, and they have proven successful [52]. In the supervised approach, the learning process is driven by the knowledge of the categories (positive/negative, in this task) and of the training instances that belong to them.

We use the supervised approach in classification of our dataset (opinions) as will be detailed later.

2.3.3.3 Classification

Classification is a supervised machine learning technique which is the process of finding a model that describes and distinguishes data classes for the purpose of being able to use the model. This model predicts the class of objects whose class label is unknown [53]. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects [54] [55].

Therefore, we have classified our dataset into different classifiers (e.g. positive or negative using methods such as Naïve Bayes (NB), Support vector machine (SVM), Decision tree (DT) and K-Nearest Neighbor (K-NN).

❖ Naïve Bayes (NB) Classifier

Naive Bayes classifiers is often used as a baseline in text classification, it predicts class membership probabilities [53]. A NB classifier assumes that the effect of an attribute value of a given class is independent of the values of the other attributes. This assumption is called class conditional independence [56]. The NB classifier, would be describe as follows [53][57]:

- Let D be training set of tuples and their associated class labels. As usual, each tuple is represented by a n-dimensional attribute vector, X = (x_1, x_2,….x_n), n measurements made on the tuple from n attribute, respectively, A_1,A_2,…, A_n.
- Assume that there are m classes, C_1, C_2,…, C_m. Given a tuple, X, the classifier will predict that X belongs to the class having the highest probability, conditioned on X. That is, the NB classifier predicts that tuple X belongs to the class C_i if and only if

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \le j \le m, j \ne i \qquad (2.1)$$

Thus, we maximize $P(C_i|X)$. The class C_i for which $P(C_i|X)$ is the maximized is called the maximum posteriori hypothesis. By Bayes theorem Equation 2.2.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{p(X)} \qquad (2.2)$$

As P(X) is constant for all classes, only $P(X|C_i)$ $P(C_i)$ needs maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equal.

- Based on the assumption is that attributes are conditionally independent (no dependence relation between attributes), $P(X|C_i)$ using Equation 2.3.

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) \qquad (2.3)$$

Equation 2.3 reduces the computation cost, only counts the class distribution. If A_k is categorical, $P(X_k|C_i)$ is the number of tuples in C_i having value x_k for A_k divided by $|C_i, D|$ (number of tuples of C_i in D). And if A_k is continuous-valued.

❖ K-Nearest-Neighbor Classifiers

K-NN is the most basic instance-based method, or lazy learning where the function is only approximated locally and all computation could be classified [58]. The k-nearest neighbor algorithms classify an instance according to a majority vote of its k most similar instances [53]. K is always a positive integer. Given a test instance x, its k closest neighbors y $_1$,…., y $_k$ are found and a vote is conducted to assign the most common class to x. That is, the class of x, denoted by c(x), is determined by Equation 2.4 [55].

$$C(x)| = \arg\max_{c\in C}\sum_{i=1}^{k}\delta(c, c(y_i)) \qquad (2.4)$$

Where c(yi) is the class of yi, and is a function that (u;v) = 1 if u = v. Usually "Closeness" is defined in terms of a distance metric, such as Euclidean distance where The Euclidean distance between two points, say, X1 = (x11, x12, …., x1n) and X2 = (x21, x22, ….., x2n), is determined by Equation 2.5 [8].

$$\mathrm{dist}(X1, X2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - X_{2i})^2} \qquad (2.5)$$

Distances for all the training samples are stored and nearest neighbor based on the K-th minimum distance is determined. Since the K-NN is supervised learning, get all the categories of your training data for the sorted value which fall under K. The predicted value is measured by using the majority of nearest neighbors. K-NN works well even when there are some missing data. K-NN is good at specified which predictions have low confidence. It also has some strong consistent results [53].

❖ Support Vector Machine

The Support Vector Machine (SVM) is a type of supervised machine learning algorithm with high generalization ability and used for both classifications and regression. SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output. For classification, SVM performs classes by finding the separating hyperplane between two classes in a high dimension feature space that maximize the separation margin between the two classes [57] the vectors that define the hyperplane are the support vectors. figure (2.8) illustrates that there are many linear classifiers (hyperplanes) which able to separate data into multiple classes. However, only one hyperplane achieves maximum separation. If such a hyperplane exists it's known as the maximum-margin hyperplane and such a linear classifier is known as a maximum margin classifier [59][60].

Equation 2.6 is dot product formula and used for the output of linear SVM, where **x** is a feature vector of classification documents composed of words, **w** is the weight of corresponding **x**, and **b** is a bias parameter determined by training process.

$$y = w.X - b \qquad (2.6)$$

SVM steps could be summarized by mapping the data to a predetermined very high-dimensional space via a kernel function. It also finds the hyperplane that maximizes the margin between the two classes. In addition, if data are not separable this would finds the hyperplane that maximizes the margin and minimizes the (a weighted average of the) misclassifications.

SVM can be used for both linear and nonlinear data. It uses a nonlinear mapping to transform the original training data into a higher dimension. With the new dimension, it searches for the linear optimal separating hyperplane (i.e., " decision boundary "). An appropriate nonlinear mapping to a sufficiently high dimension give data from two classes that separated by a hyperplane. SVM finds this hyperplane using support vectors ("essential" training tuples) and margins (defined by the support vectors). Figure (2.2) shows support vectors and how margins are maximized [60][61].

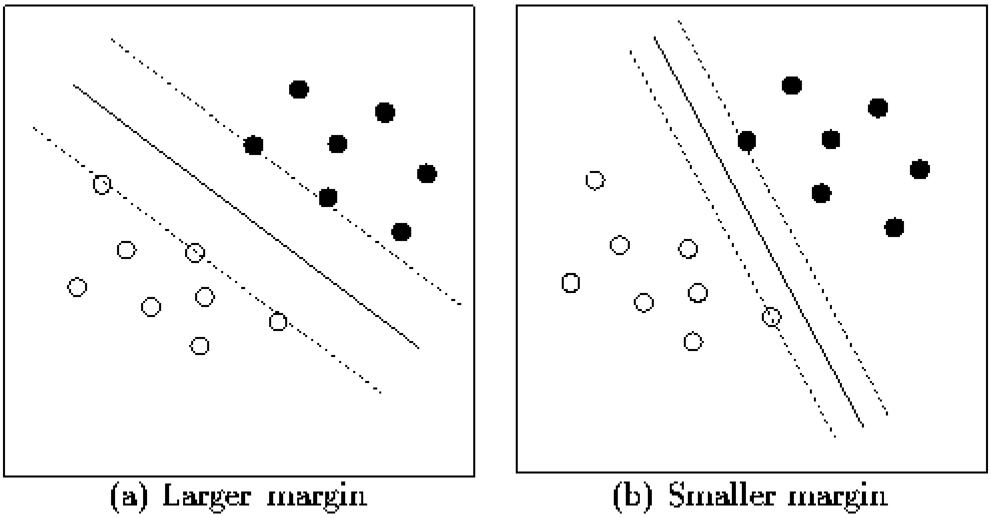


Figure (2.2) Support Vectors [60]

Moreover, SVM is effective for high dimensional data because the complexity of trained classifier is characterized by the number of support vectors rather than the dimensionality of the data. The support vectors are the essential or critical training examples, they lie closest to the decision boundary. If all other training examples are removed and the training is repeated, the same separating hyperplane would be obtained. The number of support vectors found can be used to compute an (upper) bound on the expected error rate of the SVM classifier, which is independent of the data dimensionality. Thus, an SVM with a small number of support vectors can have good generalization, even when the dimensionality of the data is high [4][53] [61].

- **Decision Tree (DT)**

Decision Tree is supervised machine learning, where it is an efficient method for producing classifiers from data. It is also a flow-chart-like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test and tree leaves represent classes. In addition, it is used in determining the optimum course of action, in situations having several possible alternatives with uncertain outcomes. A decision tree classifier is modeled in two phases: tree building and tree pruning. In tree building, the decision tree model is built by recursively splitting the training data set and assigning a class label to leaf by the most frequent class. Pruning a sub tree with branches if lower training error is obtained. Figure (2.3) presents decision tree algorithm [53].

Algorithm: Generate decision tree. Generate a decision tree from the training tuples of data partition D.

Input:

Data partition, D, which is a set of training tuples and their associated class labels;
attribute list, the set of candidate attributes;
Attribute selection method, a procedure to determine the splitting criterion that "best" partitions
the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split point or splitting subset.

Output: A decision tree.

Method:

```
(1) create a node N;
(2) if tuples in D are all of the same class, C then
(3) return N as a leaf node labeled with the class C;
(4) if attribute list is empty then
(5) return N as a leaf node labeled with the majority class in D; // majority voting
(6) apply Attribute selection method(D, attribute list) to find the "best" splitting criterion;
(7) label node N with splitting criterion;
(8) if splitting attribute is discrete-valued and
multiway splits allowed then // not restricted to binary trees
(9) attribute list attribute list splitting attribute; // remove splitting attribute
(10) for each outcome j of splitting criterion
// partition the tuples and grow subtrees for each partition
(11) let Dj be the set of data tuples in D satisfying outcome j; // a partition
(12) if Dj is empty then
(13) attach a leaf labeled with the majority class in D to node N;
(14) else attach the node returned by Generate decision tree(Dj, attribute list) to node N;
endfor
(15) return N;
```

Figure (2.3) Basic structure of Decision Tree algorithm [53]

2.3.4 Preprocessing

Data , at present , are highly susceptible to noise, missing, and inconsistent data due to their typically huge size and their likely origin from multiple, heterogeneous sources. Low-quality data will lead to low-quality mining results. Preprocessing is so important step for serious, effective, real-world data mining. There are a number of preprocessing techniques such as: cleaning, transformations, reduction, tokenization, stopword removal, and light stemming [57]. The following preprocessing process were used at the present work.

- **Tokenization**

Tokenization process is required to obtain all words that used in a given text of mining approaches. That is, a text document is split into a stream of words, phrases, symbols, or other meaningful elements called tokens. This is achieved by removing all punctuation marks and by replacing tabs and other non-text characters with single white spaces. The list of tokens becomes input for further processing such as parsing or text mining [62][63].

- **N-gram**

N-gram is a contiguous sequence of n items from a given sequence of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The n-grams typically are collected from a text or speech corpus [62][64].

- **Filter Stop words**

The stop words are meaningless of words in documents that are frequently occurred. Words like 'IT', 'AND' and 'TO' can be found virtually in every sentence in English-based documents, these words make very poor index terms. Users are indeed unlikely to ask for documents with these terms. Moreover, these words make up a large fraction of the text of most documents. The amount of information carried by these words is negligible. Consequently, it is usually worthwhile to ignore all stop word terms [65] [66].

- **Stemming**

Stemming is the process for reducing inflected words to their stem, base or root for generally a written word form. For example, the words "responsibilities" and "responsible" indicate the samething. There are many algorithms for stemming have been studied in computer science which differs in respect to performance and accuracy for example Porter algorithm, Lookup algorithm, Suffix-stripping algorithm, Lemmatization algorithm, Stochastic algorithm [47].

- **Transform cases**

It defines the basic process in most text mining that required to transform the words from the uppercase to a lowercase. This would not stop applying the analytics techniques to classify cluster and predict a new instance [64].

2.4 Part-of-Speech Tagging (POS)

POS Part Of Speech (POS) is a category used in linguistics that is define by a syntactic behavior of a word and plays a specific role in a sentence [67]. POS tagging is often used in sentiment analysis, especially due to the fact that it can be used in word-sense disambiguation. A strong correlation between the presence of adjectives and subjectivity in sentences has also been discovered [8]. Turney in [49] used POS tagging to construct conceptual sentence phrases, most of them including an adjective or adverb.

Product features are usually nouns or noun phrases in review sentences [68]. Thus the POS tagging is crucial and so important to information extraction [67][69]. We used the Stanford parser linguistic parser [70] to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is a noun, verb, adjective, etc). The process also identifies simple noun and verb groups (syntactic chunking). A part-of-speech tagger receives as input a plain text document, and returns as output document where every word and punctuation mark is associated with a tag that indicates the part of speech term. An example summarizes POS has been discussed in (section 4.3). Table (2.1) illustrates the noun and noun phrases pattern. Also, a definite linguistic filtering pattern is a noun phrase as the following patterns [71].

Table (2.1) Describes the noun and noun phrases patterns [71].

Noun and Noun phrases	
Noun	NN NNP (Proper noun) NNPS (Proper noun, plural) NNS (Plural)

2.4.1 Feature extraction

Features refer to all the components, qualities or physical characteristics of an object such as size, color, weight, speed, etc. It is also defined as features of products which customers have expressed their opinions on their reviews and feedbacks. Features refer to product features, product attributes, and/or product functions like the picture quality of the Canon IXUS 10, or the interior design of a Ford territory, or the service of hotel staff [5][68][72]. It is essential to readers that the features of the reviewed products are known as their areas of importance in different products may differ from people. For example, a reader might be more interested in the cleanliness of the hotel room, whilst the reviewer is more concerned with the quality of the customer service of the hotel staff [5][73]. Thus, Feature extraction is a process used to deduce possible product features and can be classified into two categories frequent and Infrequent features.

2.4.2 Frequent and Infrequent features extraction

Frequent features appear in majority of the customers opinions that most interested with a given product. The idea behind this technique is that feature appear on many opinions have more chance to be relevant, and therefore more likely to be a real product feature. However, infrequent feature is only appeared in a few number of customer opinion. These features can also be interesting to some potential customers [68][72].

2.5 Summary

In brief, this chapter includes an overview of theoretical knowledge that vital important to the present work. Product quality, product improvement and product quality improvement were provided in the first section. In section two we have described the six sigma DMAIC methodology and its history, definition, application as well as objectives and benefits. Section three explained the opinion mining where the definition, methods, supervised and unsupervised learning, classification and classification algorithms and preprocessing techniques were given in this section. In addition, Part of speech POS tagging, product feature extraction and frequent and infrequent features were discussed in the final section.

Next chapter will give works related to these principles and our research project.

CHAPTER THREE

Chapter Three
Related work

Our research on using opinion-mining method for product improvement. To the best of our knowledge, there are no previous work that studied opinion-mining method in product quality improvement using six sigma. So, this chapter will review related works in the two topics separately; using opinion mining in customer review in general and product quality improvement.

We will detail the historical development of opinion mining that used with many researchers in customer reviews. Therefore, many recent periodicals that relevant to the present work will be considered and given.

Many of the researches in Opinion Mining have been placing efforts on product features identification and finding opinion/sentiment orientation or making summarization of the amount opinions they gained. On the other hand, there are many process improvement methodologies are available and have many names. The related work is introduced and analyzed with respect to the thesis problem to show how far this work address to fulfill the present work objectives.

3.1 Using Opinion Mining in Customer Reviews

The following researches used opinion mining in customer reviews.

Hu and Liu in [74] proposed an opinion mining system performs the summarization in three main steps : (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results. These steps are performed in multiple sub-steps. The system first download (or crawls) all the reviews, and store it in a review database, after that POS tagger tags all the reviews which work for the mining part responsible for finding frequent features, then with the tagged sentence and features identified, opinion words are extracted and their semantic orientation are identified with the help of WordNet. The classification of sentiment depended on the words classified both as adjectives and adverbs in this step to produce a set of possible opinion words. In opinion word extraction opinion word are primarily used to express subjective opinions this system uses adjectives as opinion words. In addition, he are suggested a technique based on association rule mining to extract product features. The main idea is that people often use the same words when they comment on the same product features. Then frequent item sets of nouns in reviews are likely to be product features while the infrequent ones are less likely to be product features. This work also introduced the idea of using opinion words to find additional (often infrequent) features. The projects aims to summarize all the customer reviews of a product. This summarization task is different from traditional text summarization because it is interested in the specific features of the product that customers have opinions on and also whether the opinions are positive or negative. He do not summarize the reviews by selecting or rewriting a subset of the original sentences from the reviews to capture their main points as in the classic text summarization. The paper focuses on mining opinion/product features that the reviewers have commented on. A number of techniques were presented to mine such features. Experimental results show that these techniques are highly effective.

Lau et al. in [75] reported an automated analysis of the sentiments presented in online customers feedbacks that can facilitate both organizations' business strategy development and individual consumers' comparison shopping. He also proposed general system architecture of their Ontology Based Product Review Miner (OBPRM). The system describes a user first selects a product category and a specific product for opinion mining, based on the selected target product. The OBPRM system will use the Web services or APIs provided by e-Commerce sites and Internet Search Engines to retrieve the customer reviews for the particular product. Ontology extraction is carried offline and it must be performed before context-sensitive mining is conducted. The fuzzy domain ontology captures taxonomic information such as "iPhone" (product) "is-a" mobile phone (product category), and non-taxonomic relationship such as "screen" (product feature) is "associated with" "iPhone" (product). In addition, context-sensitive sentiment orientation (e.g., "excellent") of a product feature (e.g., "screen") is also captured in the fuzzy domain ontology. customer reviews, product ratings, and product descriptions can be retrieved from e-Commerce sites. This would illustrate the design, development, and evaluation of a novel fuzzy domain ontology based context- sensitive opinion mining system. Evaluated based on a benchmark dataset and real consumer reviews collected from Amazon.com, their system shows remarkable performance improvement over the context -free baseline.

Zhao et al. in [76] defined feature identification as the process used to deduce possible product features out of the tagged texts. Normally, the part-of-speech responsible for giving names to entities of the real world are nouns, in this case a noun gives name to the product and its features (i.e. zoom, battery life, image quality, etc.). In these works, they define two categories of features, frequent features and infrequent features. An ontology-based approach for opinion mining was proposed in their study. This describes the semantics of a domain in both human-understandable and computer-process able way. This would motivate by its success in the area of Information Extraction (IE). Their paper introduces a fine-grain approach for opinion mining, which uses the ontology structure as an essential part of the feature extraction process, by taking account the relations between concepts. The experiment result shows the benefits of exploiting ontology structure to opinion mining.

Kaiser et al. in [77] proposed a system which allows an automatic analysis of customer experience by combining methods from text mining and data mining . The system consists of four main components which are executed sequentially (1) selection (2) extraction (3) aggregation (4) analysis of product features . The evaluations and gained valuable information for product development are also considered . This approach falls into the category of feature-based opinion mining it not only considers the polarity of product features, but also the intensity of the product features. Firstly the selection component allows differentiating between relevant and non-relevant reviews those reviews are selected which are relevant for the product. The extraction component extracted the product feature there are two different kinds of product features: explicit and implicit product features. The explicit features mentioned in a text are apparent and the implicit features must be inferred from the context, the author deals with explicit feature at the present work and chosen the supervised learning uses training examples for recognizing product feature he uses support vector machines for extracting product features and evaluations then storage in database. The aggregation component operates with the structured data of the database. It connects the product features with their valuations in the sentence and summarizes by aggregate and depict evaluations of product feature. The results of this process are saved in the database and depicted by spider charts. The analysis component detects dependencies among product features by using association rules and identifies major determinants of the customer's satisfaction with a product on the whole by applying decision trees.

3.2 Product Quality Improvement

Numerous process/performance improvement techniques and models, such as six sigma , value engineering (VE) and Deming cycle (PDCA) etc., were adopted over the years. These model of application did not have a perspective or organized framework that would guarantee that the problem solutions developed. Currently, we shall discuss those applications as follows:

3.2.1 Six sigma DMAIC methodology

In this section, previous studies will be conducted with regard to available publications of six sigma methodology that will help in research problem. Six Sigma was first adopted by Motorola in 1987. In 1995, Jack Welch, has successfully established the General Electric (GE) and published Six Sigma. He implemented Six Sigma in many processes and documented significant gains in process and financial results. The simplest definition for Six Sigma is to eliminate waste and to mistake proof the processes that create value for customer. The elimination of waste led to yield improvement and production quality; higher yield increased customer satisfaction[78][79].

Coronado et al. in [78] depict six sigma is a strategic business improvement mechanism used to optimize profitability, remove waste, reduce cost of quality and enhance the effectiveness and efficiency of all operations to meet customer requirements and expectations. They point out that a defect can be classified as an imperfection that causes a shortfall or failure of a process that triggers customer complaints. The researchers also found that currently companies across the world ranging from small businesses, private and public to large organizations have adopted this philosophy to substantially improve quality level, customer satisfaction, market share, employees moral, organizational culture, people development , return on investment, and much more. In addition, the paper describes Six Sigma as an inexorable and rigorous quest of the elimination of non-value added activities and variations in core business processes to achieve continuous and breakthrough improvement in organizational performance that impact on the bottom line result.

Andel in [80] claims that Six Sigma programs should be implemented with a clear objective of improving competitive positioning and of increasing the company's value as perceived by the customer. All activities related to Six Sigma implementation should be approached from that perspective. They also reported that Six Sigma is about reducing process variability and the defects those processes create. Six Sigma follows a disciplined, data-driven methodology to reduce defects whereas below 3.4 defects per million opportunities. In addition, Six Sigma program is bound to meet requirements , self-protecting opposition of customers. This motivates organizations to keep for the existing customers as a result of process improvements.

Chou in [40] reported that Six Sigma has been widely adopted in a variety of industries in the world and it has become one of the most important subjects of debate in quality management. Six Sigma is a well-structured methodology that can help a company achieve expected goal through continuous project improvement. Some challenges, however, have emerged with the execution of the Six Sigma. For examples, feasible projects generated, critical Six Sigma projects selected given the finite resources of the organization. Their study aims to develop approach to create critical Six Sigma projects and identify the priority of these projects. Firstly, the projects are created from two aspects, namely, organization's business strategic policies and voice of customer. Secondly, an Analytic Hierarchy Process (AHP) model is implemented to evaluate the benefits of each project and from which the priority of Six Sigma projects can be determined. Finally, based on the project benefits and risk, projects can be defined as Green Belt, Black Belt, or others types of projects. Also, their study aimed to scale down waste and mistake proof processes that create value and yield to the improvement of products and services quality and tremendous customer satisfaction.

Naumann et al. in [81] have indicated that the concept of six sigma is the development of a uniform way to measure and monitor performance and set extremely high expectations and improvement goals. The authors present a high-level review of basic Six Sigma tools for gathering customer requirements, conducting customer satisfaction surveys, and managing organizational processes and opportunities based upon customer input.

Treichler et al. in [82] have concluded that six sigma is a highly disciplined process that helps an organization to focus on developing and delivering near-perfect products and services. The six sigma methodology of measuring and monitoring performance issue deals with a variety of statistical applications. The objective of six sigma is to enhance the sigma level of performance measures that reflects the needs of the customer. In addition, the Six Sigma level of performance means a product defect rate of 3.4 per million opportunities for error. In order to achieve company goals, the critical-to quality (CTQ) representatives of the product or service are identified. As the average CTQ capability increases, the capability of the corresponding process increases, make it further achieve strategic business goals. Current applications of the six sigma methodology emphasize the phases that are integrated in conducting a project, which include define-measure-analyze- improve-control (also known as the DMAIC cycle). The DMAIC cycle comes into play to meet the customer needs consistently and perfectly [83]. More related applications about Six Sigma methodology could be found in many recent publication [84][85].

The aforementioned researchers have reported about the six sigma applications. However, the following researchers described other methodology of product improvement approaches.

3.2.2 Value Engineering (VE) approach

Miles in [86] proposed Value Engineering (VE) approach. Value Engineering has been evolving for the last 60 years as a way to remove unnecessary cost from the product design before, during, and after the fact. This approach first identifies the intent or function and understands the context, then develops alternatives and implements a plan. He outlines a structured process that consists of defined steps called a “job plan,” which includes the identification of what the market furnishes and needs, as opposed to the producer’s perception of what the customer wants, and defines the priority of requirements. The underlying foundation of Value Engineering is to challenge the assumptions most people make about how the product or service satisfies the needs of the customer. Then rebuilding the product or service by identifying that the customer needs something done, they want an outcome. Customers do not want a feature, they want a function. After all, it is the function that creates a benefit for the customer. Value engineering is often done by systematically following a multi-stage job plan. Miles' original system was a six-step procedure which he called the "value analysis job plan." Others have varied the job plan to fit their constraints. Depending on the application, there may be four, five, six, or more stages. One modern version has the following eight steps:[87]

1. Preparation   2. Information   3. Analysis   4. Creation
5.Evaluation   6. Development   7. Presentation   8. Follow-up

3.2.3 Deming PDCA cycle

Deming in [88] business processes should be analyzed and measured to identify sources of variations that cause products to deviate from customer requirements. He recommended that business processes be placed in a continuous feedback loop so that managers can identify and change the parts of the process that need improvements. It's commonly known as the PDCA cycle for Plan, Do, Check, Act: [18][46][89].

- PLAN: Design or revise business process components to improve results
- DO: Implement the plan and measure its performance
- CHECK: Assess the measurements and report the results to decision makers
- ACT: Decide on changes needed to improve the process

Deming's focus was on industrial production processes, and the level of improvements he sought were on the level of production. In the modern post-industrial company, these kinds of improvements are still needed but the real performance drivers often occur on the level of business strategy. Strategic deployment is another process, but it has relatively longer-term variations because large companies cannot change as rapidly as small business units. Still, strategic initiatives can and should be placed in a feedback loop, complete with measurements and planning linked in a PDCA cycle [15][89][90].

The results of our survey reveals that there is no publications describe a model that used opinion mining model with six sigma DMAIC methodology in order to improve the product quality.

3.3 Summary

Clearly, this chapter reviewed development of opinion mining that used with many researchers. It also introduced the related work that helped to achieve our work. The previous work showed that there is no clear cut study of opinion mining using six sigma method in product quality improvement. Therefore, Chapter four will present this study of the research methodology and the tools that employed for the product quality improvement.

CHAPTER FOUR

Chapter Four
Research Methodology and Results

In this chapter, we will explain the proposed methodology which is based on Six sigma DMIAC methodology. Also, we will apply our methodology steps on Samsung Galaxy S2 and S3 as a case study. Different steps have to be performed. The main required steps are shown in figure (4.1) as follow:

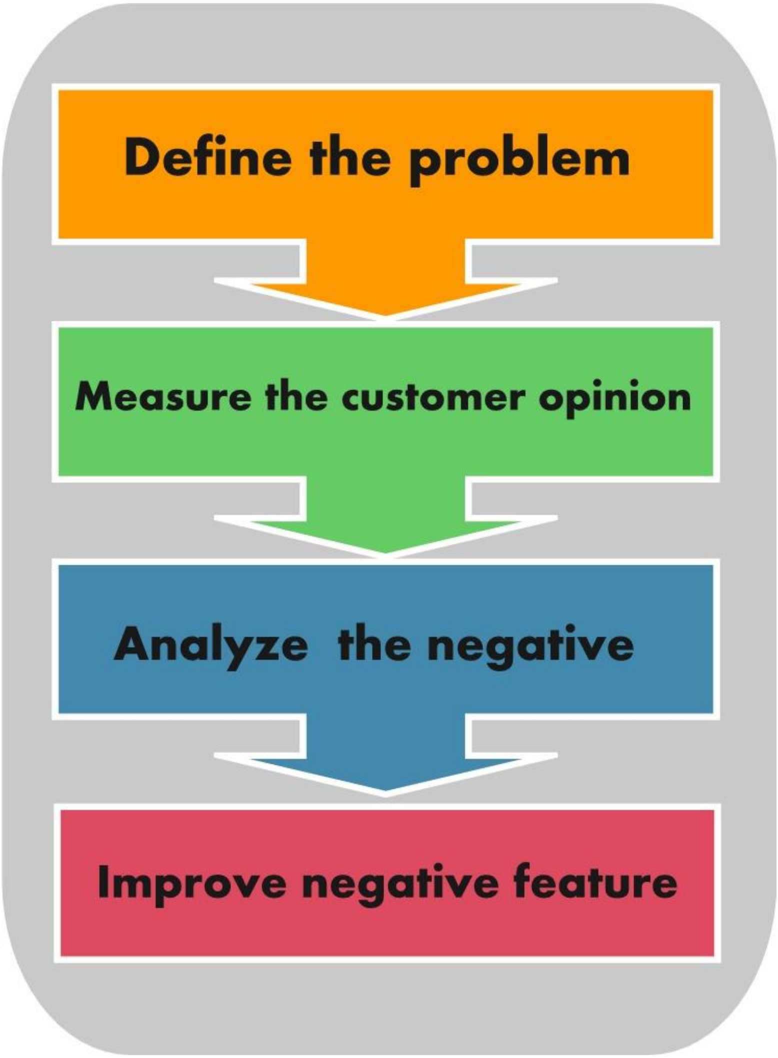


Figure (4.1) Methodology steps

This methodology focuses on improving the performance of an existing process. Improving the process can relate to a service, product, or manufacturing and transactional process. In this research, we suggest to use opinion mining as a tool to apply this methodology. As a case study of implementing this methodology, we used Samsung Galaxy S2 and compare the results with Galaxy S3 regard that Samsung Galaxy S3 is the next version and the improved product for Samsung Galaxy S2. The problem is to solve by using opinion mining to improve product quality. The research objectives can be achieved by adopted the techniques, which composed with the phases as follow:

- **Define the problem:** Define the problem through the case study.
- **Measure the opinions:** Using measuring tool such as Rapid Miner to distinguish positive opinions from negative opinions.
- **Analyze the negative:** Analyzing the data, determine the negative feature of the product, using Stanford parser tool as a feature extraction software.
- **Improve the feature:** Improve the negative feature of the product and matching the negative feature with the positive one in the next version.

The following sub sections in this chapter present the research methodology for the case study.

4.1 Define the problem

In this phase we need to set project goals and boundaries based upon the knowledge of the organization's business goals, customer needs, and the process that needs to be improved [32]. The phase consists of the following steps: Definition problem statement, Research goals, Customer and Voice of the customer (VOC) [40]. The following is how we can apply these steps in our case study.

4.1.1 Definition problem statement

Mobile phones companies seek to meet the needs of the target audience improve their product quality. So they conduct studies and devise the latest services to make its product easier for their customers and to keep them and attract new ones. The world has become totally dependent on the technology in everything of the life. Customers services constitute an important part of the services mobile and condition of widening competition in the market rival up to crowding.

4.1.2 Research goals

Research goals is to improve the product quality by analyzing the complaints of the customers and to extract the negative features of the product. The results of these features will be improved. Thus, the defects of product and customers compliant could be reduced. Accordingly, this would increase the product sale and the profits of goods. In our case, the research goal is presented to find a way to extract features for the current product in the negative opinions. In addition, to verify the product quality is in improvement phase and to increase the loyalties of customers to the company by using six sigma methodology.

4.1.3 Definition of Customers

The customer of our product is an individual or business that purchases the goods or services produced by an institution. The customer is the end goal of business, where the customer who pays for supply and creates demand. Businesses will often compete through advertisements or sales in order to attract a larger customer base [34].

4.1.4 Voice Of The Customer (VOC)

After the voice of customer VOC could be collected from the e-commerce website (e.g. www. Amazon.com). Since, VOC used to describe the customers' needs and their perceptions of the product or service [32]. Representative samples of the VOC were involved the customer opinions from the Amazon website. This would present samples of individual customer responses that represent the data gathered. The answers are to the following question:

“What is your opinion about Samsung Galaxy S2?”

Opinion 1 " This mobile was already used and I guess they change the covers, because it was with glue around the borders...also the telephone was BLOCK! So this means that they do not sell factory unblock mobiles...they unblock them at the store "....

Opinion 2 "There is no way the battery provided with this phone was the factory original - it lasted 4-5 hours at the most. I will be forced to buy an actual new battery if this phone is to work. UPDATE: Now the phone does not turn on at all after 6 days of ownership. Great. Additionally, the phone comes with Chinese bloat ware that I cannot uninstall. I have no idea what they do. Find another seller ".

Opinion 3 " If you compare Samsung`s phones with IPhone or HTC, Samsung`s phones are just trash. I owned an IPhone and HTC before but I have a Samsung galaxy s II right now. It gets freeze, menus are popping themselves up. Responding very slowly. Do not waste your money and time, just get an IPhone"

Opinion 4 " The phone is good. The only I wanted to tell is there was no Chinese language. If there is Chinese language would be better"

4.2 Measure the customer opinion

After defining the customers and based on the voice of the customer VOC in the previous section and according to an effective product sample of data, we will achieve opinion classification on the product in this phase.

The main activities in this phase are identify what to measure and make data collection plan. Data Collection Plan (DCP) includes the measure whose data needs to be collected, how much data to collect, data source, how will measure the data (step to accomplish) [17][39]. This activity gives knowledge that will assist us in narrowing the rang of potential causes that are needed to investigate in the analysis phase. In this phase, we conducted classification algorithms to measure the performance of Samsung Galaxy S2 customers opinion. Table (4.1) illustrates the DCP which describe the steps of the measurement phase.

Table (4.1) Data Collection Plan (DCP) in measurement phase

Tasks	What to Measure	Sources	Step to Accomplish	Schedule
Customers reviews about Samsung Galaxy S2 and Samsung Galaxy S3	Measure the customers opinion by classify into positive and negative classes	www.amazon.com	-Data Acquisition -Data Description -Data preprocessing -Using classification to measure the accuracy.	From 26/5/2013 To 2/6/2013

The most important activities in this phase are the identification and validation at the data accuracy. We apply measure phase as follows:

4.2.1 Data Acquisition

Data set reviewers collected from online economic website, including ***www.Amazon.Com***. We collected data between 26/5/2013 and 2/6/2013. Table (4.2) Presented general information of the collected data.

Table (4.2) General information about data sets

Data sets	Instance	Attribute #
Galaxy S_2	281	3
Galaxy S_3	601	3

4.2.2 Data Description

In this section, the dataset selected for this research is a tangible product not services where this product has another version considered as an improvement version product. The data of Samsung Galaxy is a series of Android-powered mobile computing devices designed, manufactured and marketed by Samsung Electronics. The product line includes the Galaxy S series of high-end smartphones such as Samsung Galaxy s2 and Samsung Galaxy s3. The Galaxy S2 have 4.3 inch display, 8 mega pixels camera, 1.2 GHz dual core processor with OS, v2.3.4 (Gingerbread), v4.0.4 (Ice Cream Sandwich), upgradable to v4.1 (Jelly Bean). Whereas, Samsung Galaxy S3 have 4.8 inch display, 8 mega pixels camera, 1.4 GHz Quad-core processor with Android OS, v4.0.4 (Ice Cream Sandwich), 4.3 (Jelly Bean). Table (4.3) presents the Attribute and their Description of our data sets [91][92].

- Samsung Galaxy S2 data set over 281 reviews and opinions from customers it was taken from ***www.Amazon.Com*** economic website, which consist of 3 attributes available in its data base which are Reviews, Rate, Polarity , see Table (4.3).
- Samsung Galaxy S3 data set over 601 reviews and opinions from customers it was taken from ***www.Amazon.Com*** economic website, which consist of 3 attributes available in its data base which are Reviews, Rate, Polarity , see Table (4.3).

Table (4.3) Dataset Description.

Attribute	Description
Reviews	The customer reviews about galaxy smart phone from Amazon web site
Rate	The rate of the review (from 1 to 5).
Polarity	Depend on the Rating Attribute positive or negative.

4.2.3 Preprocessing

Preprocessing is a necessary step for serious, effective, real word data mining. Real word data are generally incomplete, noisy or inconsistent because that we will apply number of preprocessing techniques to get high quality mining result. We apply Tokenization, Transform cases, Filter stop words (English), Generate n-grams (Terms), Filter Tokens (By length), Stem (porter). Figure (4.2) shows the preprocessing techniques that we applied in our datasets using Rapid Miner tool.

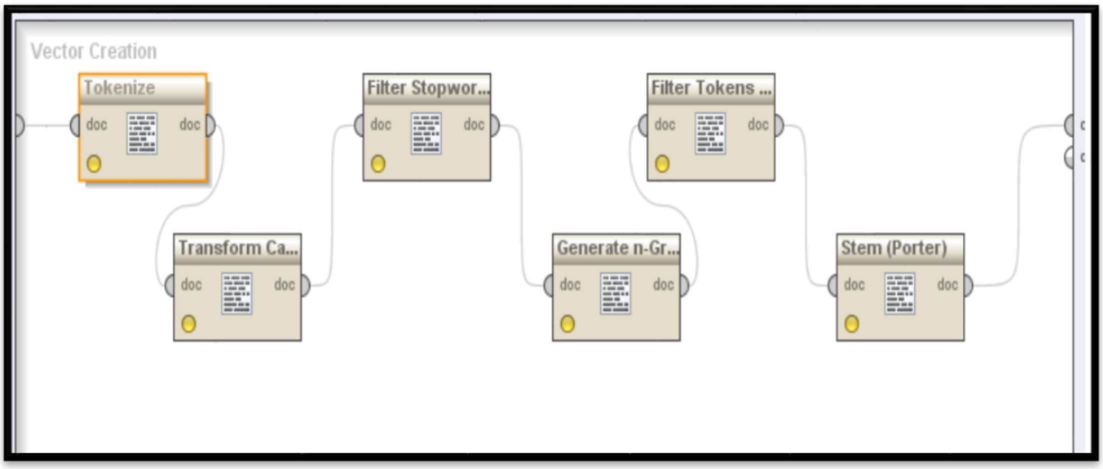


Figure (4.2) Illustrate the preprocessing process on the data sets

In this section, we apply a number of preprocessing techniques on the Samsung Galaxy S2 data set.

In the preprocessing step, documents are transformed into a representation suitable for applying the learning algorithms. (As we discussed preprocessing techniques in section (2.3.4) we applied the following):

- **Tokenize:** Usually the first step in imposing some structure to your document data is to break up the document into tokens. The simplest token is a character; however, the simplest meaningful (to a human) token is a word which we used in our data.

- **Transform cases:** The fundamental process in most text mining requires converting text from the uppercase to a lowercase.

- **Filter Stop words:** This step is the remove what are called "stop words". These are common words such as prepositions, conjunctions, articles, adverbs and so on.

- **N-gramming:** In documents, there may be pairs of words, which always go together. This process would help with the understanding of the data mining and give a more insight into the data to be in a better way.

- **Stemming:** An important pre-processing step to input documents begins the stemming of words. The term "stemming" refers to the reduction of words to their roots.

4.2.4 Data mining classification methods

In this section we describe major kinds of classification algorithms which are used in our research such as Naïve Bayes (NB), K –Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT) which are provided in Rapid Miner environment. The following subsections presented these classification algorithms and their setting which are used during experiment results.

4.2.4.1 Naïve Bayes (NB)

NB is used in our research, where it is considered as one of the most widely classifiers. Accordingly, the data is classified into classes that are presented in the training data set and estimating probabilities of individual variables values. This allows using these probabilities to classify new entities [56]. NB classifier method that used in the present work is shown in figure (4.3), where both training and testing process is achieved.

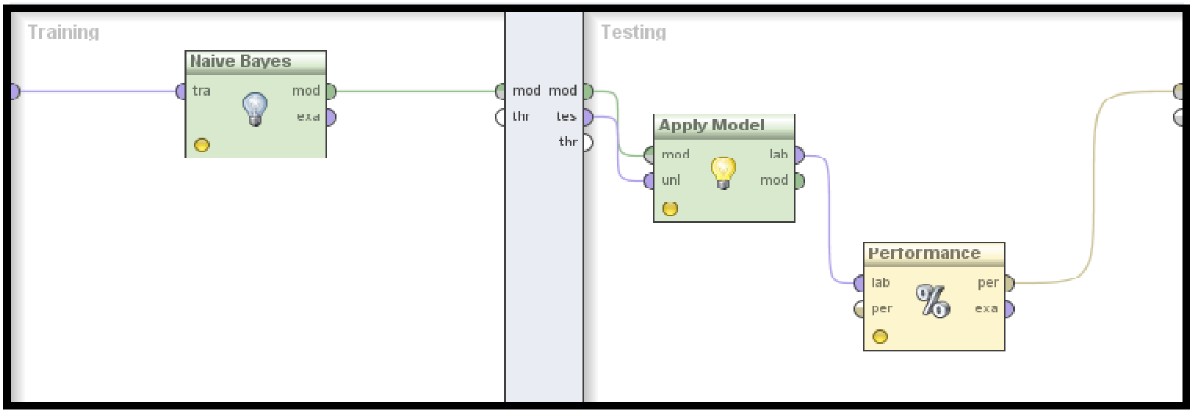


Figure (4.3) Illustrates the setting of Naïve Bayes

4.2.4.2 Decision Tree (DT)

DT is a learning model designed to predict the value of a target variable based on a number of input variables. There are several notable wages of decision tree, such as random forest classifiers, boosted trees algorithm [53]. DT classifier is presented in figure (4.4).

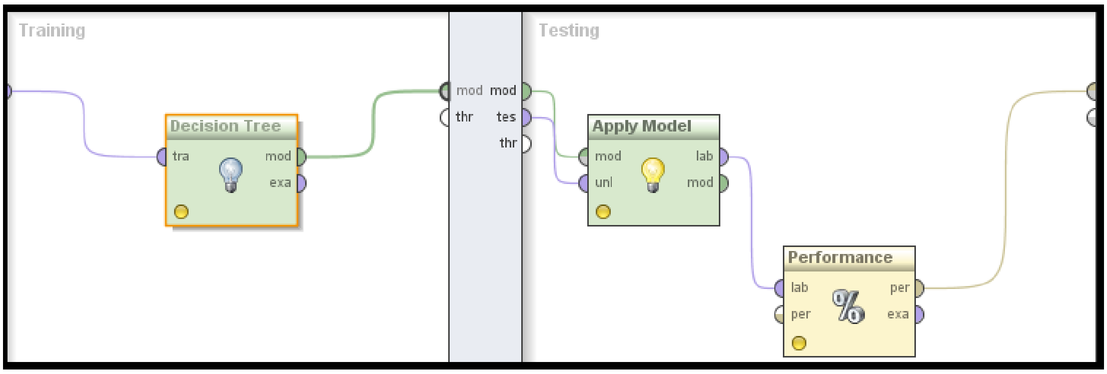
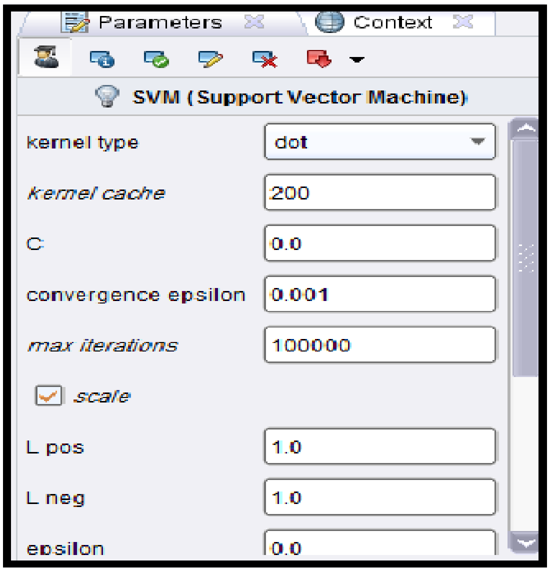


Figure (4.4) Illustrate the setting of Decision Tree

4.2.4.3 Support Vector Machine (SVM)

SVM is used in our research which a type of supervised learning method and produced a set of data vectors. SVM training algorithm is mapped into two categories that separated with largest possible margin boundary between them so that the examples of the separate categories are divided into two classes. The given data of training builds up a model that predicts which a new example belongs to a certain category [57]. Figure (4.5), (4.6) represent the setting of SVM.



SVM (Support Vector Machine)

kernel type: dot

kernel cache: 200

C: 0.0

convergence epsilon: 0.001

max iterations: 100000

☑ *scale*

L pos: 1.0

L neg: 1.0

epsilon: 0.0

Figure (4.5) Shows the setting up of SVM

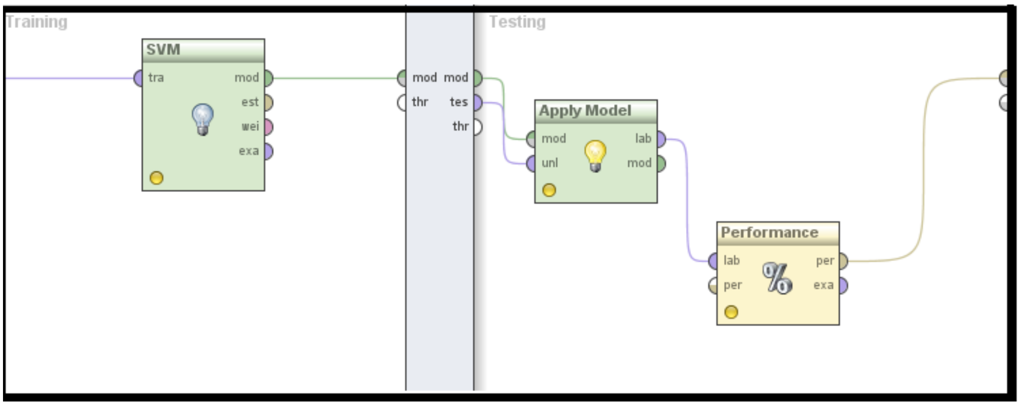


Figure (4.6) Illustrates the setting of SVM

4.2.4.4 K –Nearest Neighbor (KNN)

KNN classifier is also used in our research that describes a non-parametric method and supervised learning algorithm where a majority of KNN category are given and detailed. The purpose of this algorithm is to classify a new object based on attributes and training sample [53]. We change K value in parameter setting and try a series increasing (K=1,3,5,7,9) and take the highest accuracy. This KNN classifier is presented in figure (4.7). A series of K values that changed are also listed in figure (4.8).The results in table (4.4) indicates that the worst accuracy is for K=1, however the best accuracy is found for k=7.

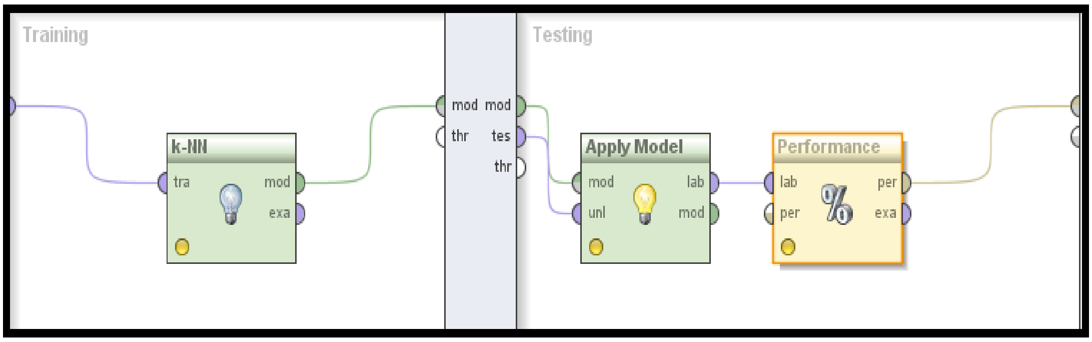
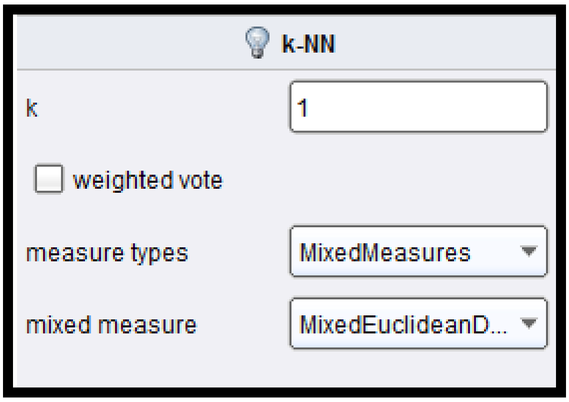


Figure (4.7) Illustrates the setting of KNN


k-NN

k 1

□ weighted vote

measure types MixedMeasures

mixed measure MixedEuclideanD...

Figure (4.8) Illustrates the setting of KNN and the input K value

- **Table (4.4) detailed and listed the accuracy result for each classifier that used in this section.**

Table (4.4) Accuracy for Samsung Galaxy S2 dataset in Data Mining Classification Experiments.

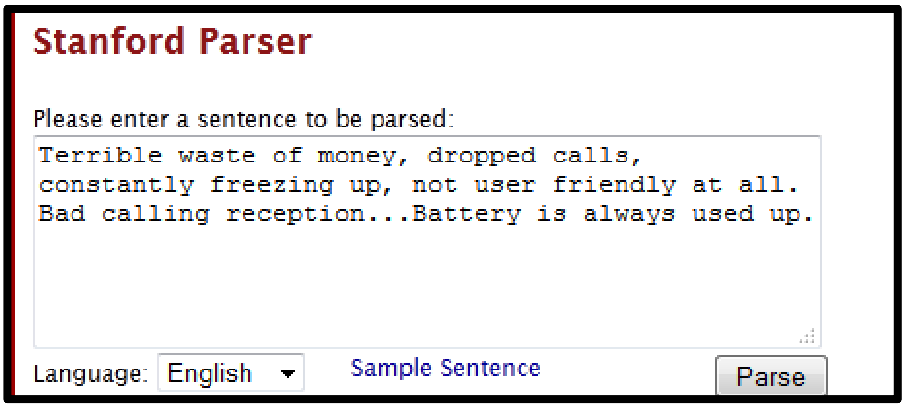
Classifier	Accuracy	
Naïve Bayes(NB)	%84.52	
Support Vector Machine (SVM)	%82.50	
K –Nearest Neighbor (KNN)	K	
	1	%78.93
	3	%80.36
	5	%82.14
	7	%83.21
	9	%80.36
Decision tree (DT)	%83.33	

Major kinds of classification algorithms are used to classify the customer's opinions. Table (4.4) shows that the best accuracy is for Naïve Bayes (NB) = 84.52% and the worst one at K –Nearest Neighbor (KNN) = 78.93% when K value = 1.

4.3 Analyze the negative

Analysis phase is defined as the process to determine root causes of variation or gaps between actual and goal performance [32][40]. To understand this gap we extract negative features from customers opinions. Customer complains come from dissatisfaction of the product, where this could be analyzed in this phase. Based upon the previous section on the measure phase, where classification of the customer's opinion is carried out, this classification gives rise both of a negative and positive opinion. In the present work, the product quality improvement will be only applied for the negative opinion, so that the positive one is excluded. Based on previous phase, we used the negative data which classified by the best classifier which is Naïve Bayes.

Product feature are usually nouns or nouns phrases in review sentence [68]. We used Stanford parser [70] to parse each review to split text into sentences and to produce the part-of-speech tag for each word (whether the word is noun, verb, adjective, etc). The following steps shows an example. Figure (4.9) illustrate the negative opinion inserted in a Stanford tool as an input.



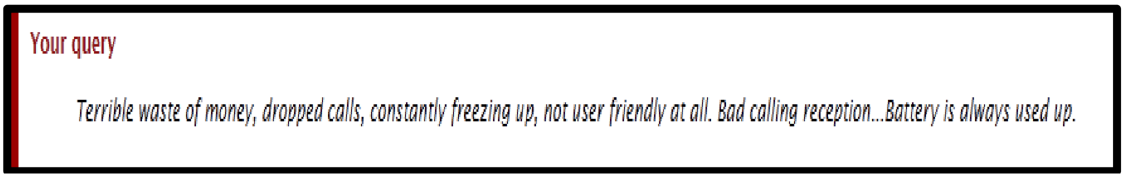
Stanford Parser

Please enter a sentence to be parsed:

```
Terrible waste of money, dropped calls,
constantly freezing up, not user friendly at all.
Bad calling reception...Battery is always used up.
```

Figure (4.9) Customer opinion inserted in the Stanford parser as input

The opinion text as a query is also shown by Stanford parser as described in figure (4.10 a) and this text could be more obvious in figure (4.10 b).



Your query

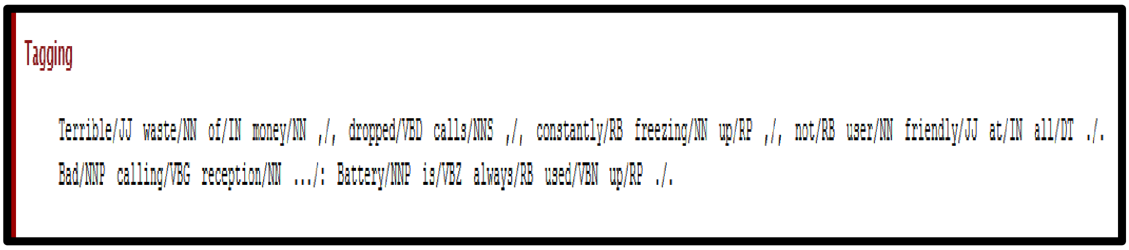
Terrible waste of money, dropped calls, constantly freezing up, not user friendly at all. Bad calling reception...Battery is always used up.

Figure (4.10 a) The text of the customer opinion as a query in Stanford parser

Terrible waste of money, dropped calls, constantly freezing up, not user friendly at all.
.Bad calling reception...Battery is always used up

Figure (4.10 b) Shows a clear text of the customer opinion as a query

The Stanford parser make tagging to the text opinion and give an output analysis format as a part of speech as described in figure (4.11 a), where this is clearly listed in figure (4.11 b)



Tagging

```
Terrible/JJ waste/NN of/IN money/NN ,/, dropped/VBD calls/NNS ,/, constantly/RB freezing/NN up/RP ,/, not/RB user/NN friendly/JJ at/IN all/DT ./.
Bad/NNP calling/VBG reception/NN .../: Battery/NNP is/VBZ always/RB used/VBN up/RP ./.
```

Figure (4.11 a) Output analysis format as a part-of-speech tagged text

Tagging

```
Terrible/JJ waste/NN of/IN money/NN
,/,
Dropped/VBD calls/NNS
,/,
Constantly/RB freezing/NN up/RP
,/,
Not/RB user/NN friendly/JJ at/IN all/DT
./.

Bad/NNP calling/VBG reception/NN .../:
Battery/NNP is/VBZ always/RB used/VBN up/RP
./.
```

Figure (4.11 b) Shows the clearly output analysis format as a part-of-speech tagged text

The aforementioned process is applied for all negative opinion as discussed in the previous example. Thus, feature extraction is obtained that required to improve the Samsung Galaxy S2 according to customer complains.

- **Product feature extraction**

As discussed earlier in, it was found that noun and noun phrases were extracted from the POS tagging. Table (2.1) in chapter two illustrates the noun and noun phrases that represent the product feature extraction that we used in this phase to extract the frequent feature of the product. Table (4.5) shows an examples of the obtained feature extracted as concluded from the analysis.

Table (4.5) Illustrated feature extraction resulted in the analyze phase.

Feature Extraction				
Battery	OS	Cover (body)	Touch Key Board	4G internet
SW(Android)	IMEI	Handset	Internal memory (large space)	Accessories
Box	Camera	Customer career	Typing on the Key Board	Cheap Quality
Signal	Display screen	Sound	Warranty	Mobile processor
Hot if it charge	Restart for the cell phone (Battery devote)	Freezing up (Froze) Continuously	Not friendly	Connect with the computer SW
Bad calling reception	Samsung galaxy easy to broke Expensive to replace	Version (Model)	Search for connects	Vibrator motor
Menus (Icons)	Responding very slow	Map app	Amazon career for sailing Samsung phones	

Feature extraction is also classified into both frequent and infrequent, frequent feature appears in majority of the opinions, infrequent feature is only appeared in a few number of customer opinion [5]. Table (6.4) illustrated the most Frequent Feature Extracted from customer's opinion.

Table (4.6) Illustrated frequent feature extraction.

Frequent Feature Extraction			
Battery	Freezing up	OS	Warranty
Processor	Internal memory	Customer career	

The results exhibits the most important steps of the analysis phase that extract the negative feature, where it is useful for the product quality improvement process. This will be explain and discuses in next section.

4.4 Improve negative feature

Improve phase can be defined, as the process performed by eliminating the defects is capable of producing the customer requirements. This is considered to be iterative until improvements are identified and verified with facts, data and performance metrics [32][33][40]. In fact, we proposed the improvement steps in figure (4.12) that are applied in the present work. These steps will be explained throughout this section .

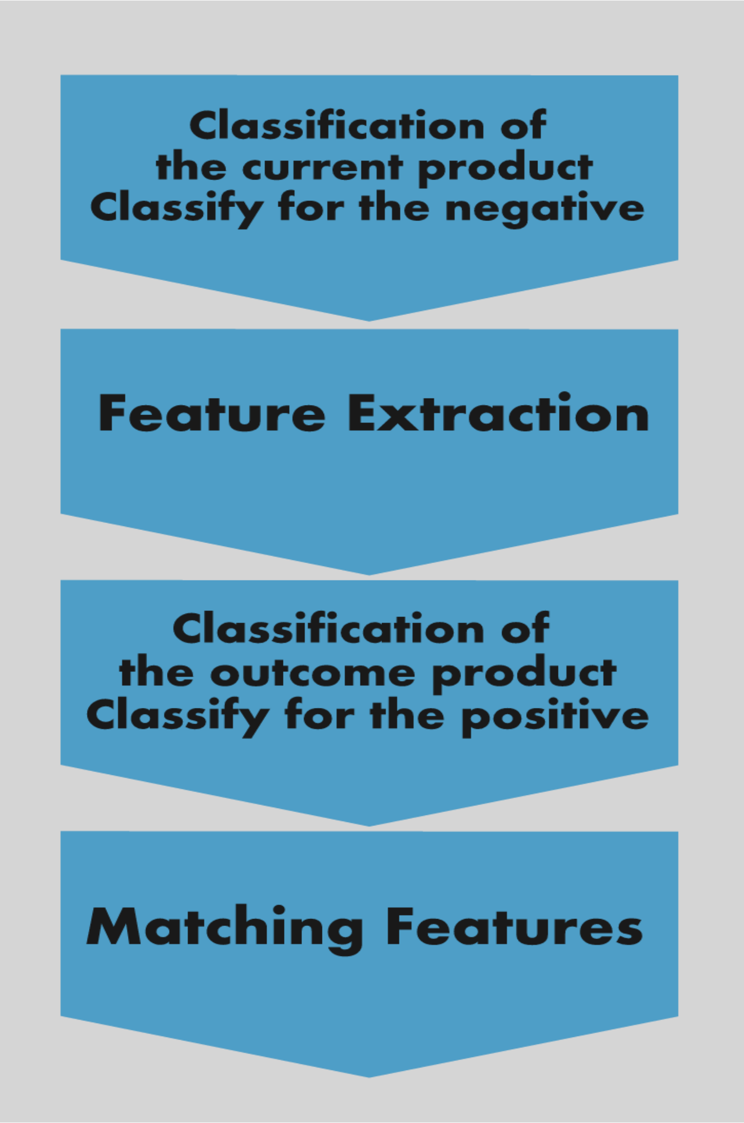


Figure (4.12) The improvement phase steps

The improvement steps are carried out as follow:

4.4.1 Classification of the current product

In first step, we get the negative opinion of the current product. For example in our case the negative of Samsung Galaxy S2. In our case, customer opinion is classified into positive and negative for this product. Positive opinions are discarded and we concentrate at negative opinions. This is carried out throughout different classification method by using Rapid Miner environment. This step was carried out as part of measure phase.

4.4.2 Feature Extraction

By analyzing the negative opinion, we have obtained the negative features (as defects) according to the customer complains as described in the analysis phase, where Stanford parser is used for this tasks.

4.4.3 Classification of the outcome product

In this step, we get the positive opinion of the outcome product. For example in our case the positive of Samsung Galaxy S3. In our case, customer opinion is classified into positive and negative for this product. Negative opinions are discarded , in fact negative opinion can be used to improve than next version of the product since improve phase is an iterative process. However, in this thesis we concentrate at positive opinions.

At the beginning, preprocessing techniques is used for this data set to get high quality mining results and effective real data. The preprocessing steps is presented in figure (4.13) (as this is applied in section (4.2.3) for Samsung Galaxy S2). Then, customer's opinions is also classified into positive and negative opinions and results a positive one for the outcome product, where Rapid Miner is used too. The classification process is carried out through different method as shown in Figures (4.14) and (4.15).

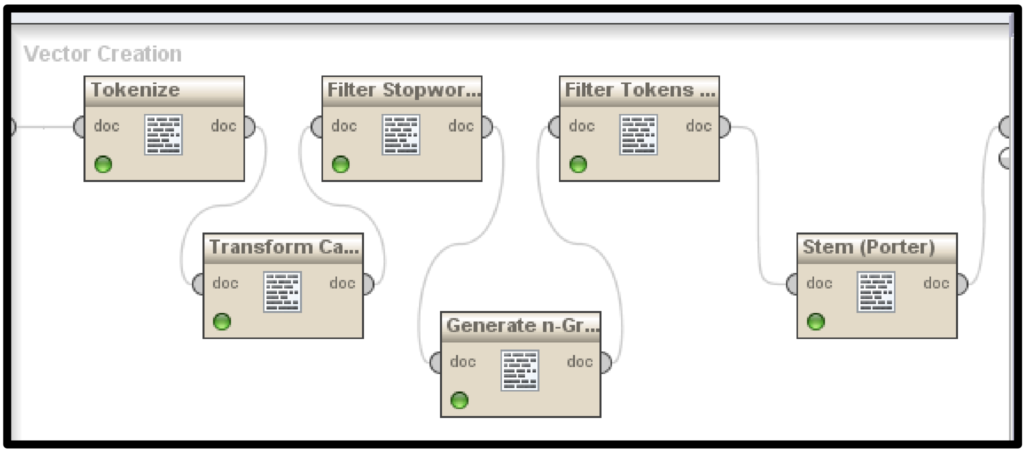


Figure (4.13) Describes preprocessing techniques used for the outcome product (Samsung Galaxy S3)

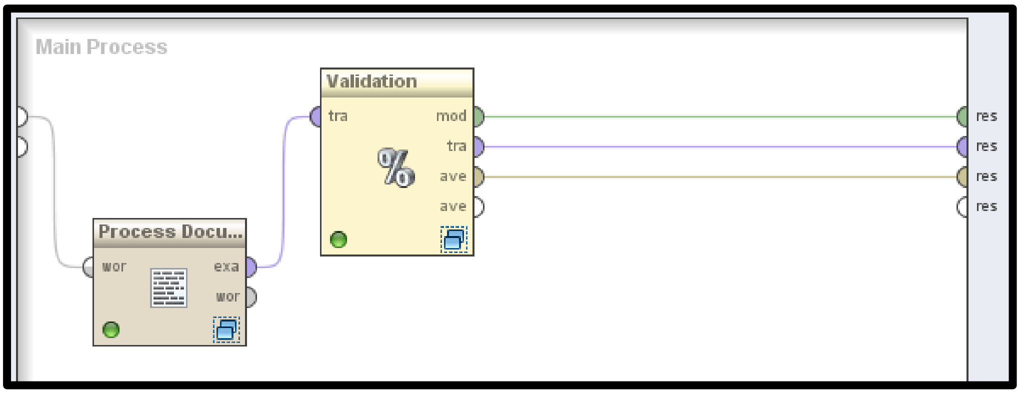


Figure (4.14) Exhibits the classification step of the output product (Samsung Galaxy S3)

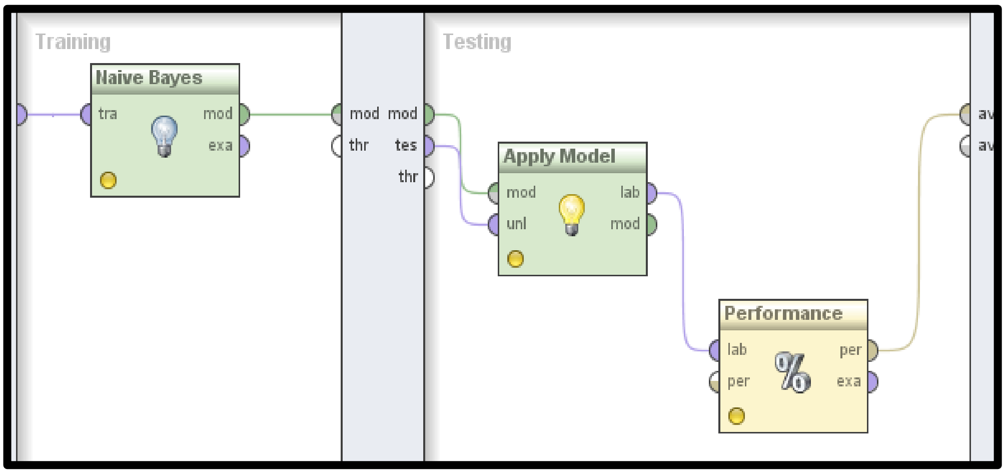


Figure (4.15) Exhibits the classification for (Samsung Galaxy S3) using Naïve Bayes

4.4.4 Matching features

Based on the previous classification process for both current product (e.g. Samsung Galaxy S2) and outcome product (e.g. Samsung Galaxy S3), a matching features is obtained. This matching is produced from the negative and positive opinions results of both products as well as to the features. Accordingly, the improvement of the outcome product is presented by this matching features. Examples of negative and positive opinions and features are listed in table (4.7).

Table (4.7) Describes sample matching improvements of the outcome products that presented by negative, positive opinions and features.

S2 Negative	S3 Positive	Feature
Misleading advertising. There is no international warranty. I like my Iphone better because the screen is sharper. The touchscreen also seems slower. I've only had it a week and will continue testing.	The phone is awsome i love it...they delivered as per my expectations but it will be great if i received a cover and screen protector..i have to buy it...	screen
There is no way the battery provided with this phone was the factory original - it lasted 4-5 hours at the most. I will be forced to buy an actual new battery if this phone is to work.UPDATE: Now the phone does not turn on at all after 6 days of ownership. Great.Additionally, the phone comes with Chinese bloatware that I cannot uninstall. I have no idea what they do.Find another seller	excellent product nice screen now i know what they call s,art phones battery last long and speakers are loud enough great value	battery
The order I placed was for the Samsung Galaxy S II GT-I9100 (International version). The one I received today has the correct box but the cell phone deliberatly placed inside is the GT-I9100G which is not the international version. For those who doesnt know, this model has a different processor.I am returning the phone right now!!!!I hope the seller will comment on this entry.	I was looking for a good cell phone and I ended with a perfect one!No complaints at all.Perfect size, processor, memory and performance.	processor
I don't know why but the camera and battery didn't work. It fell down from a chair and the display screen broke. Don't buy from this seller EVER!	Best Smartphone,1- Large crisp clear bright screen.2- Great Camera with LED and front facing camera. Image Quality is superb.3- Sound and voice are very good.4- Wifi, GPS and Bluetooth work fine.5- Fast, No freezes so far. Smooth flow,6- Battery could last 1-2 days with mild use in battery saving mode, charges fast.7- Great Android OS and Google Market.8- Large Memory. Upgradable with card.9- Uses Mini Sim Card and not the standard one!.10- Responsive screen.11- Little bigger than I like but still very thin and light.	camera
I bought two phones: A Samsung Omnia 7 with Windows Phone and this Samsung Galaxy SII.I can't believe how cheap the quality of SII is. Also the OS is zero friendly compared to my Windows Phone 7.Good things are the camera and the screen size...I like that.	Super phone. Very slick looking, fast and user friendly. Nice colour. The only problem is the international version will arrive set in a different language...this video shows you how to change to English or another preferred language[...]Otherwise perfect!	friendly
Terrible waste of money, dropped calls, constantly freezing up, not user friendly at all. Bad calling reception...Battery is always used up.	I have been using Samsung Galaxy SIII for more than 6 months now. There were some teething troubles in moving from iOS to Andriod. But over all Galaxy SIII is the best phone from android stable.	freezing

I could not get android upgrades since I bought the device. Kies says that the device is not eligible for upgrade. Some offer to root the phone for upgrade but I don't want to root it. This problem could be mentioned on the page at amazon.com. If I knew device has such a problem I would not have bought it.Secondly, the internal (system) memory gets full in every 2-3 months and I have to reset the device to factory settings again and again. There is no solution for this and Hey! I had not been warned that I was about to buy a defective device.	Overall, The Galaxy S III is the best phone I have ever owned. It is perfect for any college student. It is easy to use and its internal memory is excellent. It takes amazing photos with the rear camera, the front camera is also very good, and the camera setting are very easy to use and offer a variety of effects. The update for the phone now has a driving mode setting which reads you who is calling or texting you. The personalization features are very good as well. The only drawbacks of this phone are the battery life and durability. I usually run out of battery by noon everyday. I recommend buying a backup battery and charging station if you use your phone heavily. This phone is very fragile too. Even though the screen is nearly scratch proof, a short drop on the pavement or floor will crack the screen. The otterbox defender case is a sure way to ensure the safety of the phone. They are a little pricy but make the phone waterproof and shock absorbent. Occasionally there is a lag in the phone. This phone is an excellent buy overall and can meet the needs of any person.	internal memory
There are two major issues with this handset which I received: I received I9100P1. I am not able to connect to AT&T GoPhone(prepaid) data plans. AT&T customer care says its not possible to use this handset.2. The battery backup is not as expected. Screen display is consuming lot of battery even if the settings is in low brightness. My earlier Samsung smartphone had better battery backup for the same usage.	Surprisingly fast shipment, phone worked flawlessly, all the accessories where there and worked as they should. I had to return it not because there was anything wrong with the phone, I wasn't aware that International versions of phones doesn't really work well with US carriers because the bandwidth they use is different as the rest of the world. I contacted them and explained why I had to return it and in 5-10 mins they approved me to return it. Even issuing the refund was pretty fast and without any issue. I wouldn't hesitate if I had to buy from this company again, really professional and responsible.	accessories
The phone OS can't be updated, is refurbished, not new, came in a box not original, is a scam or rip-off you named	Having owned a vast number of cellphones myself from different brands and using different OS I'm forced to recommend this one, not only because it absolutely met my expectations but because it shows how a great product, with a great design and a fantastic software implementation can be put out there for a fair price.	os

I RECEIVED THE PHONE ON TIME AND THATS THE ONLY GOOD THING. THE PHONE WOULD NOT UPDATE ON SAMSUNG KIES SOFTWARE, RADIO DOES NOT WORK, SWIPE DOES NOT WORK. I WILL ADMIT THAT THE PHONE WILL MAKE AND RECEIVE CALLS.BUT THATS IT. DONT KNOW WHY SELLER PLAYED WITH PHONE. THE PROBLEM IS THAT ITS NOT WORTH RETURNING THE PHONE FOR A REFUND AS I AM OVERSEAS LIVING IN THE CARIBBEAN. PHONE WAS BOUGHT AND SHIIPED TO A FREIGHT FORWARDER. WOULD NEVER BUY FROM THIS SELLER AGAIN.	awesome, its a gift for my wife....Hope she will like it too.....Thanks Hasslefree cell.........thanks for your quick customer feed back	customer care
Very True. Even I was told that we get a 1 year warrenty which was not true. AMAZON.COM Please stop promising people on Warranty Advertisement.	The product is as advertised. I like the way the seller put a small note about how to change the language of the phone from German to English :)The only problem is the Warranty which is really important to me and I expected them to fill out and sign the Warranty card, because without date of purchase and signature of the dealer, the Warranty is not valid (I think so). I couldn't find any way to contact the seller. That's why I finally mentioned the Warranty issue here.Everything else is perfect. The phone seems awesome (for one-day working) and the shipping was fast and on time. I'll update this review after working with the phone for a while.Thanks,Maysam	Warranty

4.5 Summary

The proposed methodology consist of many steps to achieve and evaluate the results. Define the problem is the first step where problem statement, research goals, define of our customer as well as the voice of the customer VOC are described. The second step is the measure the opinion, where in the customer opinions are classified into positive and negative classes using classification algorithm in Rapid Miner environment. An accuracy could be obtained in this phase. In analyze the negative, we extract the negative product feature from customer opinions. This would be improved and given the most frequent features as discussed throughout this chapter. Finally, improvement phase, we match the negative opinion from the current product (e.g. Samsung Galaxy S2) with the positive opinion of the output product (e.g. Samsung Galaxy S3). This phase gives rise a sample of results that detailed throughout the present work.

CHAPTER FIVE

Chapter Five
Conclusion and Future work

5.1 Discussion and conclusion

In the present work, we have studied the opinion mining model to understand and analyze negative reviews so as to improve the product quality through customer opinions and obtain customer satisfaction improvement.

This work has been carried out by using opinion mining in a proposed methodology which we based on six sigma DMAIC methodology, since our survey there is no clear evidence that this method is previously used to improve the product quality. The major value of the present work lies in the detection of this model, opinion mining on the product quality improvement. Opinions and reviews can be easily expressed on the web such as in merchant sites, review portals, blogs, Internet forums and much more rather than a traditional market survey that seeks out customer opinions with questioners or interviews. This data reveals that both product manufactures and potential customers are very interested in this customer feedback or customer opinions. It provides a knowledge of customers likes and dislikes as well as the positive and the negative comments on the products. This also gives a better knowledge of the product quality improvement, limitation and advantages over competitors. In addition, customer reviews provide potential customers with useful information on the products and services to aid purchase decision making process and increase their motivation. Furthermore, the loyalty of the existing customers would be largely increased.

Therefore, our methodology has been applied in this research since it is noticed that the customers complain and product defects are reduced by six sigma implementation. This is due to negative feature has significantly improved on the quality product . The proposed methodology consist of many steps to achieve and evaluate the results of the present work. Define the problem is the first step where problem statement, research goals, define of our customer as well as the voice of the customer VOC are collected to our case study (Samsung Galaxy s2 mobile phone). The second step is measure the customers opinion, where data collection plan (DCP) are carried out throughout this phase. Data Acquisition, Description, Preprocessing and using classification algorithm KNN, DT, NB, SVM are applied then obtained where the accuracy of this results is also deduced. Customer opinions are classified into positive and negative classes using classification algorithm in Rapid Miner environment. Analyze the negative, is the third step, the negative product feature from customer opinions were extracted by Part of Speech POS tagging using Stanford parser to analyze each customer opinion in our dataset. Nouns and nouns phrases are obtained that represent the product features, where the frequent features is deduced. This would be improved as discussed throughout this research. Finally, negative opinion from the current product (Samsung Galaxy S2) is matching with the positive opinion of the output product (Samsung Galaxy S3 as a next version). This phase gives rise a sample of opinion matching results that detailed throughout the present work.

In brief, we believe that the opinion mining model is well implemented to improve the product quality. Clearly, our case study (Samsung Galaxy S2) illustrates the importance of using opinion mining so that product quality improvement can be easily obtained. Thus, this model can be customize to several organizations, products, process, services and management in order to achieve high quality improvement.

5.2 Future work

The future work could be proposed as follows:

- The present work is limited for English language opinion dataset and this would suggest for other language opinion dataset such as Arabic .
- The theoretical approach that used in the present work could be generalized for other products and give more quality results.
- The product could be selected either tangible product or services for customers where it leads to improve services and product quality.
- Clearly, different classification method can be applied and mostly are suitable to give an adequate results. The method based on a larger dataset could be more efficient and effective.
- Generalizing our approach to other kinds of user-generated content (UGC), e.g., Internet forums, discussion groups, social media , E-commerce websites and blogs, so that the customers opinions and reviews can be easily obtained.
- Our approach extended and applied to other domains such as news, sports, politics ……etc.
- Product feature from negative opinion is extracted to be improved of the product quality in the present work. Hopefully, the extracted feature of the positive opinion would give more efficient and effective results of product or services quality to meet the customers satisfaction.

References

[1] Linderman, K., Schroeder, R. G., Zaheer, S., & Choo, A. S. (2003). Six Sigma: a goal-theoretic perspective. Journal of Operations management, 21(2), 193-203.

[2] Pande, P. S., Holpp, L., & Pande, P. (2002). What is six sigma? (Vol. 1). New York: McGraw-Hill.

[3] Chen, Y. Y., & Lee, K. V. (2011). User-Centered Sentiment Analysis on Customer Product Review. World Applied Sciences Journal, 12, 32-38.

[4] Somprasertsri, G., & Lalitrojwong, P. (2008, September). A maximum entropy model for product feature extraction in online customer reviews. In Cybernetics and Intelligent Systems, 2008 IEEE Conference on (pp. 575-580). IEEE.

[5] Lo, Y. W., & Potdar, V. (2009, June). A review of opinion mining and sentiment classification framework in social networks. In Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on (pp. 396-401). IEEE.

[6] Harb, A., Plantié, M., Dray, G., Roche, M., Trousset, F., & Poncelet, P. (2008, October). Web Opinion Mining: How to extract opinions from blogs?. In Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology (pp. 211-217). ACM.

[7] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[8] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

[9] Juran, J. M. (1962). Quality control handbook. In Quality control handbook. McGraw-Hill.

[10] Mandelbaum, J. (n.d.). Product Improvement -- Which Approach? Retrieved June 9, 2010, from http://leaninsider.productivitypress.com/2012/05/product-improvement-which-approach.html

[11] Mitchell, A. (1994, Oct). The Quality Pocketbook. Management Pocketbooks.

[12] Marash, S. (2000). Six sigma: business results through innovation. In ANNUAL QUALITY CONGRESS PROCEEDINGS-AMERICAN SOCIETY FOR QUALITY CONTROL (pp. 627-630). ASQ; 1999.

[13] Kwak, Y. H., & Anbari, F. T. (2006). Benefits, obstacles, and future of six sigma approach. Technovation, 26(5), 708-715.

[14] Tonini, A. C., de Mesquita Spinola, M., & Laurindo, F. B. (2006, July). Six Sigma and Software Development Process: DMAIC Improvements. In Technology

Management for the Global Future, 2006. PICMET 2006 (Vol. 6, pp. 2815-2823). IEEE.

[15] Sokovic, M., & Pavletic, D. (2007). Quality Improvement- PDCA Cycle vs. DMAIC and DFSS. Strojniski Vestnik, 53(6), 369-378.

[16] Garvin, D. A. (1983). Quality on the line. Harvard business review, 61(5), 65-75.

[17] Masaaki, I. (1986). Kaizen: The key to Japan's competitive success. New York, itd: McGraw-Hill.

[18] Namasivayam, S. Al-Atabi, M.., Chong, C. H., Choong, F. , Hosseini, M. Gamboa, R. A., & Sivanesan, S. (2013). A blueprint for executing continual quality improvement in an engineering undergraduate programme. Journal of Engineering Science and Technology, Special Issue on Engineering Education, (pp. 31-37).

[19] Decker, S. (n.d.). Why Negative Reviews are a 'Gift'. Retrieved May 18, 2010, from http://blog.bazaarvoice.com/2010/05/18/why-negative-reviews-are-a gift/#sthash.ZqQnbyME.dpuf

[20] Chaovalit, P., & Zhou, L. (2005, January). Movie review mining: A comparison between supervised and unsupervised classification approaches. In System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on (pp. 112c-112c). IEEE.

[21] Bad reviews are good for business The power of negative reviews, Simplifying social commerce. (n.d.). Retrieved October 5, 2010, from http://www.reevoo.com/resourcess/ebooks/bad-reviews-are-good-business

[22] Nave, D. (n.d.). Bad reviews are good for business The power of negative reviews, Simplifying social commerce. Retrieved November 16, 2010, from dave@davenave.com

[23] Sokovic, M., Pavletic, D., & Pipan, K. K. (2010). Quality improvement methodologies–PDCA cycle, RADAR matrix, DMAIC and DFSS. Journal of Achievements in Materials and Manufacturing Engineering, 43(1), 476-483.

[24] Goffnett, S. P. (2004). Understanding Six Sigma: implications for industry and education. Journal of Industrial Technology, 20(4), 2-10

[25] Antony, J. (2002). Design for Six Sigma: a breakthrough business improvement strategy for achieving competitive advantage. Work Study, 51(1), 6-8.

[26] Forrest, W., & Breyfogle, J. (2003). Implementing Six Sigma: Smarter solutions using statistical methods.

[27] Snee, R. D. (2004). Six–Sigma: the evolution of 100 years of business improvement methodology. International Journal of Six Sigma and Competitive Advantage, 1(1), 4-20.

[28] Brewer, P. C., & Bagranoff, N. A. (2004). Near zero : defect accounting with Six Sigma. Journal of Corporate Accounting & Finance, 15(2), 67-72.

[29] Pan, Z., Park, H., Baik, J., & Choi, H. (2007, December). A Six Sigma framework for software process improvements and its implementation. In Software Engineering Conference, 2007. APSEC 2007. 14th Asia-Pacific (pp. 446-453). IEEE.

[30] Chakrabarty, A., & Tan, K. C. (2007). The current state of six sigma application in services. Managing Service Quality, 17(2), 194-208.

[31] Neuman, R. P., & Cavanagh, R. (2000). The six sigma way: How GE, Motorola, and other top companies are honing their performance. McGraw Hill Professional.

[32] Desai, T. N., & Shrivastava, R. L. (2008, July). Six Sigma: A Break through Business Improvement Strategy for Achieving Competitive Advantage A Case Study. In Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on (pp. 777-780). IEEE.

[33] Six Sigma.(n.d).Quality Resources for Achieving Six Sigma Results. Retrieved July 20, 2010 from isixsigma.com http://www.isixsigma.com/index.php?option=com

[34] Giese, J. L., & Cote, J. A. (2000). Defining consumer satisfaction. Academy of marketing science review, 1(1), 1-22.

[35] Collica, R. S. (2011). Customer Segmentation and Clustering Using SAS Enterprise Miner. SAS Institute.

[36] Akshay , Ankit , Laksha , Nilesh & Omkar (n.d.). Samsung mobile. Retrieved October 10, 2011, from http://www.slideshare.net/nileshrkalmegh1/samsung-mobile-ppt?qid=93b3f05b-da3c-4883-a76c-48f3066d7db6&v=default&b=&from_search=4

[37] Frank. (n.d.). Samsung Galaxy S2 Target Market Discussion . Buy custom Samsung Galaxy S2 Target Market Discussion essay. Retrieved January 17, 2012, from essay http://special-essays.com/essays/Business/samsung-galaxy-s2-target-market-discussion.html

[38] Pyzdek, T. (2003). The Six SIGMA Handbook: A Complete Guide for Green Belts, Black Belts, and Managers at all Levels, Revised and Expanded.

[39] Waddick, P. (n.d.). Quality Resources for Achieving Six Sigma Results. Retrieved January 17, 2012, from http://www.isixsigma.com/tools-templates/sampling-data/building-sound-data-collection-plan/

[40] Su, C. T., & Chou, C. J. (2008). A systematic methodology for the creation of Six Sigma projects: A case study of semiconductor foundry. Expert Systems with Applications, 34(4), 2693-2703.

[41] Maddox, K. (2004). Marketers embrace Six Sigma strategies. B to B, 89(10), 1-32.

[42] Rylander, D. H., & Provost, T. (2006). Improving the odds: combining six sigma and online market research for better customer service. SAM Advanced Management Journal, 71(1), 13.

[43] McInerney, D. (2006, January). Slashing product development time in financial service. In iSixSigma Magazine , 3(2), 1-3.

[44] Adams, C. W., Gupta, P., & Wilson, C. E. (2003). Six sigma deployment (Vol. 4). Routledge.

[45] Pak, A., & Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In LREC.

[46] Cios, K. J., Pedrycz, W., & Swiniarski, R. W. (1998). Data Mining and Knowledge Discovery (pp. 1-26). Springer US.

[47] Lovins, J. B. (1968). Development of a stemming algorithm. MIT Information Processing Group, Electronic Systems Laboratory.

[48] Ding, X., & Liu, B. (2010, August). Resolving object and attribute conference in opinion mining. In Proceedings of the 23rd International Conference on Computational Linguistics (pp. 268-276). Association for Computational Linguistics.

[49] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.

[50] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-(Vol. 10, pp. 79-86).Association for Computational Linguistics.

[51] Carroll, T. Z. J. (2008, January). Unsupervised Classification of Sentiment and Objectivity in Chinese Text. In Third International Joint Conference on Natural Language Processing (p. 304).

[52] Joachims, T., & Sebastiani, F. (2002). Guest Editors' Introduction to the Special Issue on Automated Text Categorization. Journal of Intelligent Information Systems, 18(2), 103-105.

[53] Han, J., & Kamber, M. (2006). Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan kaufmann.

[54] Goebel, M., & Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explorations Newsletter, 1(1), 20-33.

[55] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.

[56] Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003, August). Tackling the poor assumptions of naive bayes text classifiers. In ICML (Vol. 3, pp. 616-623).

[57] Sun, Y., Kamel, M. S., Wong, A. K., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. Pattern Recognition, 40(12), 3358-3378.

[58] Chang, R., Pei, Z., & Zhang, C. (2011). A modified editing k-nearest neighbor rule. Journal of Computers, 6(7), 1493-1500.

[59] Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah?. ACM SIGKDD Explorations Newsletter, 2(2), 1-13.

[60] Duda, R. O., Hart, P. E., & Stork, D. G. (1999). Pattern classification. John Wiley & Sons,.

[61] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[62] Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A Brief Survey of Text Mining. In Ldv Forum (Vol. 20, No. 1, pp. 19-62).

[63] Corney, M., de Vel, O., Anderson, A., & Mohay, G. (2002). Gender-preferential text mining of e-mail discourse. In Computer Security Applications Conference, 2002. Proceedings. 18th Annual (pp. 282-289). IEEE.

[64] McMeen, C. (n.d.). Text Processing Tutorial with RapidMiner. Retrieved May 15, 2013, from http://auburnbigdata.blogspot.com/2013/03/text-processing-tutorial-with-rapidminer.html

[65] Jardine, N., & van Rijsbergen, C. J. (1971). The use of hierarchic clustering in information retrieval. Information storage and retrieval, 7(5), 217-240.

[66] Francis, W. N., & Kucera, H. (1982) . Frequency analysis of English usage: Lexicon and grammar .

[67] Esuli, A., & Sebastiani, F. (2006, May). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of LREC (Vol. 6, pp. 417-422).

[68] Ghorashi, S. H., Ibrahim, R., Noekhah, S., & Dastjerdi, N. S. (2012 , July). A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews Extraction of Customer Reviews. IJCSI International Journal of Computer Science Issues, Vol. 9, 4(1) 1694-0814.

[69] Qiu, G., Liu, B., Bu, J., & Chen, C. (2009, July). Expanding Domain Sentiment Lexicon through Double Propagation. In IJCAI (Vol. 9, pp. 1199-1204).

[70] The Stanford Parser: A statistical parser. (2012, April 8). Retrieved from http://nlp.stanford.edu/software/lex-parser.shtml

[71] Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. Computational linguistics, 19(2), 313-330.

[72] Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In AAAI (Vol. 4, No. 4, pp. 755-760).

[73] Popescu, A. M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In Natural language processing and text mining (pp. 9-28). Springer London.

[74] Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). ACM.

[75] Lau, R. Y., Lai, C. C., Ma, J., & Li, Y. (2009). Automatic domain ontology extraction for context-sensitive opinion mining. In Proceedings of ICIS 2009, 35-53.

[76] Zhao, L., & Li, C. (2009). Ontology based opinion mining for movie reviews (pp. 204-214). Springer Berlin Heidelberg.

[77] Kaiser, C., & Bodendorf, F. (2009, September). Opinion and relationship mining in online forums. In Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on (Vol. 1, pp. 128-131). IET.

[78] Coronado, R. B., & Antony, J. (2002). Critical success factors for the successful implementation of six sigma projects in organisations. The TQM magazine, 14(2), 92-99.

[79] Pfeifer, T., Reissiger, W., & Canales, C. (2004). Integrating Six Sigma with quality management systems. The TQM Magazine, 16(4), 241-249.

[80] Andel, T. (2007). Lean & six sigma traps to avoid. Material Handling Management, 62(3 S 23428).

[81] Naumann, E., & Hoisington, S. H. (2001). Customer centered six sigma: Linking customers, process improvement, and financial results. ASQ Quality Press.

[82] Treichler, D., Carmichael, B., Kusmanoff, A., Lewis, J., & Berthiez, G. (2002). Design for six sigma: 15 lessons learned. Quality Progress, 35(1), 33-42.

[83] Kuei, C. H., & Madu, C. N. (2003). Customer-centric six sigma quality and reliability management. International Journal of Quality & Reliability Management, 20(8), 954-964.

[84] Harry, M., & Schroeder, R. (2005). Six sigma: the breakthrough management strategy revolutionizing the world's top corporations. Random House LLC.

[85] Goh, T. N., & Xie, M. (2004). Improving on the Six Sigma paradigm. The TQM Magazine, 16(4), 235-240.

[86] Miles, Lawrence D., Washington D.C., Eleanor Miles Walker,(1989) Techniques of Value Analysis and Engineering, 3rd ed Executive Director, Value Foundation.

[87] Cooper, R., & Slagmulder, R. (1997). Target costing and value engineering. Portland, OR: Productivity Press.

[88] Arveson, P. (n.d.). The deming cycle. Retrieved September 15, 2011, from http://www.balancedscorecard.org/thedemingcycle/tabid/112/default.aspx

[89] Deming, W. E. (2000). The new economics: for industry, government, education. MIT press

[90] Basu, R. (2004). Implementing quality: a practical guide to tools and techniques: enabling the power of operational excellence. Cengage Learning EMEA.

[91] Samsung GALAXY S III designed for humans inspired by nature. (n.d.). Retrieved September 15, 2011, from http://www.samsung.com/global/galaxys3

[92] Samsung GALAXY S II designed for humans inspired by nature. (n.d.). Retrieved September 15, 2011, from http://www.samsung.com/global/galaxys2