SENTIMENT ANALYSIS AND FEATURE-BASED MINING OF CUSTOMER PRODUCT REVIEWS

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering

in

Computer Science and Engineering

Submitted By

Zeenia Singla

(Roll No. 801532060)

Under the supervision of:

Dr. Sushma Jain

Assistant Professor

Ms. Sukhchandan Randhawa

Lecturer

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY PATIALA – 147004

July 2017

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, **"**Sentiment Analysis and Feature-Based Mining of Customer Product Reviews**"**, in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Sushma Jain* and *Ms. Sukhchandan Randhawa* and refers other researchers' work which are duly listed in the reference section. The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

Signature:
(Zeenia Singla)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Sushma Jain)
Assistant Professor,
CSED

(Ms. Sukhchandan Randhawa)
Lecturer,
CSED

ACKNOWLEDGEMENT

The tenure from the beginning and the completion of my thesis has been a great learning paradigm, which has not only honed my research and problem-solving skills but has also widened my knowledge base. My thesis would not have been completed without the constant support and right guidance of my advisors, whom I would like to thank for.

First and foremost, I would like to offer my sincerest gratitude to my supervisors, **Dr. Sushma Jain** and **Ms. Sukhchandan Randhawa**, who have always motivated and supported me throughout my thesis. They have always shown me the right path to achieve my objectives with their vast knowledge and insurmountable patience. Both of my supervisors have provided constant support and made all the resources available at my disposal.

I would also like to thank **Dr. Maninder Singh**, Head, Computer Science and Engineering Department, Thapar University for his support and cooperation. I acknowledge the efforts of the complete staff and faculty of Computer Science and Engineering Department, Thapar University to provide me with adequate facilities required for the completion of this thesis.

Finally, I would like to express my gratitude to my peers for stimulating discussions and family members for always inspiring me which kept me going.

Date: July 2016 Zeenia Singla

Place: Thapar University, Patiala 801532060

ME (CSE)

ABSTRACT

Online reviews have the potential to provide an insight to the buyers about the product like its quality, performance and recommendations; painting a clear picture of the product in front of the future buyers. Sentiment Analysis is a computational study to extract subjective information from the text. In this research, data analysis of a large set of online reviews for mobile phones is conducted. Variegated techniques have been used to perform classification of the text, namely, Lexicon-based approach using sentiment dictionary, Supervised Learning using Naïve Bayes, Support Vector Machine (SVM) and Decision Tree classifiers, Deep learning and Feature-based extraction of reviews using association rule mining. The text is not only classified into positive and negative sentiments but also represents eight different emotions using lexicon-based approach. This delineated classification of reviews is helpful to evaluate the product holistically, hence enabling better-decision making for consumers. Supervised learning is performed using binary and multi-class classification. The performance of SVM is the best in both with higher accuracy in multi-label data. To improve the efficiency of classification, deep learning has been used which classifies the data into two classes. It extracts subjective meaning of the text along with the negation effect. Two different techniques, Vocabulary-based Vectorization and Feature Hashing have been employed. Vocabulary-based Vectorization outperforms Feature Hashing and there is a considerable rise in accuracy as compared to supervised learning. Feature-based extraction is employed to determine the performance of various features of mobile phones which aids the customers to make well-informed decisions and also highlights various flaws on which a product designer can work upon to improvise the product.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Due to the rapid growth of electronic commerce, online reviews have replaced the traditional “word-of-mouth” and have been playing a vital role in influencing the consumer’s buying patterns and sales of a product. Reviews act as a trust-building platform for the consumers where by judging the previous buyers' experience they are able to make informed decisions. From the manufacturer’s point of view, helpful online reviews are crucial to mine customer requirements for improving a product or designing a new product. By capturing relevant online reviews, manufacturers can adhere to the customer requirements in the target market. Manufacturers also get an insight to the competitive market and ongoing trends influencing their marketing decisions.

Retail websites like Amazon.com offer different options to the reviewers for writing their reviews. For instance, the user can provide rating in the form of numerical stars (usually ranging from 1 to 5 stars) or open-ended customer-authored comments about the product as shown in Figure 1.1.

Figure 1.1: Sample of Reviews posted on Amazon.com

The presence of online reviews on a website is believed to increase the user credibility, attract consumer visits, augment hit ratio and time spent on the site. The discovery platforms like Zomato and Trivago are booming just on the basis of user reviews provided on restaurants and hotels. Reliable customer reviews build a trust factor among the novice users and help to enlarge the customer base. Both positive and negative reviews help the consumers and the manufacturers where manufacturers can take negative feedback constructively and know about the areas which they need to work upon to improve their product or service.

Traditionally, companies used to rely upon online surveys and questionnaires to understand consumers' requirements. However, in today's digitalized world where online reviews have covered a vast sphere of e-commerce, conducting online surveys has become far more time-consuming and at the same time providing less information. In the case of online reviews, companies can get hold of real-time data where the users are participating voluntarily. This volunteer participation is more accurate than coercing the users to provide reviews by filling surveys.

1.1 Statistical Analysis

Statistical analysis of the data helps to analyze and visualize data providing deeper insights and better understanding of data. It aids to realize the trends among different features and make data-driven decisions. The main aim of statistical analysis is to identify the ongoing trends in data and find patterns in structured, semi-structured and unstructured data which can be utilized to enhance user experience. Statistics is applied in almost every field be it research, industry, real estate or government sector. Exploratory analysis helps to determine which part of the data is non-essential or unimportant and is usually represented through outliers and unusual clusters, etc. The two methodologies followed for statistical analysis are *Descriptive* and *Inferential Statistics* [1].

1. **Descriptive Statistics:** It is the mean or standard deviation derived from the data. Descriptive statistics aids in comparing statistics between co-related features of the dataset, thereby highlighting the differences or similarities between the different columns in the data.
2. **Inferential Statistics:** Inferential statistics involves the use of different techniques to perform exploratory analysis and draw conclusions. For example, election polling results are predicted by sampling the voting population of

different constituencies and then a part of the population is asked who they will vote for. Their responses are fed into the prediction model to draw conclusions.

1.2 Helpfulness of Online Reviews

With an ongoing increase in the number of consumers choosing to buy online, the number of reviews is also increasing tremendously. Such a large volume of reviews becomes cumbersome for the users to go through and form an opinion about the product. Out of this large database of reviews, there are ample number of reviews which might not be that useful to the consumers. So, filtering out reviews that will be helpful to the users is essential to provide access to quality data in less-time. *Helpfulness of reviews* is determined by using some textual features in the dataset like number of question marks, exclamation marks, words appearing in all caps, helpful votes received, length of the review, reviewer rating and number of negative words included in the text, etc. Reviews having many exclamation marks or question marks or too many words with all caps are said to be too *extreme* and hence less helpful. Amazon.com usually asks the user "Was this review helpful to you?" to collect the number of helpful votes associated with a particular review and displays reviews in decreasing order of their helpfulness rating. So, helpful reviews can be filtered out by keeping a threshold value of helpful votes received. Even reviewer's information can be a useful criterion for knowing the credibility of a review like how often the reviewer posts reviews, number of up votes received by the reviewer and writing style of the review.

Mudambi and Schuff [2] explain that the *type of the product* also helps to determine helpfulness of online reviews, for instance, *experience goods* like movies and music have extreme ratings and vey less moderate ratings as the reviews are subjective in nature and consumers simply like the product or don't like it. So, in these cases objective reviews or reviews with moderate rating are considered to be more insightful. *Search goods* are those goods whose quality and performance of functional attributes can be known before actually buying the product. Goods like cars, mobile phones and electronic devices come under this category. So, reviews of search goods provide more useful feedback regarding the performance of the product and functional performance of different attributes like camera, battery, screen and speakers, etc. is determined in case of mobile phones. For search goods extreme reviews can be more credible and would provide better judgement than objective reviews or moderate reviews.

1.3 Machine Learning

Artificial intelligence has paved a way for machines to learn without the need of programming them explicitly. This ability of the machine to learn by itself is known as *Machine Learning*. The learning helps computers to find new insights because when a machine is fed with new data it is able to adapt itself accordingly. The previous learning of the machine helps to make valuable predictions and results. Complex mathematical computations are performed on large volumes of data which has been used to make meaningful predictions, for instance, online recommendation offers on e-commerce websites like Amazon.com have made online shopping experience much more exhilarating based on the items viewed recently. Machine learning is the commonly used approach for text mining and is based on statistical and mathematical models.

Feature selection is the process of selecting a subset of relevant features to train and test the model. Feature selection aids in improving the performance of the machine learning models as relevant features will produce useful outcomes and predictions. It also reduces training time considerably and overfitting of data. Feature selection [3] can be classified mainly into three categories, *Filter methods, Wrapper methods* and *Embedded methods*.

- **Filter methods:** Filter methods means that the features are selected on the basis of some statistical analysis and their correlation. These features are independent of any machine learning algorithms.
- **Wrapper methods:** In these methods, a subset of features is trained on a machine learning algorithm. It is a sequential method in which inferences drawn from the previous model are used to add or remove features and train the model again.
- **Embedded methods:** It is the combination of both filter and wrapper methods in which models like LASSO and Ridge regression are used. These models have feature-selection methods embedded in them.

Classification is the process of classifying data into classes or categories. Supervised learning of data is conducted using classification. The observations are divided into a set of quantifiable properties or features. These classes may be *categorical* (A, B or C), *ordinal* (large, medium or small), *integer-valued* or *real-valued*. There are several classifiers that help to classify data in machine learning. Classification can be *binary* or *multiclass classification*. Binary classification involves only two classes whereas multiclass classification involves more than two classes. Commonly known classifiers are linear classifiers (Naïve Bayes), Support Vector Machine (SVM), Decision Tree, Neural Network, etc. The evaluation metrics used to measure the performance of a classifier system are *precision* and *recall*.

In some scenarios, a particular class of data dominates the other class that can yield biased predictions and misleading accuracies. Data having such anomalies is considered to be imbalanced. This is because the machine learning model is not able to learn about the minority class efficiently. The imbalanced classification of data is prevalent mostly in binary classification rather than multi-level classification. So, in order to produce accurate predictions data needs to be balanced where the weightage of each class is almost the same. This enables the classifier to extract all the necessary information required for every class present in the dataset. The methods available to convert imbalanced data into a balanced one are enlisted in Figure 1.2.

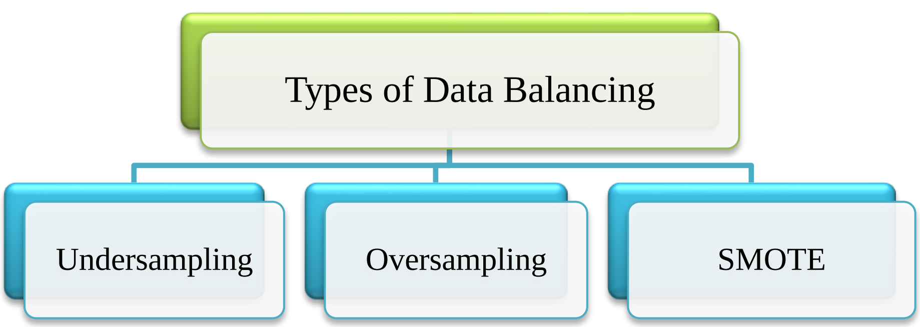


Figure 1.2: Data Balancing Techniques

- **Undersampling**: In this method, the observations of majority class are reduced to balance the dataset. It can be further classified into two types: *Random* and *Informative.* In the former undersampling technique, a random set of observations is chosen and eliminated from the dataset. In the latter case, the set of observations to be eliminated is selected on the basis of some pre-defined criteria.

- **Oversampling:** Oversampling is just the opposite of undersampling in which the observations of minority class are replicated to balance the data. It is also of two types: *Random* and *Informative*. In random oversampling, the number of observations to be replicated are chosen randomly from the minority class, whereas, in informative oversampling, the observations are selected using some criteria.
- **Synthetic Minority Oversampling Technique (SMOTE):** SMOTE is a type of oversampling technique which generates data synthetically for data balancing. It creates artificial data using feature space and K-nearest neighbors. Difference between each feature vector and its nearest neighbor is calculated which is then multiplied by a random number between 0 and 1. This number is added to the feature vector. SMOTE technique is highly used and the most preferred one for data balancing in most of the scenarios.

Cross validation helps to evaluate the ability of the learning algorithm to classify the unseen data. There are three ways to cross validate the learner's performance, i.e., *Holdout Method, K-fold Cross Validation* and *Leave-one-out-cross validation*.

- **Holdout Method:** It is the naïve method of cross validation in which the data is divided into two subsets, training and testing dataset (unseen data). The machine learning algorithm is trained using the training set and then it makes predictions for the testing dataset.
- **K-fold Cross Validation:** In this method, *k*-subsets of the dataset are formed and the holdout method is repeated *k*-times. For every iteration, *k*-1 subsets are used for training purposes and one of the *k* subsets is used for testing. It is the improved version of holdout method.
- **Leave-one-out Cross Validation:** This cross validation technique uses all the data for training leaving one point. That one point is used to test the model.

1.5 Sentiment Analysis

Sentiment analysis is the process of identifying the sentiment of the text using natural language processing. The polarity of the text can be binary or multiclass. Binary classification of the text is divided into positive and negative. Multiclass classification of the text involves more than two classes like positive, negative, neutral, very positive and very negative. Sentiment analysis can be conducted at document level, phrase level or word level. It helps to know about people's viewpoint regarding a particular object or service. Sentiment analysis has a wide scope in terms of application in variegated areas like manufacturing, food services, government and medical sector, etc. For example, if a phone's rating is low, sentiment analysis can help to determine what all negative reviews that particular phone has received and even Figure out why the particular product could not meet users' expectations. Different types of sentiment analyses are demonstrated in Figure1.3.

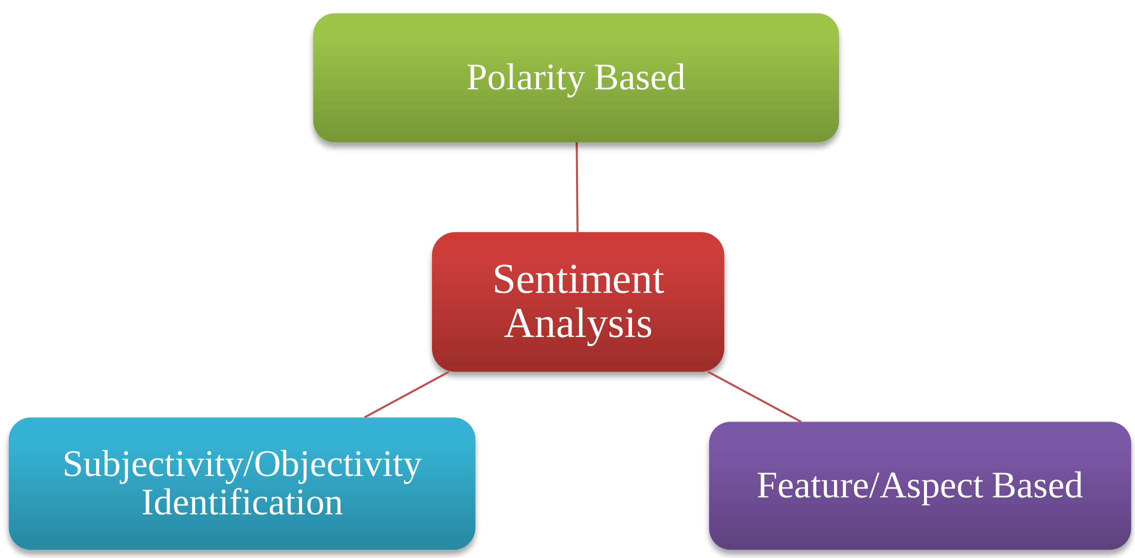


Figure 1.3: Types of Sentiment Analysis

- **Polarity-Based:** In polarity based sentiment analysis, feelings of the users are determined using sentiment oriented words like angry, happy, dislike, and unsatisfied. Phrases having more positive words than negative words are generally classified as positive and vice-versa. Pang and Lee [4] described that the sentiment can also be determined by rating the reviews on a scale of -10 to +10. Using numbers to represent sentiments gives a clearer picture as sentiment value can be adjusted as per the environment of the phrase.

- **Subjectivity/Objectivity Identification:** The text is classified into two classes: subjective or objective as performed by Pang and Lee [5]. Subjective identification implies reviews influenced by emotions. Objective identification implies reviews not influenced by personal emotions or views.
- **Feature/Aspect Based:** Liu [6] implemented feature based classification which is driven by opinions associated with different features of a product. For example, battery life of a mobile phone, quality of food of a restaurant and average speed of a car. This type of analysis helps to target a particular segment of customers and provide information to the users regarding a particular feature of the product. Feature based sentiment analysis is very useful from product designer's perspective in order to improvise the product.

Various techniques like machine learning approaches and lexicon-based approaches [7] are used to conduct sentiment analysis as shown in Figure 1.4. Machine learning approaches used for sentiment analysis are:

- **Supervised Learning:** In supervised learning, models are trained using labelled or classified data. The set of inputs along with the outputs is fed into the model. The model then compares its actual output with the correct output and hence learns from the data. After training the model, unseen data is fed into the system without knowing the class labels and the algorithm makes predictions using patterns. It can be used to make weather prediction, anticipate fraudulent transactions, increase in demand of a particular product and rise in inflation, etc. from historical data. Classifiers used in supervised learning are linear classifiers like SVM, rule-based classifiers and probabilistic classifiers like Naïve Bayes.

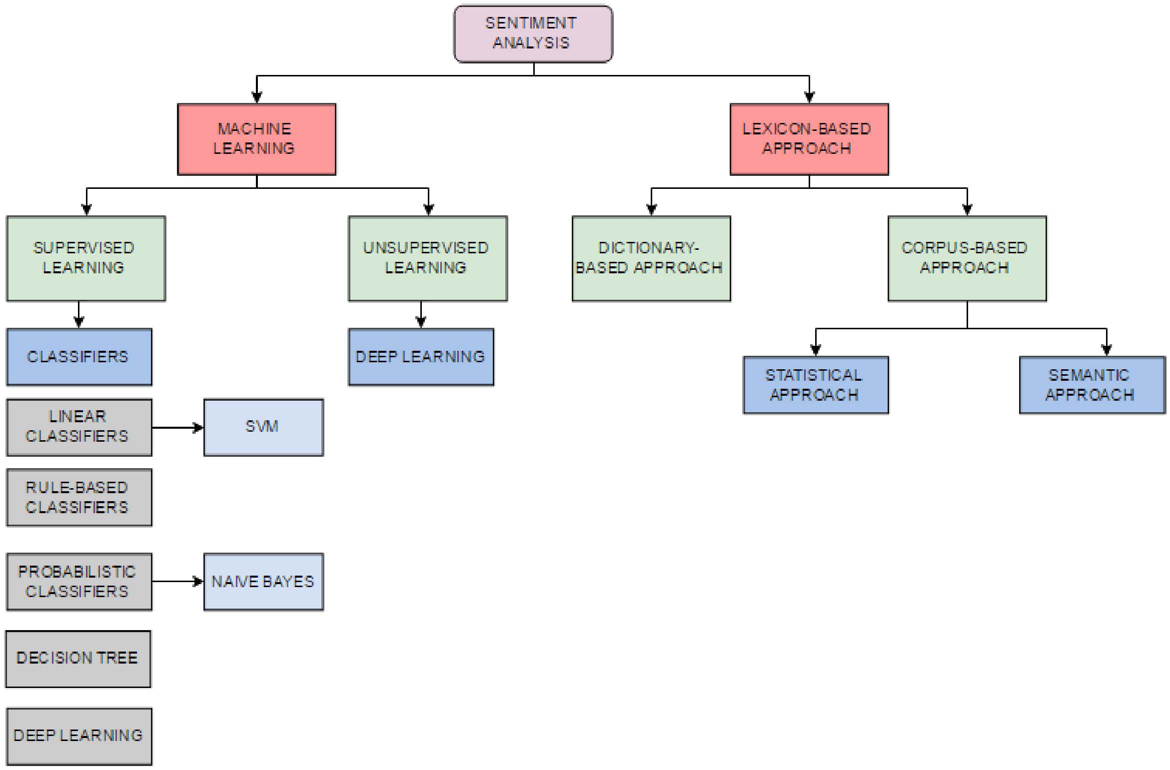


Figure 1.4: Sentiment Analysis Techniques

- **Unsupervised Learning:** Unsupervised learning is employed on data with no labels. It performs exploratory analysis on its own and tries to find hidden patterns of data. Unsupervised learning is most suitable for transactional or market-basket data. Techniques used for this kind of learning are: *self-organizing maps*, *k-nearest neighbors*, *k-means clustering* and *deep learning*. This particular learning can be useful to determine mobile phones having similar characteristics or similar flaws which can then be recommended to the consumer on the basis of his/her preference.
- **Lexicon-Based Approach:** In this approach, lexical features of the text are used to classify the text. Lexicon-based approach uses vocabulary of words or parts-of-speech to understand the subjectivity of an opinion. In dictionary-based approach positive and negatives words are listed in the dictionary which are then matched with the lexicon provided [8]. In corpus-based approach [9], domain dependent orientation of the text is determined. Statistical approach determines if the word occurs more frequently in the positive text, then its polarity is considered as

‘positive’ and vice-versa. If the frequency of occurrence of words in both positive and negative text is same then it is considered as ‘neutral’. Semantic approach determines sentiment consistency on the basis of conjunctions to classify words. Adjectives conjoined with conjunctions like and, or, either-or and neither-nor are considered to be on similar constraints.

1.6 Deep Learning

Deep learning is a novice technique used in machine learning that includes more than one layer between input and output in which output from the previous layer is taken as input in every successive layer. It is a big neural network whose performance gets better with larger data. Deep learning can be implemented for both supervised and unsupervised learning and is known to give better efficiency than simply using classifiers for classification purposes. It follows a hierarchical structure to do feature learning and make intermediate representations as depicted in Figure 1.5. Deep learning has the ability to learn intermediate concepts between raw input and target which would produce far better understanding of the textual data. Deep learning is best used for clustering analog data. It can also be used for raw text.

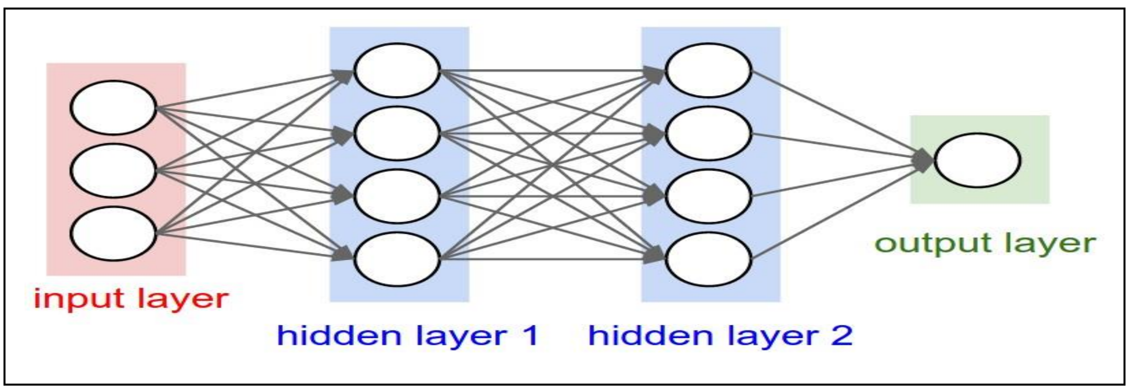


Figure 1.5: Structure of Deep Learning

In brief, deep learning extracts features from the data on its own without any need of human intervention. The ability of deep learning to learn from unlabeled data with good accuracy is what makes it different from the previous algorithms available in machine learning.

Text processing in deep learning can be implemented by using Word2Vec and Doc2Vec [10]. Word2Vec is used to convert textual data into numerical data (vectors) hich can be understood by deep neural networks. Doc2Vec considers every review as a separate document and attaches labels to them. To perform classification, labeled data are required in Doc2Vec. Doc2Vec establishes relationships between labels and words instead of words with other words. Every document is converted to a vector representing the meaning of a document. Doc2Vec tool is helpful in training the model. Mukku et al. [11] explicated the use of Doc2Vec in which fixed-length features are created from sentences having variable length, whereas in Word2Vec, variable length vectors are created. In machine learning algorithms fixed-length features are required to train the model, so Doc2Vec is more feasible than Word2Vec.

Different architectures that can be used to implement deep learning are: *Deep Dense Networks, Recurrent Neural Networks, Recursive Neural Networks* and *Convolutional Neural Networks.*

- **Deep Dense Networks:** These networks uses Word2Vec model which are used by Mehr [12]. This model can make clusters of words having similar meanings and classify them by topic. It makes use of average of word vectors which is trained using a word2vec model. The vector representation of sentences encapsulates the meaning of the sentence which is passed as a feature to the classifier. Deep dense networks do not determine the relationship between words in a sentence; hence its performance is not so good in comparison to other deep learning techniques but is better than other state-of-the-art classifiers.
- **Recurrent Neural Networks:** These networks make use of sequential word vectors whose output depends on previous vectors. Every word in a sentence is represented by single layer in the network. If a sentence has six words then there will be six layers in the network. This model helps to generate new text by using output probabilities. Mehr [12] demonstrates the performance of this network low for sentiment classification.
- **Recursive Neural Networks:** In this network, same set of weights is applied recursively over a structure. These networks are able to understand the

relationship between words better. Recursive neural networks are better than recurrent neural networks.

- **Convolutional Neural Networks:** It is a feed-forward artificial neural network which consists of multilayer perceptron to convert input into output. Different filters are applied to generate different outputs. It is a generalized form of recursive neural network.

1.7 Structure of Thesis

The structure of the thesis is as follows.

- **Chapter 2:** This chapter includes the literature survey of the presented work and the different techniques used for sentiment classification. It helps to analyze the gaps in previous work done.
- **Chapter 3:** This chapter lists the gaps in previous work, explains the problem statement and the main objectives of the thesis.
- **Chapter 4:** Based on the problem statement, the chapter deals with the approach followed to solve the problem and meet the objectives. It provides the detailed workflow model and various techniques applied for sentiment classification such as Lexicon-Based Approach, Supervised Learning and Deep Learning.
- **Chapter 5:** This chapter demonstrates the implementation results of the different techniques used for sentiment classification. The results are then analyzed to find out the most technique for sentiment analysis. It also lists the results of feature-based mining of product reviews.
- **Chapter 6:** Finally this chapter provides the conclusion drawn from the results obtained and the summary of contributions made through this thesis. It also explains the future scope of the proposed work.

Chapter 2

STATE OF ART

Data Analytics has enabled users to unravel the hidden patterns in data. Big data provides insights on consumer behavior which can be used to make informed decisions. Erevelles *et al*. [13] discussed about the three resources, human, physical and organizational capital, that organizations can exploit to better understand the role of Big Data in marketing activities. This involves collecting consumer data, extracting consumer insights and thereby, using these insights to improve the functionalities. An average consumer is generating both structured and unstructured data which are transforming the way decisions are taken in the industry. Big Data so generated are defined using three dimensions: *Volume, Velocity* and *Variety* as mentioned in George *et al.* [14], Strong and Colin [15]. The volume and the relentless rapidity at which data are being generated every day are exceeding the computing capacity of many IT departments. Four more Vs that play an important role in explaining big data are: *Veracity, Value, Variability* and *Visualization*. Veracity adds to the noise and abnormality in data that degrades the quality of data in question. Value of Big Data lies in the information and insights revealed about the data using exploratory analysis. Variability implies the whimsical nature of Big Data whose meaning keeps on changing over time. After the processing of data, to be able to visualize such a large amount of data with dozens of variables and parameters is a big challenge in itself. By filtering the irrelevant data, remaining data can be utilized to provide valuable business insights.

Big Data has enabled businesses to flourish and improvise on the basis of evidence rather than intuition. It aids in gaining insights on better targeted social influencer marketing, segmentation of customer base, recognition of sales and marketing opportunities, detection of fraud, quantification of risks, better planning and forecasting and understanding consumer behavior, etc. as proposed by Russom [1], Stone and Merlin [16]. With the increasing reliability of consumers on online reviews, large and voluminous amount of reviews are being generated with relentless rapidity which need to be processed and extracted to provide useful information.

2.1 Extraction of Reviews

In the previous studies it was found that the information extracted from other sources like online or offline surveys was not so rich as compared to the information gained from online reviews as shown by Korfiatus *et al*. [17] and Sher and Lee [18]. Several research papers in the past have illustrated that large-scale information sharing on digital networks has been able to build a community of trust among buyers and sellers in online markets influencing the probability of purchase. For instance, Resnick and Zeckhauser [19] exhibited how a seller's reputation gained from online reviews affects the probability of the sale of the product on eBay.

There are a variety of measures which can used to measure the performance of online reviews but helpfulness is the most widely used measure as described by Ghose and Ipeirotis [20], Mudambi and Schuff [2]. Ghose and Ipeirotis [20] mentioned that reviews can be classified into two categories: *objective reviews* and *subjective reviews* and it was observed that for experience-based goods, like movies, users preferred to have an access to subjective reviews giving them an insight of their views regarding the movie. On the other hand, for feature-based goods such as electronics, users require objective reviews that confirm to the features of the product stated, thereby increasing the viability of the product.

The research conducted by Mudambi and Schuff [21] elucidates that even the extremity level of reviews affects the helpfulness. It was determined that for experience goods reviews with extreme ratings are less helpful than reviews with moderate ratings, which is not the case in feature-based goods. Chen *et al*. [22] suggests that online "rate the reviewer" systems like on Amazon.com help to strengthen the trust of potential buyers apart from aggregate quality scores retrieved from product reviews. While Zhu *et al*. [23] posits that the primary determinants of perceived helpfulness are reviewer expertise and helpfulness. The length of the reviews is also said to be a determinant of review helpfulness where longer reviews are found to be more helpful to users as compared to short and concise reviews as shown by Kuan *et al*. [24]. The more the number of up votes given to a review, the more helpful it is.

It has also been noted that the publication date of a review has positive correlation with its total number of votes, commonly referred to as *e-bias* as demonstrated by Liu *et al*. [25]. It implies that the older reviews receive more votes because they have been on the website for a longer time and are more likely to be read by the users. Amazon.com sorts reviews by the "most helpful first" instead of the "newest first" Salehan and Kim [26], so the latest reviews are less likely to be observed by the users. From the designer's perspective, latest reviews will be more preferable because they will provide him with the latest demands and the ongoing trends in the market.

Identifying fake reviews has also been an issue of concern on these giant web forums like TripAdvisor.com and Expedia.com where users post negative reviews for their competitors as explained by Mazylin *et al*. [27] and study of Hu *et al*. [28] reports that 10.3% of book reviews are fake or manipulated. In Mazylin *et al*. [27], fake reviews were eliminated by allowing only those consumers to post a review who had at least stayed overnight in the hotel. So, determining these fake reviews and filtering them from the dataset is also a need of the hour while determining helpfulness of reviews.

Most of the research has been carried out on determining helpful online reviews keeping in mind consumers' needs. However, a very limited work has been done to find helpfulness of online reviews from designer's perspective. One such research has been carried out in Liu *et al*. [29] proposing four types of features that can address the designer's concerns. The four features defined are: *linguistic features, product features, features based on information quality* and *features based on information theory*.

2.2 Review of Sentiment Analysis

Sentiment analysis, also known as *opinion mining*, means identifying the sentiments of the users on the basis of positive, negative and neutral connotations. Opinion mining can be classified into three different levels: *document level, sentence level* and *phrase level* as described by Liu and Bing [30]. Sentiment analysis is conducted on the basis of product features and sentiments expressed in reviews to extract the summary of reviews. Yang *et al*. [31] explained two major approaches are followed to perform feature extraction:

Supervised and *Unsupervised learning*. Supervised product feature extraction involves training and testing dataset. Training dataset is a set of pre-annotated data which is used to train the model. Yang *et al.* [32] employed product association rules and Naive Bayes classifier as two supervised learning algorithms for feature-based extraction of reviews. A lot of prior research has been done in this field, for instance, Hatzivassiloglou and McKeown [33] classifies words and phrases with prior positive or negative polarity. This prior classification is helpful in many cases but when contextual polarity comes into the picture, the meaning derived from positive or negative polarity can be entirely different. For example, the word 'amazing' has a prior positive polarity and the word 'degrade' has a prior negative polarity. However, they may be used with negation words like 'not' change the context completely and sometimes phrases containing negation words intensify rather than changing the polarity. For instance, the product delivered was not only good but also amazing in terms of looks. This contextual polarity of the phrases was taken into consideration in Wilson *et al*. [34] and ambiguity was removed. Pang and Lee [35] used a refined method to establish contextual polarity of phrases by using subjective detection that compressed reviews while still maintaining the intended polarity.

Delineated study has been conducted on tweets available on Twitter, movie reviews, etc. to build the grounds on sentiment analysis and opinion mining. A sentiment classifier has been built to categorize positive, negative and neutral sentiments not only in English but also for other languages using corpus from Twitter like shown in Pak and Paroubek [36]. Tan and Zhang [37] performed sentiment analysis of Chinese text by implementing four feature selection methods (Mutual Information (MI), Information Gain (IG), Chi Square (CHI) and Document Frequency (DF)) and five classifiers viz. Centroid classifier, K-nearest neighbor, Winnow classifier, Naive Bayes and SVM. Through this learning paradigm it was concluded that IG is the best feature selection method and SVM outperforms all the other learning methods in terms of sentiment classification. Similarly, Sharma and Dey [38] performed sentiment analysis on movie reviews by using five feature selection methods (DF, IG, Gain Ratio (GR), CHI and Relief-F) and seven classifiers (Naive Bayes, SVM, Maximum Entropy, Decision Tree, K-Nearest Neighbor, Window, Adaboost). In this case also SVM performed better than all the other techniques. Ye *et al.* [39] performed sentiment analysis on travel reviews using three machine learning models namely, Naive Bayes, SVM and character based N-gram model in which SVM and N-gram approaches have better performance than Naive Bayes.

Most of the times, SVM showcases best performance in comparison to other classification models. However, the case study of Zhang *et al.* [40] demonstrated otherwise. In this research work, Cantonese (Chinese) restaurant reviews were classified using Naive Bayes and SVM in which knowledge independent machine learning method is implemented. Here, the Naive Bayes classifier astonishingly performed better than SVM. Wahyudi and Kristiyanti [41] determined the polarity of smartphone product reviews only on the basis of positive and negative orientation of the review. Devi *et al.* [42] established a system using SVM where sentiment analysis is carried out by taking into consideration sarcasm, grammatical errors and spam detection. Narayanan *et al.* [43] performs sentiment analysis by using an enhanced Naïve Bayes model by combining methods like effective negation handling, word n-grams and feature selection.

Sometimes, the lexicon of sentiment words is not able to accommodate words related to a particular domain accurately which affects the classification accuracy of classifiers like Naive Bayes and SVM. In order to overcome this problem, Kang *et al.* [44] came up with a new sentiment lexicon for restaurant reviews and used unigrams and bigrams together as features, thereby narrowing the gap between positive accuracy and negative accuracy. Deriu *et al.* [45] predicted sentiment orientation of multi-lingual short texts using deep learning. Huge chunk of weakly supervised data is used to train multi-layer convolutional network and it is found that the single-language approach is better as compared to the multi-language approach in terms of performance. Zárate *et al.* [46] put forth an aspect-based sentiment analysis approach by filtering words around the aspect using N-gram methods like N-gram after, N-gram before and N-gram around. The performance of N-gram around was the best in terms of precision.

On the other hand, unsupervised feature-based extraction extracts product features without explicitly training the model. In such feature-based extraction techniques, Part-of-Speech (POS) tagging is done to fetch features on the basis of nouns or noun phrases present in sentences. Then, association rule mining algorithm is implemented to extract rules on the basis of reviews as explained in Hu and Liu [47]. Hu and Liu [48] further built a system called Opinion Observer which clearly distinguishes strengths and weaknesses of a product as per different features. Here, the feature-extraction depends on the count of occurrence of terms. Apriori algorithm is used for association rule mining to extract frequent item sets. Popescu *et al.* [49] devised a system that fetches noun phrases having frequency greater than the threshold frequency and then extracts product features and opinions. Wei *et al*. [50] proposed a semantic-based feature extraction on the basis of positive and negative adjectives to determine opinion words and extract features mentioned in the reviews. Zhang *et al*. [51] explicates an algorithm to rank different products according to their features. Firstly, relevant product features are established. Then subjective and comparative sentences are extracted and a feature-specific product graph is constructed to display relative quality of products. Finally, a page-rank algorithm is used to rank the products.

Xu *et al.* [52] combines chunk features and heuristic position information along with word features, part-of-speech features and context features, hence improving the product opinion mining considerably. Yang and Shih [53] employs class association rule mining algorithm to extract rules that can help to extract opinion-based reviews related to particular features of a product. Khan *et al.* [54] proposed a novel approach to extract product features and opinionated words through auxiliary verbs like is, are, was, has, have, had and were. The findings of this research established that more than 80% of features and opinionated words have auxiliary verbs, hence proving that auxiliary verbs are also good indicators of features and opinions apart from nouns, verbs, adjectives and adverbs. Zhai *et al.* [55] performs structure analysis of reviews having high values of precision and recall. A database of sentiment patterns has been constructed and matched with each review sentence, thereby extracting corresponding features of products. Apart from Sentiment Analysis, information quality of the review is also essential for meaningful extraction. To eliminate noisy reviews, Chen and Tseng [56] have implemented an information quality framework to perform feature-based extraction.

Tang *et al.* [57] conducts aspect level sentiment classification using deep memory network. Every context word is taken into consideration while determining sentiment polarity of an aspect which is calculated using multiple computational layers. Luo *et al.* [58] leads feature based sentiment classification using SVM. The proposed work employs sentiment vector-space model to represent the text and addresses different issues faced during sentiment classification like selecting classification algorithms, feature extraction techniques and feature dimension. Feature selection method of Chi-square Difference between the Positive and Negative Categories (CDPNC) was made that improved the performance of SVM classifier.

Deep Learning is known to be one of the finest methods for sentiment analysis that outperforms all the previous methods of sentiment analysis. Deep learning performs unsupervised learning and classifies text with high accuracy. Mehr [12] compares different deep learning techniques viz. recurrent neural network, recursive neural network, convolutional neural network along with state-of-the-art Naïve Bayes' classifier's performance in classification of movie reviews. Convolutional neural network fed with word2vec vector of words is found to be the most accurate one among all. In Socher *et al.* [59] Recursive Neural Tensor Network has been used to classify sentences into negative and positive and it outperformed all the basic classifiers used to classify the text. Mukku *et al.* [11] uses Doc2vec to convert sentences in the text into vectors to perform sentiment classification of Telugu text and then different machine learning techniques are applied to classify the data. Abdelwahab and Elmaghraby [60] used both word2vec and doc2vec models to train the tweets. Word2vec model was trained using bag-of-words architecture and doc2vec model was trained on paragraphs of positive, negative and neutral tweets. It was then used to calculate the average word vector for each tweet which outperformed other classifiers like SVM, Logistic Regression and Boosted Trees.

Sentiment Analysis is not only confined to reviews or Twitter data but is also applicable on stock markets Yu *et al.* [61], Hagenau *et al.* [62], news articles Xu *et al.* [63] or political debates Maks and Vossen [64]. In Yu et al. [61] the intensity of emotion words has been calculated to improve the classification performance by using contextual entropy model that measures the similarity between two words. Hagenau *et al.* [62] selects semantically relevant features for classification and reduces the effect of overfitting on machine learning approaches. Sentiment analysis can be used to flourish consumer products related business as proposed by Qiu *et al.* [65]. It uses rule-based approach for sentiment analysis to extract topic words of negative opinion sentences and thus promote the competitors of the products receiving negative feedback. Similarly, relevant ads based on a person's liking or disliking are displayed on various blogging sites targeting bloggers. Fan and Chang [66] concocted a Blogger-Centric Contextual Advertising Framework to determine users' personal interests and display those ads that intersect with them. Table 2.1 showcases the summary of research work previously done on sentiment analysis.

Table 2.1: Summary of Sentiment Analysis

S.No.	Studies	Classification Technique Used	Models Used	Outcome
1.	Tan and Zhang [37]	Supervised Learning	Centroid, K-NN, Window, Naïve Bayes and SVM	SVM outperformed others in sentiment classification
2.	Sharma and Dey [38]	Supervised Learning	Naïve Bayes, SVM, Maximum Entropy, Decision Tree, K-NN, Window and Adaboost	SVM performed best in sentiment classification and performance of Naïve Bayes improved with less features
3.	Ye *et al.* [39]	Supervised Learning	Naïve Bayes, SVM and character based N-gram model	SVM, N-gram model better than Naïve Bayes in sentiment

				classification
4.	Zhang *et al.* [40]	Supervised Learning	Naïve Bayes and SVM on Chinese text	Naïve Bayes performed better in sentiment classification
5.	Devi *et al.* [42]	Lexicon-based Approach	SVM using sarcasm, grammatical errors and spam detection	Efficient SVM model for sentiment analysis
6.	Narayanan *et al.* [43]	Lexicon-based Approach	Naïve Bayes model using negation handling, word n-grams and feature selection	Fast and accurate Naïve Bayes model for sentiment analysis
7.	Kang *et al.* [44]	Lexicon-based Approach	Unigram and Bigram as features	Gap narrowed between negative and positive accuracy
8.	Deriu *et al.* [45]	Deep Learning	Convolutional Network on multi-lingual text	Single-language approach better than Multi-language approach
9.	Zárate *et al.* [46]	Lexicon-based Approach	N-gram after, N-gram before and N-gram around.	N-gram around outperformed others in classification
10.	Hu and Liu [47]	Lexicon-based Approach	Apriori algorithm for association rule mining	Determines strengths and weaknesses of a product

11.	Popescu *et al.* [49]	Lexicon-based Approach	Extracts noun phrases	Sentiment classification
12.	Yang *et al.* [50]	Lexicon-based Approach	Semantic based feature extraction	Sentiment classification
13.	Zhang *et al.* [51]	Unsupervised Learning	Page-rank algorithm	Quality-based ranking of products
14.	Xu *et al.* [52]	Lexicon-based Approach	Chunk features, heuristic position, word features, part-of-speech features and context features	Better product opinion mining
15.	Yang and Shih [53]	Lexicon-based Approach	Class Association Rule Mining	Extraction of opinion-based reviews
16.	Khan *et al.* [54]	Lexicon-based Approach	Use of auxiliary verbs to extract opinion-based reviews	80% features have auxiliary verbs, hence good indicators of features and opinions
17.	Zhai *et al.* [55]	Lexicon-based Approach	Sentiment patterns matched with reviews	Feature extraction of products
18.	Chen and Tseng [56]	Lexicon-based Approach	Information quality framework	Elimination of noisy reviews
19.	Tang *et al.* [57]	Deep Learning	Deep memory network	Aspect-level sentiment classification

20.	Luo *et al.* [58]	Supervised Learning	Feature selection method of CDPNC	SVM performance improved for sentiment classification
21.	Socher *et al.* [59]	Deep Learning	Recursive neural tensor network	Improved accuracy, accurately captures effect of negation
22.	Mukku *et al.* [11]	Deep Learning	Vector of sentences using Doc2vec model fed as features	Better performance than classifiers using Supervised Learning
23.	Abdelwahab and Elmaghraby [60]	Deep Learning	Word2vec and Doc2vec model used to make vectors	Outperformed SVM, Logistic Regression and Boosted Tree
24.	Yu et al. [61]	Deep Learning	Contextual entropy model to measure similarity between words	More useful emotion words and their intensity, better accuracy for classifying data
25.	Hagenau *et al.* [62]	Lexical-based Approach	Selection of semantically relevant features	Overfitting reduced

2.3 Balancing the Data

In some scenarios, while using classifiers to conduct sentiment analysis or predict classes for classification problems, inaccurate results and unbiased predictions are produced. This is prevalent in cases where one class dominates the other class in data. Cross validation performed on imbalanced data produces highly inaccurate results as the classifiers are sensitive to the majority class. Marco Altini [67] explained several approaches to balance the data namely, Undersampling, Oversampling and Synthetic Techniques. In undersampling, some of the majority class observations are disregarded until the data is balanced. The problem faced in undersampling is that while eliminating some observations, there is a high probability of losing essential information, thereby degrading the performance of the classifier. In oversampling, overfitting can be an issue during cross validation. As it copies the observations of minority class and when cross validation is performed on the oversampled data, the one sample left for validation might be the replica of one or more observations in N-1 sample. Thus, actual accuracy of the classifier will not be measured. So, oversampling should be performed excluding the validation set for every iteration. It has been found that integrating both undersampling and oversampling has the capability of achieving better classifier performance as showcased by Chawla *et al*. [68]. One of the better approaches for data balancing is SMOTE. It does not replicate data directly from the dataset but employs synthetic techniques to generate data by considering feature-space and neighbors of the minority class. Synthetic examples of data are constructed corresponding to every minority class data along the line of k-nearest neighbors.

Chapter 3

PROBLEM STATEMENT

3.1 Barriers in Previous Work

With an ever increasing demand of smart phones, the mobile phone market is expanding at an exponential pace. With such a boom in the smart-phone industry, there is a need to realize the holistic review of the brand. Reviews available on e-commerce platforms act as a guiding tool for the consumers to make informed decisions. As there are innumerable products manufactured by many different brands, so providing relevant reviews to the consumers is the need of hour.

While an extensive amount of research work has been done related to the sentiment orientation of a review, however, most of the times classification of reviews is performed either into two classes: Negative and Positive. In binary classification, the sentiment of a review is determined simply on the basis of number of positive and negative words. One major drawback of classification of reviews into two classes using supervised learning is that when negation is present in a sentence it completely changes the essence of meaning and classifies incorrectly. For instance, words like 'good', 'nice', and 'correct' are labeled positive, however, presence of 'not' in a sentence can completely change the meaning of that sentence. To exemplify, the mobile phone not only worked great but also had amazing features. Here, the presence of 'not' does not make the sentence negative but intensifies the positivity of the sentence. Sometimes, the presence of 'not' completely changes the polarity of text, like, the product was not working. So, determining contextual polarity of the text accurately is one of the core issues of sentiment classification.

There are many instances when sarcasm is used to express the feeling of an individual. Use of sarcasm in a sentence is the use of words that mean the opposite of what a person is actually saying. Hence, capturing sarcasm, while determining the sentiment of some text, is one of the most tedious tasks in hand. Spam detection is also one of the emerging issues while determining helpful reviews as some sellers buy reviews and post them on their websites to lure more and more customers. At times, competitors also post negative reviews to degrade a product's performance in front of the users and drop the sales and increase the user's proclivity towards their product. Therefore, some kind of mechanism is required to filter the spam reviews from the database and display genuine reviews on the electronic front.

Linguistic features of a sentence are also considered while classifying the text and judging the performance of different features of a product. Determining key structures representing a sentence is essential to extract the meaning of a sentence. In most of the cases, only nouns or noun phrases have been used to determine different features of a mobile. But judging the performance of a product only on the basis of nouns is not accurate as other parts-of-speech like verbs, adjectives and adverbs also form an integral part of sentence structure.

3.2 Problem Statement

Providing information rich reviews act as a testimony to the consumers to know about the real-world performance of a product. Going through millions of reviews related to a product is not feasible for the users to evaluate the actual working of a product. So, filtering useful reviews and then classifying helpful reviews on the basis of their sentiment depicting different emotions can provide a clear picture about the product.

The primary aim of this thesis is to construct a framework which can extract helpful reviews from a large database of online reviews and perform sentiment classification of text using supervised and unsupervised learning. Lexicon-based approach is used to assign labels to the text. The classification of reviews is performed not only into two classes but also multiple classes to get a more detailed understanding of the review and improve the performance of the classifier models. The models are verified and their accuracy is validated. Lexical features of reviews are taken into consideration to extract dominant attributes of a mobile phone and evaluate their performance on the basis of user feedback. In this research, we intend to address the following questions:

- How to extract helpful online reviews for manufacturers?
- How to deal with imbalanced data?
- How to classify the text into different sentiments associated with the reviews?
- Which sentiment classification technique is most efficient for text classification?
- How to determine the quality of features of mobile phones?

The proposed work will help future buyers to make better decisions on the basis of analysis of feedback received by a particular smartphone brand. It will help to determine how well the feature is liked by the user and what are its drawbacks. It will also allow manufacturers to meet consumer expectations better on the basis of the feedback received and improvise the product.

3.3 Objectives of Proposed Work

The objectives of the research are:

- To extract meaningful reviews and perform sentiment classification of the text using Supervised Learning, Deep Learning and Feature-based Extraction.
- To cross validate and compare the performance of the techniques applied for sentiment classification.
- To determine the quality of different features of mobile phones using linguistic features of the text through rule-base.

Chapter 4

PROPOSED WORK

4.1 Framework

The main objective of the proposed framework is to perform sentiment classification of the text using different techniques and determine the performance of different features of mobile phones. The proposed framework of the research work is shown in Figure 4.1. The framework is divided into six different modules, viz., *Data Collection* and *Pre-Processing*, *Data Analysis* and *Selection*, *Sentiment Analysis*, *Supervised Learning*, *Unsupervised Learning* and *Feature-based Extraction*. The sentiment classification of reviews is done using three different approaches, namely, Lexicon-based Approach, Supervised Learning and Unsupervised Learning. Initially, the experimental data is collected from an e-commerce website Amazon.com. Data are pre-processed to remove stop words, punctuation marks, whitespaces, digits and special symbols. In the second step, statistical and exploratory analysis is conducted. Helpful reviews are extracted from the dataset. Finally, feature selection is performed to extract relevant features from the data set. In the given data set out of the six features, only three features, i.e., *Product Name*, *Brand Name* and *Reviews* have been considered for sentiment classification. In the third step, lexicon-based approach is employed to classify the data. *'Pos/Neg'* tags are appended to the dataset corresponding to each review to conduct supervised learning. In the fourth step, training and testing of the classified data is conducted using Naïve Bayes, SVM and Decision Tree models using supervised learning. The accuracy so obtained is validated using 10-fold cross validation. The fifth step involves sentiment classification of reviews using unsupervised learning in which two techniques are used: Vocabulary-based Vectorization and Feature Hashing. The last and final step includes feature-based extraction of reviews to determine the properties of different features of mobile phones.

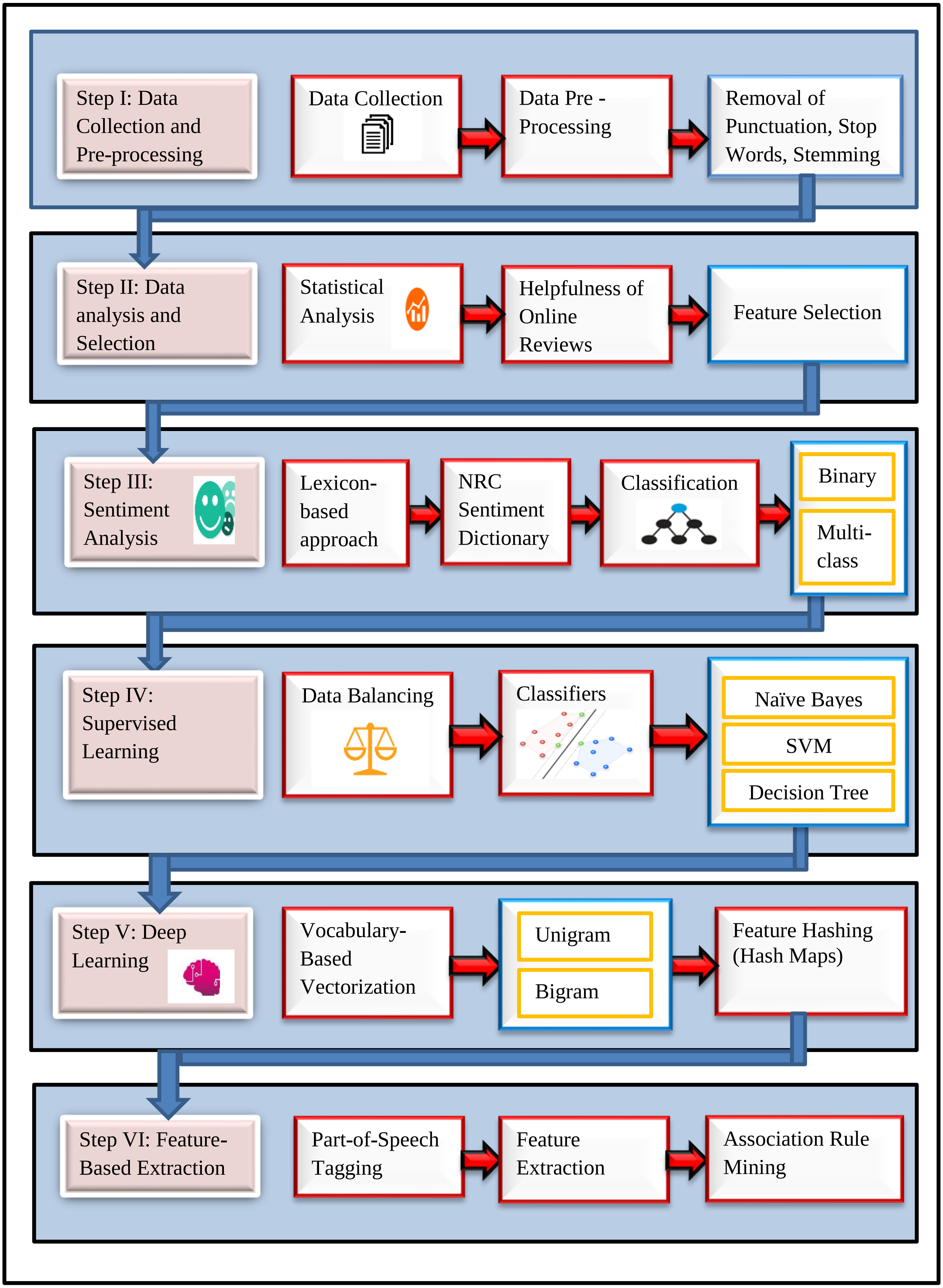


Figure 4.1: Proposed Framework

4.2 Data Annotation

Once the delineated analysis of the data is completed, the reviews of top three brands Samsung, Apply and BLU are selected to further filter out non-helpful reviews. So, helpfulness of the reviews is determined on the basis of some fixed criteria which are defined below.

- Number of Review Votes>0
- Number of Question Marks should not be greater than 3
- Number of Exclamation Marks should not be greater than 3
- Number of words in all caps not greater than 3
- Length of the review >= 2
- Character count>5

The reviews not fulfilling the above mentioned criteria will not be considered for further research and will be filtered out. After the extraction of helpful reviews, dataset is extracted with relevant features using filter method. Out of the six attributes as explained in Table 4.1, only three attributes are chosen, namely, Product Name, Brand Name and Reviews.

Data cleansing is an integral part of analysis as raw data is first converted into information suitable for conducting analysis. Raw data is the data directly collected from a particular source. It is first converted into a data frame which is the readable form in R. This readable form is known CSV format. It contains unprocessed data with missing values, incorrect labels, and outliers and so on. So, reviews are pre-processed in which blank lines, punctuation marks, numbers and stop words are removed. There are about 127 stop words in English language like (and, is, are, just). Stemming of reviews is then performed to make words having same stem but different forms to be considered the same. Words like ‘surprise’, ‘surprising’ and ‘surprisingly’ are reduced to the stem ‘surprise’. After pre-processing, the data are converted in the form of matrix to perform classification.

4.3 Sentiment Classification of Reviews

The different approaches applied in sentiment analysis are: *Lexicon-based Approach*, *Supervised Learning* and *Deep Learning*. The steps followed in each of the technique are explained below.

4.3.1 Lexicon-based Approach

Sentiment classification of unlabeled reviews is conducted using three sentiment dictionaries. NRC sentiment dictionary is used to extract eight different emotions and their corresponding valence in the text including all the reviews. Eight different emotions represented are: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. Positive and negative valence is also calculated. The steps followed to conduct sentiment analysis and attach labels to the data are:

Step 1: Convert text into string

The data uploaded is firstly converted into string by implementing get_text_string() function to divide the sentences into tokens.

Step 2: Tokenization

After converting the text into string, every review present in the text is broken into tokens of words by implementing get_tokens() function which is then collected into a vector.

Step 3: Extracting sentiments

The token of words of each review is then passed into the get_sentiment() function in which 'nrc' method is called to extract sentiments. After successful classification of all the reviews, sentiment extraction of reviews of top three brands, i.e., Samsung, BLU and Apple is performed using the above stated steps.

Step 4: Classification

Classification of the reviews is performed on the basis of overall valence of the sentence. Classification is done into two types: Binary classification and Multiclass classification.

- **Binary Classification**

  Binary classification involves classification of data into two classes. The sentiment score established for each positive and negative polarity using National Research Council Canada (NRC) sentiment dictionary is then added to the reviews dataset. The dictionary has words representing eight different emotions viz. anger, anticipation, disgust, joy, sadness, surprise and trust and two sentiments namely, positive and negative. This individual score is used to calculate the overall polarity as given by Eq. (4. 1) of the sentiment.

  $$Polarity = Positive\ Score - Negative\ Score \tag{4.1}$$

  After calculating the polarity corresponding to each review, different subsets of positive and negative sentiment are formed having polarity 0 to 10 and -1 to -10 respectively. The sample of polarity of reviews is shown in Table 4.1. The subsets so obtained are combined together in a CSV file along with the Pos/Neg tags as depicted in Table 4.2.

Table 4.1: Sample of Overall Polarity of Reviews

Review	**Positive**	**Negative**	**Polarity**
Very Pleased	1	0	1
Phone was in great condition and was able to be activated when received.	1	0	1
The Phone does not work it's crap! It slowly got worse now I'm to the point where I cannot hear my calls!	0	3	-3
I bought this for my mother to replace the iPhone 4 she had been toting around for years. This review is a reflection of how she feels about the phone. She does love the bigger screen and the fact that it moves faster overall- but the battery does die way quicker than the 4's did. Even after I did lots of energy-saving settings modifications, it still dies quickly.	3	3	0

It came without 2gb sd card :S shipped in 1 day, phone is good, love keyboard but battery life is too short , has no internal memory.	2	1	1
there was no 2 GB microSD card !! also no headphones, but the phone is perfect. i'm giving 4 star for this cheat :)	3	1	2

Table 4.2: Sample of Binary Classification of Data

Class (Pos/Neg)	**Reviews**
Pos	Very pleased
Pos	Phone was in great condition and was able to be activated when received.
Neg	The Phone does not work it's crap! It slowly got worse now I'm to the point where I cannot hear my calls!
Pos	I bought this for my mother to replace the iPhone 4 she had been toting around for years. This review is a reflection of how she feels about the phone. She does love the bigger screen and the fact that it moves faster overall- but the battery does die way quicker than the 4's did. Even after I did lots of energy-saving settings modifications, it still dies quickly.
Pos	It came without 2gb sd card :S shipped in 1 day, phone is good, love keyboard but battery life is too short , has no internal memory.
Pos	there was no 2 GB microSD card !! also no headphones, but the phone is perfect. i'm giving 4 star for this cheat :)

- **Multi-label Classification**

  Multiclass classification means classification of data into three or more classes. Multiclass classification is done into five different classes namely, ‘Positive’, ‘Very Positive’, ‘Neutral’, ‘Negative’ and ‘Very Negative’. The classes are determined on the basis of the polarity of the review calculated. The range of polarity considered for each class is determined in the Table 4.3.

Table 4.3: Distribution of Classes over Polarity

Class	**Polarity Range**
Positive	1 to 5
Very Positive	6 to10
Neutral	0
Negative	-1 to -5
Very Negative	-6 to -10

Table 4.4 elucidates the classified data in five different classes where ‘Pos’ stands for ‘Positive’, ‘Very Pos’ stands for ‘Very Positive’, ‘Neg’ stands for ‘Negative’, ‘Very Neg’ stands for ‘Very Negative’.

Table 4.4: Sample of Multi-label Data

Reviews	**Class**
Very pleased	Pos
Phone was in great condition and was able to be activated when received.	Very Pos
The Phone does not work it's crap! It slowly got worse now I'm to the point where I cannot hear my calls!	Neg
This phone was a huge disappointment and a fraud. I can't wait until I have the cash for a new smartphone. So many headaches. First, it is NOT	Very Neg

LTE capable. Total lie and I've read many other reviews that say the same .Horrible battery life. Locks up regularly Slow processor Totally overpriced for such a hunk of junk.	
I bought this for my mother to replace the iPhone 4 she had been toting around for years. This review is a reflection of how she feels about the phone. She does love the bigger screen and the fact that it moves faster overall- but the battery does die way quicker than the 4's did. Even after I did lots of energy-saving settings modifications, it still dies quickly.	Neutral
love, love this phone...quite fun and easy to use..going on my third month of usage and so far, no problems at all, and using the phone at the same time...but its been well worth..the $$$ i paid for it...ooo.have i mentioed that the frosted gold is just divine??...it is lovely..	Very Pos
there was no 2 GB microSD card !! also no headphones, but the phone is perfect. i'm giving 4 star for this cheat :)	Pos

4.3.2 Data Balancing

Class distribution is checked after classification of the data. If this class distribution is not found to be equal or almost equal, then data balancing is done. This implies that the data are imbalanced as the target variable has imbalanced proportion of classes. In this research, the four techniques of data balancing are applied: Undersampling, Oversampling, both undersampling and oversampling and SMOTE. Their accuracies are compared and the technique with best accuracy is chosen.

4.3.3 Classifiers Used

A classifier is a function *f* devised to map input features to output class labels. There are numerous classifiers available in machine learning. In this research, three classifiers have been used to perform sentiment classification of data. The classifiers used are: Naïve Bayes, Support Vector Machine and Decision Tree.

- **Naive Bayesian Classifier:** A statistical classifier or a probabilistic model that maps input feature vectors to output class labels. For a set of training data D, each row is represented by an n-dimensional feature vector, $X = x_1, x_2, ..., x_n$. There are K classes, $K_1, K_2, ..., K_m$ in the output class label. For every tuple X, the classifier will predict 2 as given by Eq.(4.2) that X belongs to K_i if and only if: $P(K_i/X) > P(K_j/X)$, where $i, j \varepsilon [1,m]$ and $i \neq j$.

$$P(K_i \,|X) = \prod_{a=1}^{n} P(x_k|K_i) \tag{4.2}$$

  Naïve Bayes classifier works on the assumption that the features are not dependent on each other given the class label, hence termed as Naïve. These classifiers can be trained efficiently in the case of supervised learning. The main advantage of Naïve Bayes classifier is that it does not require very large amount of training data to learn and can estimate the parameters required for classifying the data with good efficiency.

  As mentioned in Murphy [69], bag of words is a common technique employed for document classification in Naïve Bayes, where occurrence of each word is noted. If the occurrence of a particular word is above the threshold value then the document is classified accordingly. For instance, in sentiment classification of data into positive and negative, number of positive and negative words will be counted in a particular sentence. If the number of positive words is more than the number of negative words, then the sentence will be classified as positive and vice-versa. Therefore, this naïve method does not take order of words into consideration.

- **Support Vector Machine (SVM):** SVM is used for a labelled training data that categorizes testing dataset using an optimal hyperplane. A hyperplane is separates data of one class from another which is defined as given in Eq. (4.3).

$$W - X + b = 0 \tag{4.3}$$

Figure 4.2 shows the hyperplane between two classes. SVM chooses the hyperplane which segregates the classes better. The hyperplane having maximum distance from closest data point is known to be the best segregator.

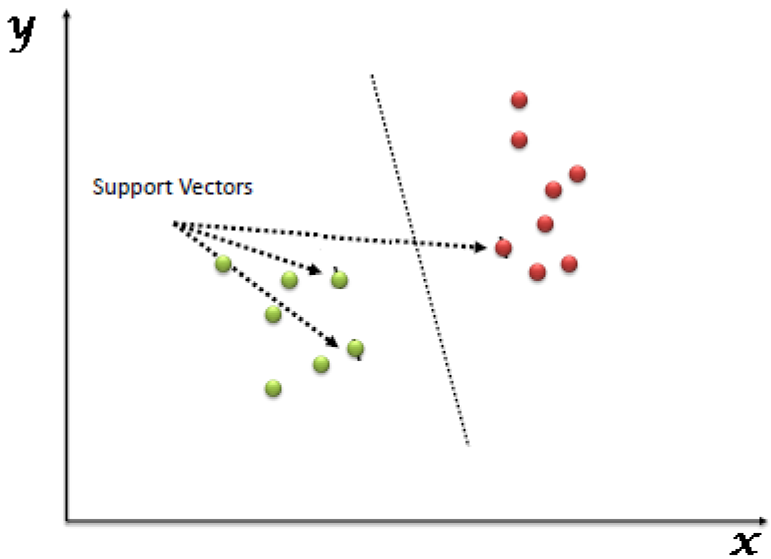


Figure 4.2: Segregation of classes using hyperplane in SVM

SVM is mainly used for supervised learning. It can perform both linear and non-linear classification using kernel trick, i.e., it works on an implicit feature space to transform data without any need of human intervention.

- **Decision Tree:** A hierarchical tree structure encompassing decision nodes for representing attributes and edges for denoting attribute values. This representation in the form of a tree allows to construct decision rules that classify new instances of the data. It has decision nodes and leaf nodes. Decision nodes have two or more branches and leaf nodes depict the classification. Top-down greedy approach is generally applied for building decision trees where subsets of a large dataset are created. This subdivision of data is performed with the help of some rules. Edges of the tree represent the rules. Figure 4.3 shows the diagrammatical representation of decision tree.

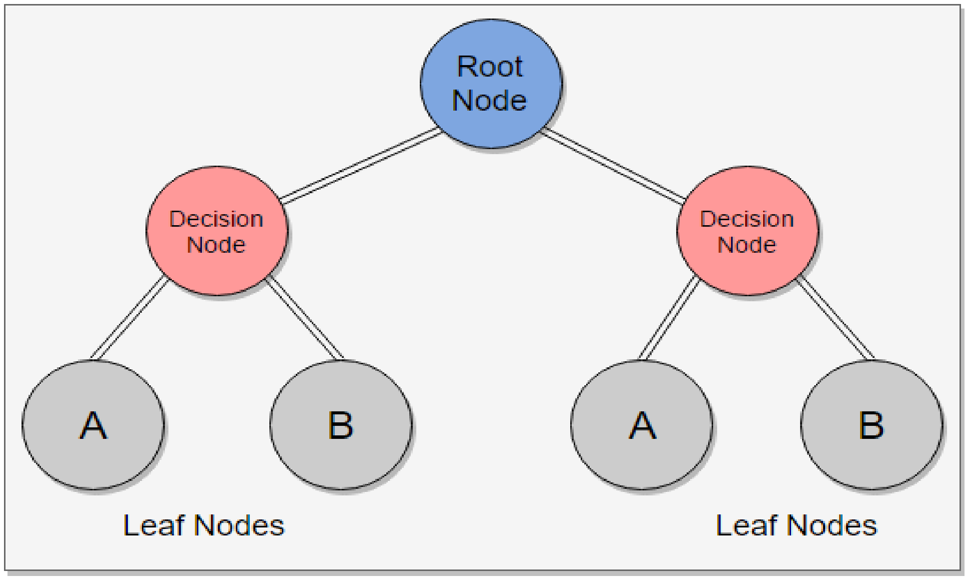


Figure 4.3: Decision Tree

Steps followed in classification of text using machine learning models are shown in Figure 4.4.

Cross Validation

When the model is trained and tested, 10-fold cross validation is performed to determine how well the learner has learnt and how well can it make accurate predictions.

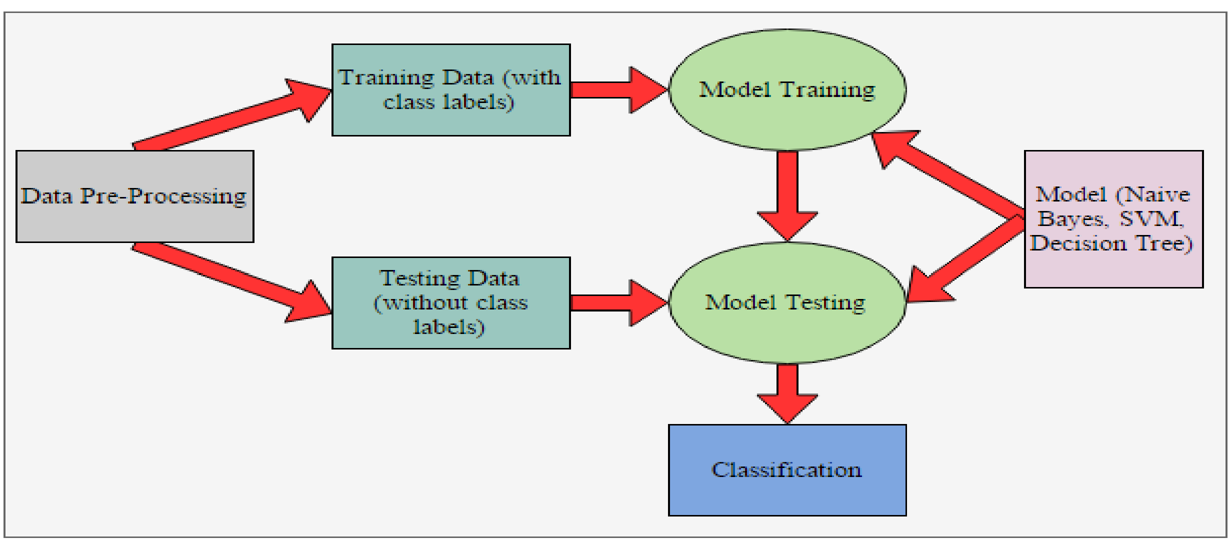


Figure 4.4: Classification Process

4.4 Sentiment Classification using Deep Learning

Deep learning takes contextual polarity of the sentence into consideration. It implies that a phrase having the word not before an adjective, like 'not good', will be predicted as negative. The dataset used to train the model has two classes: 0 and 4. 0 stands for a negative review and 4 stands for a positive review. Values lying between 0.35 and 0.65 will be considered neutral.

Text2vec package has been used to conduct sentiment analysis. The text is converted into vectors of n-grams by using *vocabulary* and *feature hashing*. *Global Vectors for Word Representation (GLoVe)* algorithm is used to extract vector form for words. A vector converts every word into some numerical value which can then be fed as features to train a model. It determines the similarity between two words on the basis of their frequency of occurrence together in a corpus. GLoVe learns those word vectors whose dot product will be equal to the logarithm of the probability of words occurring together. Nearest neighbors reveal not only the common words associated with a particular vector but also the rare relevant words that are not so easy to guess by human mind. For example, the closest words obtained for the word 'happy' are: glad, surprise, great, satisfied, love, excellent, perfect, good, etc. Model 'glmnet', as explained in Friedman *et al.* [70], is used to fit the lasso model in which Gaussian Linear model has been employed to train the data. 'glmnet' is very fast and has the ability to exploit sparsity in the input matrix.

- **Vocabulary based Vectorization**

The following steps are undertaken for conducting sentiment analysis using vocabulary.

Step 1: Vocabulary based Vectorization

A set of unique terms are extracted from all the reviews which are then mapped with a unique id. Unigrams and bigrams are used to create this vocabulary. A Document Term Matrix (DTM) or sparse matrix is created of the unique vectors which are traversed using an iterator. The number of reviews is represented by rows and the number of unique terms is represented by columns.

Step 2: Model Training

DTM obtained from the vocabulary based vectorization is used to train the model. A linear model is used to perform training and testing. Since, it is binary classification so ‘binomial’ family of ‘glmnet’ model has been used.

Step 3: Cross Validation

5-fold Cross validation has been performed to determine the accuracy of trained model. In 5-fold cross validation, the dataset is divided into five equal size subsamples out of which four samples are used to train the data and the single subsample left is used to test the data.

Step 4: Model Prediction

Steps 1 to 3 are repeated for using this model to make predictions on unclassified data.

- **Feature Hashing**

It is the process of creating hash maps to represent different features of the data. It is used to compress sparse matrices in which no dictionary is required.

Steps followed to perform sentiment analysis using feature hashing are:

Step 1: Sentence Tokenization

Sentences are broken into token of words and are traversed using an iterator.

Step 2: Hash Vectorization

A vector of hash keys is created for each word which is then used as features.

Step 3: Model Training

A linear model of binomial family is used to learn from the features and ‘glmnet’ package is used to fit the model.

Step 4: Cross Validation

5-fold cross validation is performed to check the accuracy of the trained model.

Step 5: Model Testing

Steps 1 to 4 are repeated to test the model on unseen data.

4.5 Feature-Based Extraction

Reviews are extracted on the basis of some pre-defined features which are used to make rules. The rules so obtained are then used to determine the performance of various features of a mobile phone. The two modules of feature based extraction are: *Part-of-Speech Tagging* and *Association Rule Mining*.

4.5.1 Part-of-Speech (POS) Tagging

POS tagging is a part of data annotation in which each lexical item is classified according to the part of speech it belongs to. Every word in a sentence can be classified into noun, verb, adjective, adverb and determiner, etc. For example,

Ann drives a car.

After POS tagging the sentence will be classified and will be returned as:

Ann\NNP drives\VBZ a\DT car\NN.

Here, NNP denotes proper noun (singular), VBZ denotes third person singular present tense verb, DT denotes determiner and NN denotes noun.

The various part-of-speech tags available and their category have been depicted in Table 4.5. Two different ways exist to add POS tags to the data: Using *openNLP package* and *Tree Tagger*.

- **OpenNLP Package:** The library implements the Apache OpenNLP Maxent Part-of-Speech Tagger which comes with pre-trained models. So, one does not need to explicitly train the POS taggers as generally required. It is usually used for newspaper texts and POS tags are assigned based on the highest probability of possible POS tags associated with a word.
- **TreeTagger:** Another method of POS tagging is using the koRpus library. This package makes use of TreeTagger, a third-party software to break the text into tokens which are then tagged according to their part-of-speech.

Installing TreeTagger is usually complex in windows, so windows users generally use OpenNLP package for POS tagging. Table 4.5 elucidates the different part-of-speech tags produced using this methodology.

Table 4.5: POS Tags

Part-of-Speech Tag	**Part-of-Speech Category**
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol

TO	To
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBZ	Verb, 3rd person singular present
WDT	Wh-Determiner
WP	Wh-Pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

4.5.2 Association Rule Mining

Association rule mining is used to extract rules or associations between variables from market basket data having unique transaction IDs. The rule depicts that the items on the right hand side will be present if particular items on the right hand side are present. The rule is of the form as shown in Eq. (4.4):

$$A \Rightarrow B \tag{4.4}$$

They are represented in the form of if-then statements where 'if' is the antecedent part of the rule and 'then' forms the consequent. For example, *rechargeable* ⇒ *battery* means that whenever rechargeable word occurs in the transaction, battery will be present along with it. Such rules can help to make feature-based extraction of reviews, thus providing users with the desired information.

- **Apriori Algorithm**

Apriori algorithm has been used to generate rules for *frequent item sets*. Frequent item sets are sets having items with minimum support. Item sets not fulfilling the minimum support criteria are filtered out. Frequent item sets are denoted by L_i for i^{th}

item set. In order to find L_i, set of candidate I item sets is formed by joining L_{k-1} with itself.

Bottom-up approach is implemented in this algorithm in which frequent item sets are extended one at a time. It works on a set of transactions. Apriori property states that any subset of frequent item set must also be frequent. It means that if {XY} is a frequent item set, then both {X} and {Y} should also be a frequent item set. Item sets can be 1-frequent item set, 2-frequent item set, 3-frequent item set and so on. Steps followed to perform feature-based extraction are given below.

Step 1: POS Tagging

In this research work, openNLP package has been used to tag along parts-of-speech of the words present in each review. After POS tagging, 4-gram phrases having only meaningful words like nouns, verbs, adverbs and adjectives are extracted from each sentence.

Step 2: Feature Extraction

Relevant features are extracted from the list of nouns obtained in Step 1.

Step 3: Generation of Rules

Rules are generated using Apriori algorithm where threshold values of support and confidence are fixed.

4.6 Evaluation Parameters

The actual and predicted classification done by the system is checked using the confusion matrix. The accuracy is determined based on the percentage of observations correctly classified. The confusion matrix has the following components:

- **True Negative (TN):** Number of reviews which are 'negative' and predicted as 'negative'.
- **True Positive (TP):** Number of reviews which are 'positive' and predicted as 'positive'.

- **False Negative (FN):** Number of reviews predicted as 'negative' but are not 'negative'.
- **False Positive (FP):** Number of reviews predicted as 'positive' but are not 'positive'.

  Table 4.6 shows the representation of confusion matrix with the outcomes explained above.

Table 4.6: Confusion Matrix

		Predicted Class	
		Negative	Positive
Actual class	Negative	TN	FP
	Positive	FN	TP

Evaluation parameters derived from the confusion matrix are:

- **Accuracy:** It is the ratio of total number of reviews correctly classified to the total number of reviews classified. It is given by the Eq. (4.5).

$$Accuracy = \frac{TP+TN}{TN+FP+FN+TP} \tag{4.5}$$

- **True Positive Rate (TPR):** It is also known as *sensitivity or recall*. It is the rate at which positive class is predicted correctly as shown in Eq. (4.6).

$$TPR = \frac{TP}{TP+FN} \tag{4.6}$$

- **True Negative Rate (TNR):** It is also known as *specificity*. It is the rate at which negative class is classified accurately and is given by Eq.(4.7).

$$TNR = \frac{TN}{TN+FP} \tag{4.7}$$

- **False Positive Rate (FPR):** It is the ratio of positive class misclassified as negative as shown in Eq. (4.8).

$$FPR = \frac{FP}{FP+TN} \tag{4.8}$$

- **False Negative Rate (FNR):** It is the proportion at which number of negative class was predicted as positive and is given by Eq. (4.9).

$$FNR = \frac{FN}{FN+TP} \tag{4.9}$$

- **Precision:** It is defined as the measure of correctness achieved in positive prediction and is given by Eq. (4.10).

$$Precision = \frac{TP}{TP+FP} \tag{4.10}$$

- **Receiver Operating Characteristics (ROC) curve**: It is a graph used to summarize the performance of a classifier. The plot is drawn between TPR on y-axis and FPR on x-axis.

Association rules represent correlation between item sets. Association rule evaluation metrics are given below:

- **Support:** Support is the number of transactions having all items of the antecedent and consequent part of the rule and is given by Eq. (4.11).

$$support(A \Rightarrow B) = P(AUB) \tag{4.11}$$

- **Confidence:** Confidence is the ratio of the number of transactions having all items of the antecedent and the consequent to the number of transactions having items in the antecedent as shown by Eq. (4.12).

$$confidence(A \Rightarrow B) = P(B|A) = \frac{P(AUB)}{P(A)} \tag{4.12}$$

The evaluation parameters listed are used to evaluate the performance of the different techniques used for sentiment classification.

Chapter 5

SIMULATION RESULTS

The proposed approach is applied and tested on testing dataset to classify reviews accurately and perform feature based extraction.

5.1 Dataset and its features

A large sample of online reviews is collected from the e-commerce giant Amazon.com. The data set consists of 41,3840 reviews for approximately 4500 mobile phones. It includes six features as explained in Table 5.1. The dataset consists of 385 different brands with 4410 unique products.

Table 5.1: Features included in the Dataset

Feature	Description
Product Name	Model name of mobile phone
Brand Name	Manufacturing brand
Price	Price of the mobile in dollars
Rating	User rating between 1 to 5
Reviews	User reviews provided for every mobile phone
Review Votes	Number of people who found the review helpful

5.2 Statistical Analysis

Different features present in the dataset are compared and plotted to understand the relationship between them. The different outcomes of running the statistical analysis are listed below.

- **Number of Review Counts by Brand**

  After analysis, the count of number of reviews is determined for each brand. Table 5.2 represents the count of top 10 reviews in descending order starting from Samsung (highest). Through the bar chart it can be seen that Samsung, BLU and Apple are the top three brands having the highest number of reviews. Brands HTC and CNPGD have received the minimum reviews from the consumers. Thus, one can conclude that the brands Samsung, BLU and Apple have the highest customer base among all the other brands.

Table 5.2: Top 10 Brands with highest number of reviews

Brand Name	**Review Count**
Samsung	65747
BLU	63248
Apple	58186
LG	22417
Blackberry	16872
Nokia	16806
Motorola	13417
HTC	12724
CNPGD	12613
Otter Box	7989

- **Rating Distribution by Brand**

  Rating distribution corresponding to each brand is found out to determine the brands receiving highest and lowest rating. Figure 3 provides the rating distribution by each brand. User rating for each review is provided on a scale of 1 to 5 where 1 to 2 depicts negative rating, 3 is neutral and 4 to 5 depicts positive rating. The top three brands having highest rating 5 are Samsung, Apple and BLU. The negative ratings of these three brands are almost the same. It can be concluded that Samsung is the most favored brand in terms of rating. CNPGD brand has almost equal rating of 1, 3 and 5 respectively showcasing that CNPGD brand is an average brand. OtterBox brand has the lowest rating of 1.

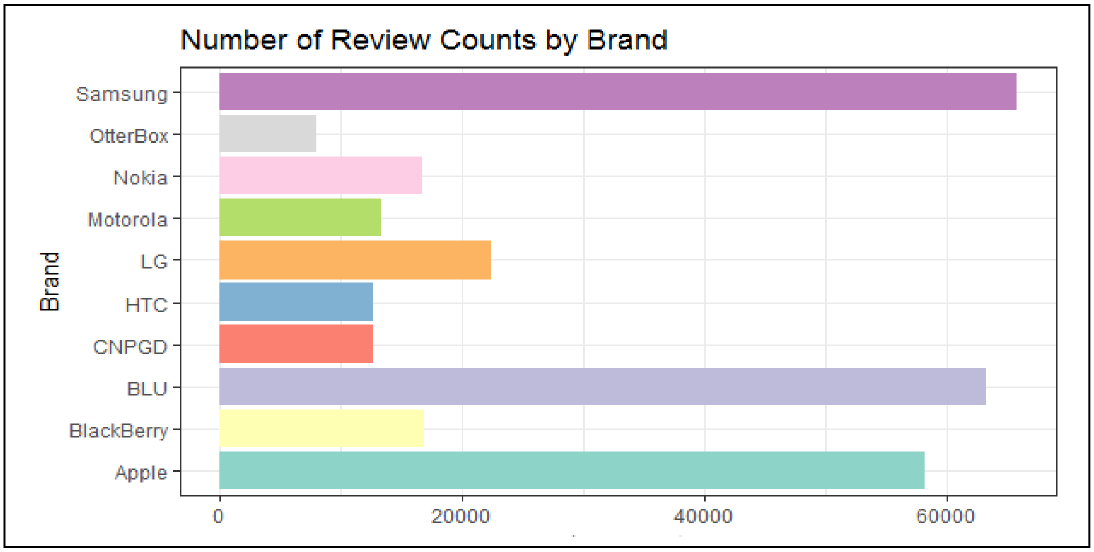


Figure 5.1: Number of Review Counts by Brand

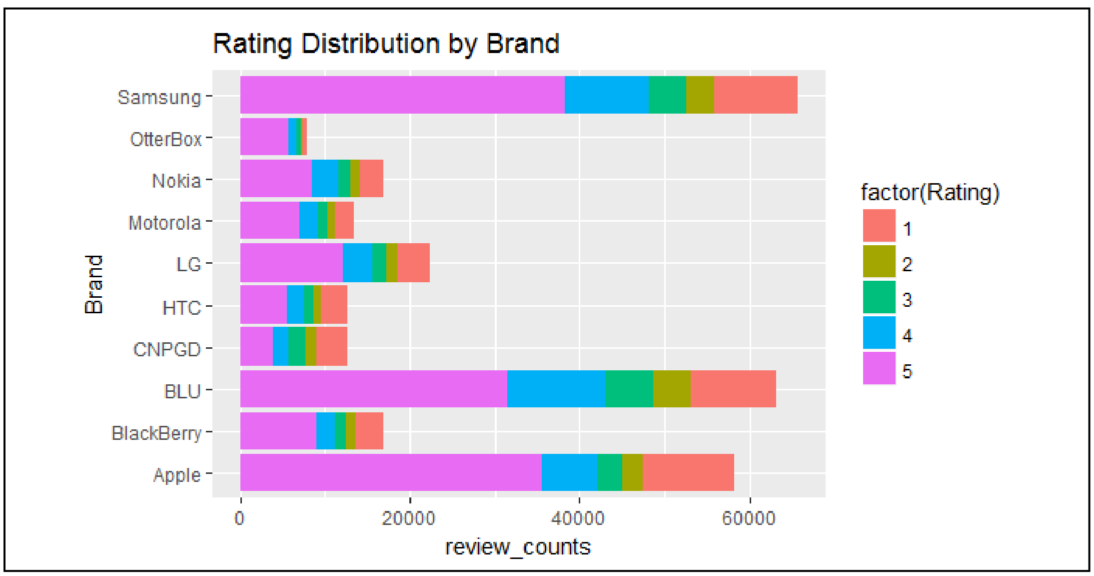


Figure 5.2: Rating Distribution by Brand

- **Average Rating for Top 10 Brands**

  The average rating of top 10 brands is calculated. OtterBox has the highest rating followed by Samsung and Apple. CNPGD has the lowest rating among these 10 brands. Table 5.3 shows the average rating of the brands and Figure 5.3 depicts the graphical representation of the same.

Table 5.3: Average Rating for Top 10 Brands

Brand Name	Mean Rating
OtterBox	4.383778
Samsung	3.962356
Apple	3.924415
LG	3.841460
Nokia	3.819291
Motorola	3.812849
BLU	3.792262
Blackberry	3.741465
HTC	3.465420
CNPGD	3.106002

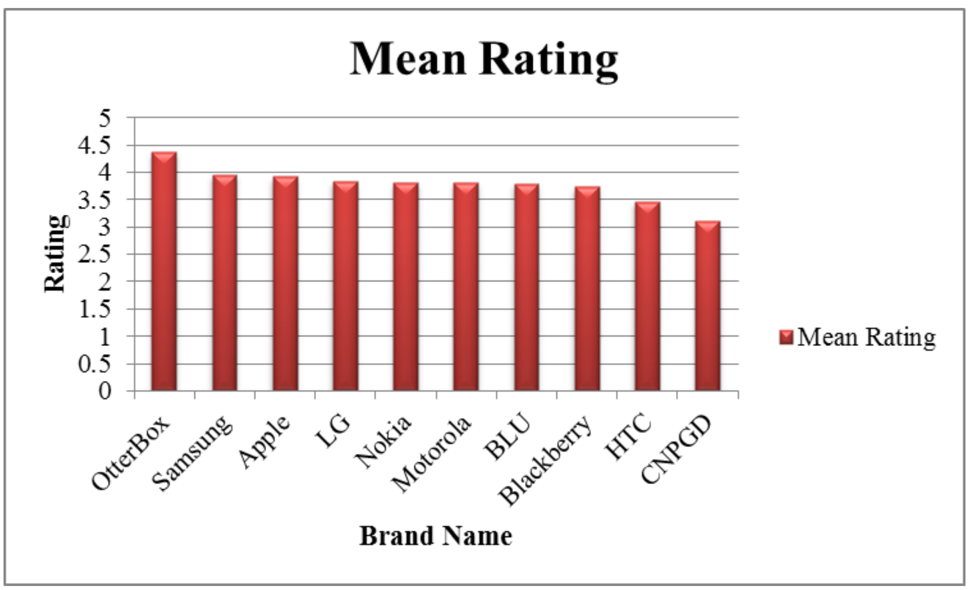


Figure 5.3: Average Rating of Top 10 Brands

- **Positive and Negative Reviews Distribution by Brand**

  Classification of reviews is done in terms of positive and negative reviews with respect to each brand. Reviews having rating between 0 and 3 have been classified as negative and reviews having rating between 3 and 6 have been classified as positive. This analysis will influence the consumer buying patterns, as the consumer tends to choose the brand having maximum positive reviews. In Figure 4, it can be observed that the brand Samsung has the highest share of positive reviews among all the brands. Brands BLU and Apple stand second and third in the stack. CNPGD brand has almost equal share of positive and negative reviews. Although, OtterBox brand has minimum number of reviews, yet its positive feedback is far greater than the negative feedback. This elucidates that even though mobile phones manufactured by OtterBox haven't been used by a large number of users, yet the consumers' experience has been fairly good with its handsets.

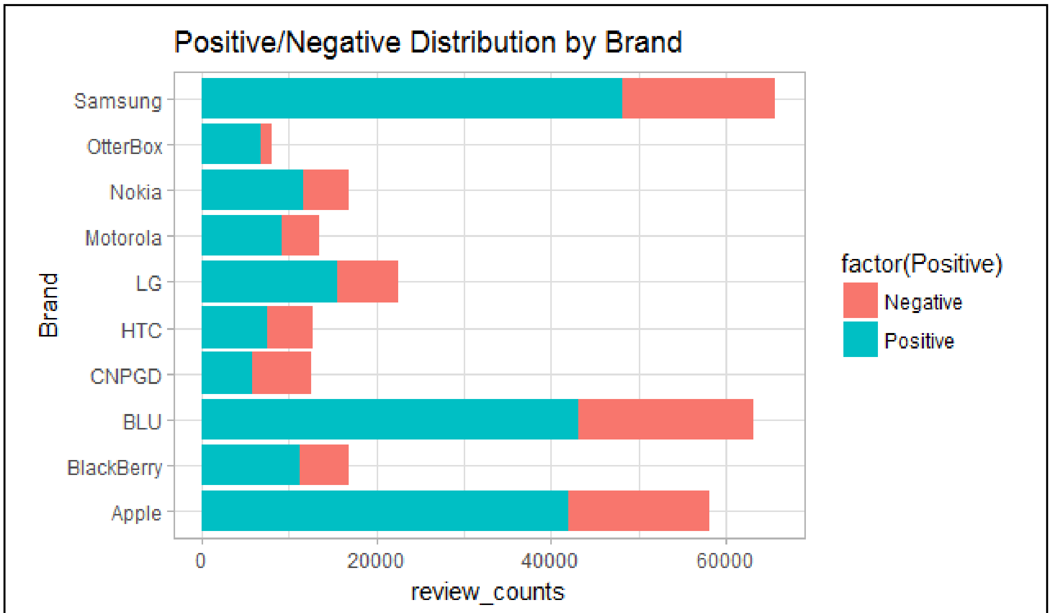


Figure 5.4: Positive/Negative Reviews Distribution by Brand

- **Review Length**

  Review length is calculated to determine its relationship with other attributes. In this data set, there are over 400,000 reviews. Length of each review is calculated. Maximum length of a review is 29,624 characters. Mean length of all the reviews comes out to be 216.67 characters. The analysis of relationship between review length and dataset features like product rating and product price is presented below.

➢ *Review Length and Product Rating*

Relationship between review length and product rating enabled us to determine whether detailed reviews affect product rating or not. Figure 5.5 elucidates that there is no correlation between the two.

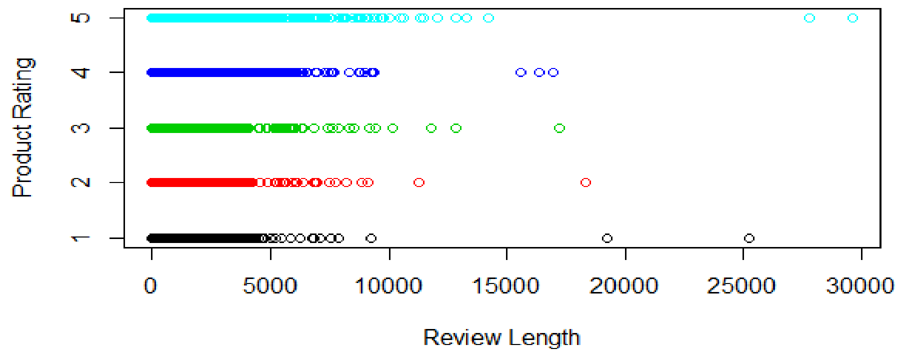


Figure 5.5: Relationship between Review Length and Product Rating

➢ *Review Length and Product Price*

The Figure 5.6 shows that with the increase in price, the length of the reviews does not increase. As the top models of different brands have high price, so one expects to have detailed reviews regarding these products. However, no such trend was observed.

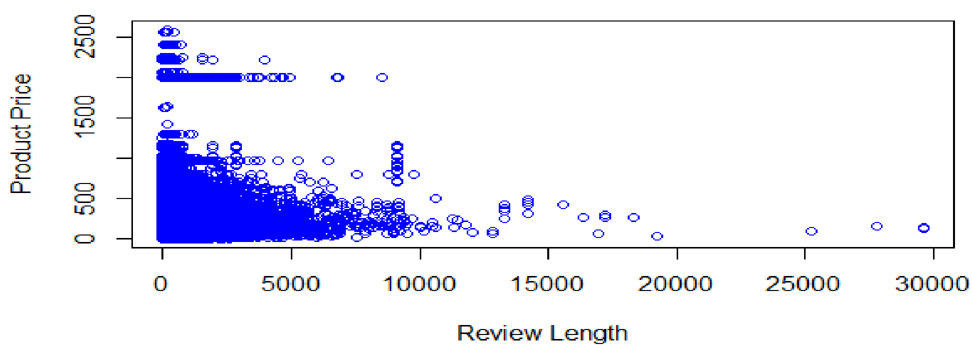


Figure 5.6: Relationship between Review Length and Product Price

- **Price and Rating**

Higher the price of the product, higher the expectations of the consumers and better is the quality. As seen in Figure 5.7, high priced products attract higher ratings. This illustrates that there is higher satisfaction among buyers of expensive products.

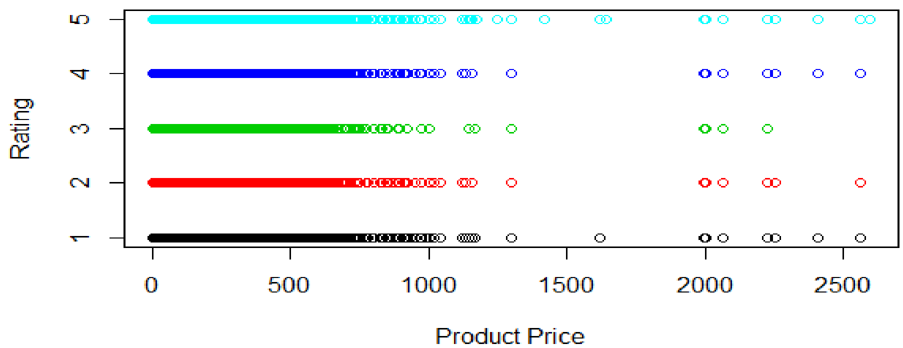


Figure 5.7: Relationship between Product Price and Rating

- **Word Cloud**

The most frequent occurring words are found out in this analysis which can give both the consumer and the designer an idea of what the users are feeling about the product or what are the key aspects of the product. The height of the word represents its frequency. The word cloud of words having minimum frequency of 2000 is created by using 'SnowballC' [11] and 'WordCloud' [12] packages for each of the top three brands namely Samsung, BLU and Apple as shown below.

In Figure 5.8, highlighted words like 'great', 'good', 'quality', 'happy' and 'love', etc. Most of the words present in the cloud are positive in nature, hence giving a subtle idea to the consumers that the performance of mobile phones manufactured by Samsung is good. Similarly, in Figure 5.9, highlighted words like 'great', 'good', 'excellent', 'easy' paint a positive picture before the consumer about the BLU brand. Also, in Figure 5.10, highlighted words like 'great', 'good', 'works', 'perfect', 'recommend' are assuring the consumer that the products of apple are worth-buying.

Figure 5.8: Word Cloud for Samsung Reviews

Figure 5.9: Word Cloud for BLU Reviews

Figure 5.10: Word Cloud for Apple Reviews

5.3 Results

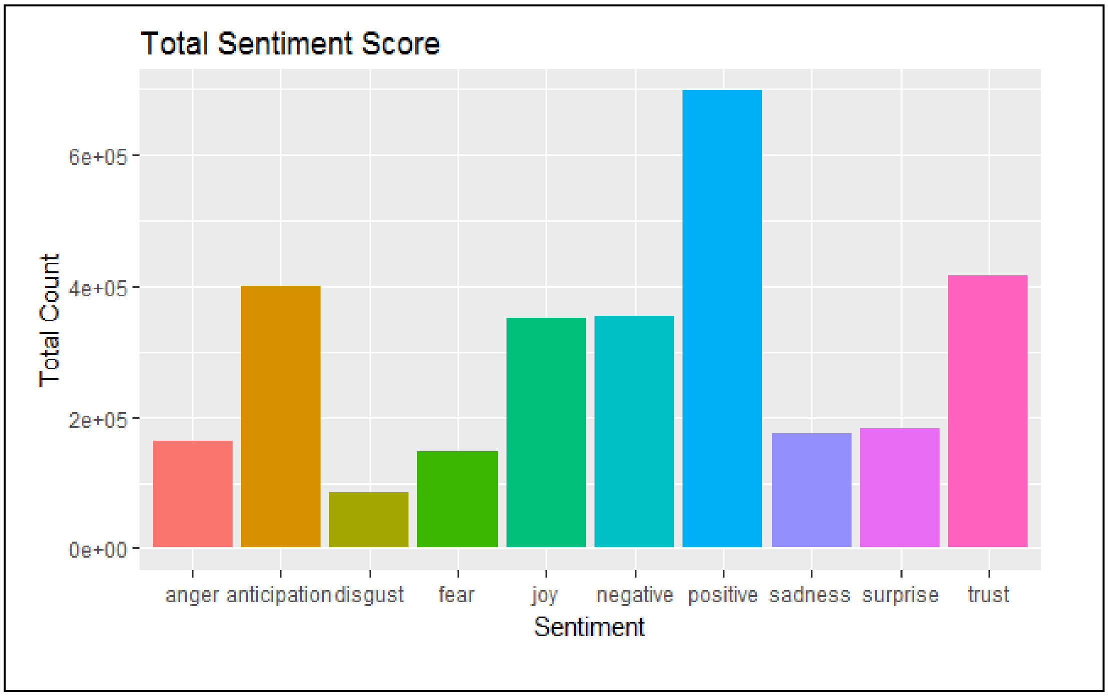
Sentiment classification of text and feature-based extraction of reviews is performed as explained in chapter 4. The results obtained are shown below.

5.3.1 Sentiment Analysis using Lexicon-Based Approach

Table 5.4 demonstrates the mean values of emotions extracted from the text. The sentiment ‘Positive’ has the highest mean value and ‘Disgust’ has the lowest mean value. The mean values determine the distribution of sentiments throughout the text. The graphical representation of the distribution of sentiments is shown in Figure. In Figure 5.11 it is seen that overall polarity of the text is positive with high valence of trust, joy and anticipation. Negative polarity is almost half the positive polarity as shown in Figure 5.12 with average feelings of anger, disgust, fear and sadness.

Table 5.4: Summary of Sentiment Values

Sentiment	Minimum Value	Maximum Value	Mean
Anger	0	24	0.3941
Anticipation	0	39	0.9626
Disgust	0	17	0.2075
Fear	0	30	0.3583
Joy	0	32	0.8501
Sadness	0	36	0.4212
Surprise	0	23	0.4428
Trust	0	47	1.003
Negative	0	64	0.8527
Positive	0	86	1.683



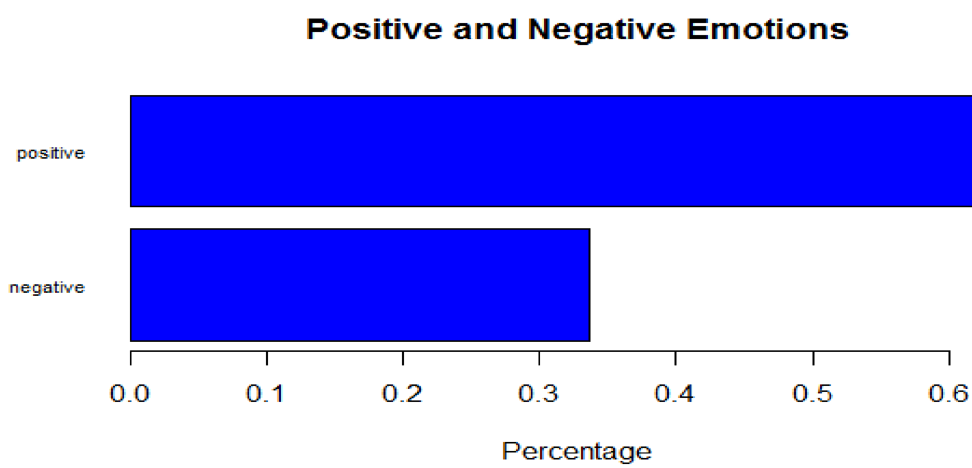


Figure 5.12: Percentage of Positive and Negative Emotions

- **Sentiment Analysis of Samsung Reviews**

According to Figure 5.13, feelings of anticipation, joy, trust and surprise are highly greater than the feelings of disgust, fear and sadness. As per Figure 5.14, number of the positive reviews is more than double the number of negative reviews. Thus, the overall sentiment of mobile phones manufactured by Samsung is positive.

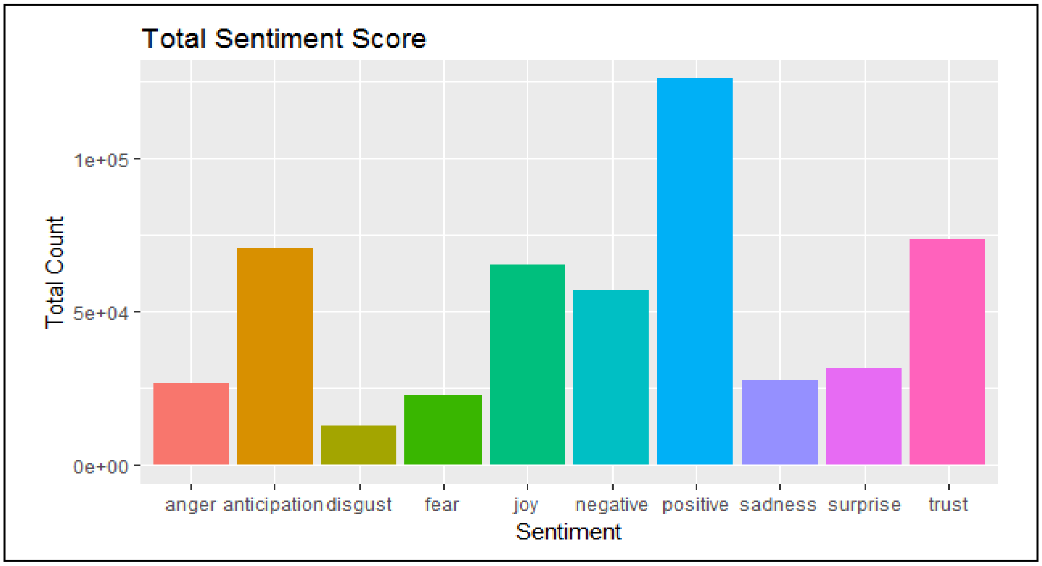


Figure 5.13: Sentiment Analysis of Samsung Reviews

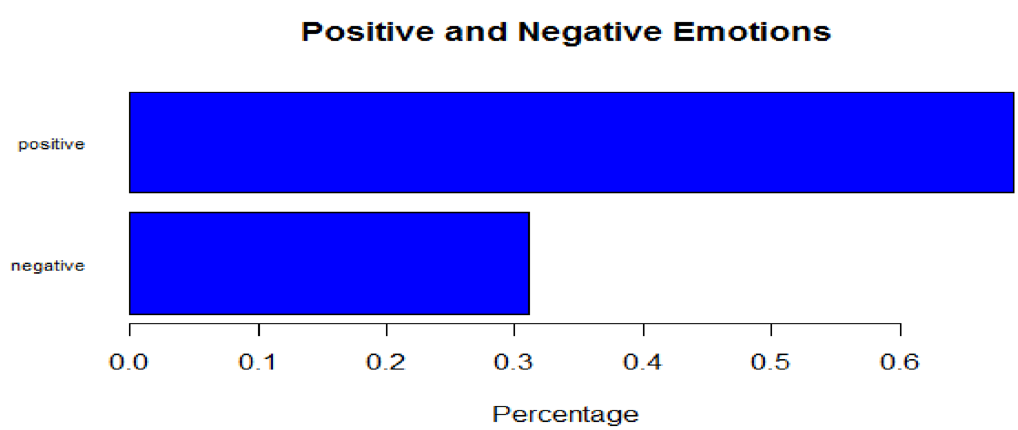


Figure 5.14: Percentage of Positive and Negative Emotions of Samsung Reviews

- **Sentiment Analysis of BLU reviews**

It can be observed in Figure 5.15 that the positive sentiments of joy, trust, anticipation and surprise associated with mobile phones manufactured by BLU are high than the negative sentiments of disgust, fear and sadness. Likewise, the overall polarity of the BLU brand is positive and is more than double the negative reviews as seen in Figure 5.16.

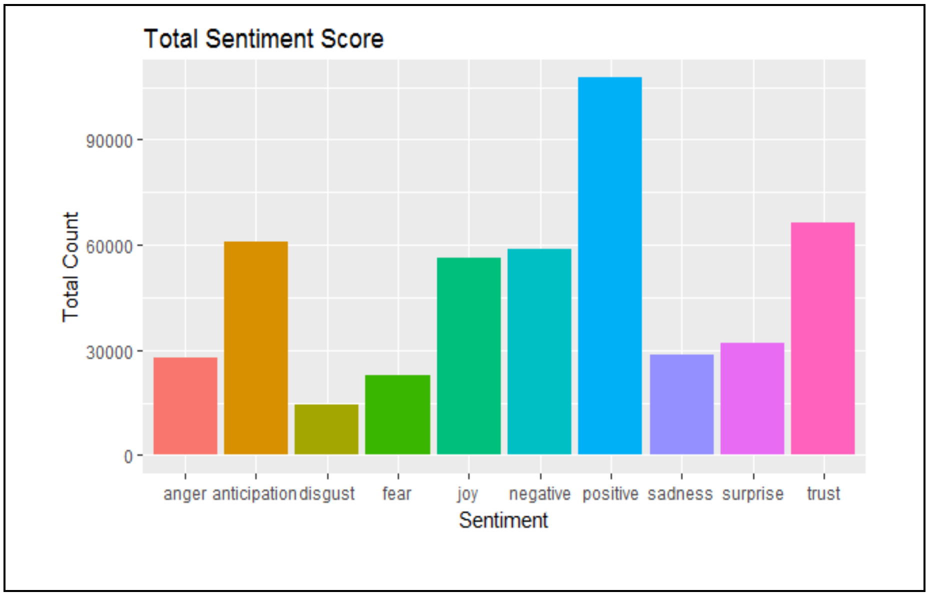


Figure 5.15: Sentiment Analysis of BLU Reviews

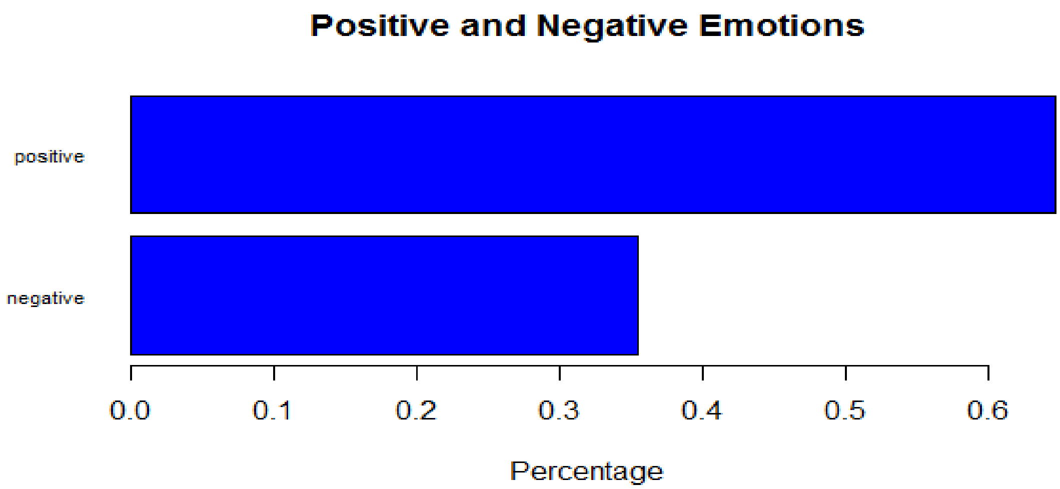


Figure 5.16: Percentage of Positive and Negative Emotions of BLU Reviews

- **Sentiment Analysis of Apple Reviews**

In Figure 5.17, feelings of anticipation, joy, trust and surprise are far greater than the feelings of disgust, fear and sadness. In Figure 5.18 the number of positive reviews is almost 130% more than the number of negative reviews. Thus, it can be said that the overall sentiment of mobile phones manufactured by Apple is positive.

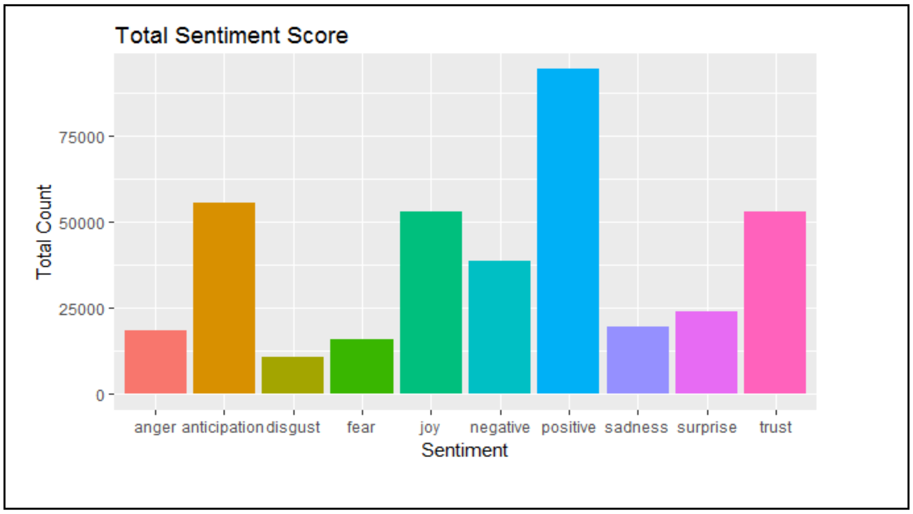


Figure 5.17: Sentiment Analysis of Apple Reviews

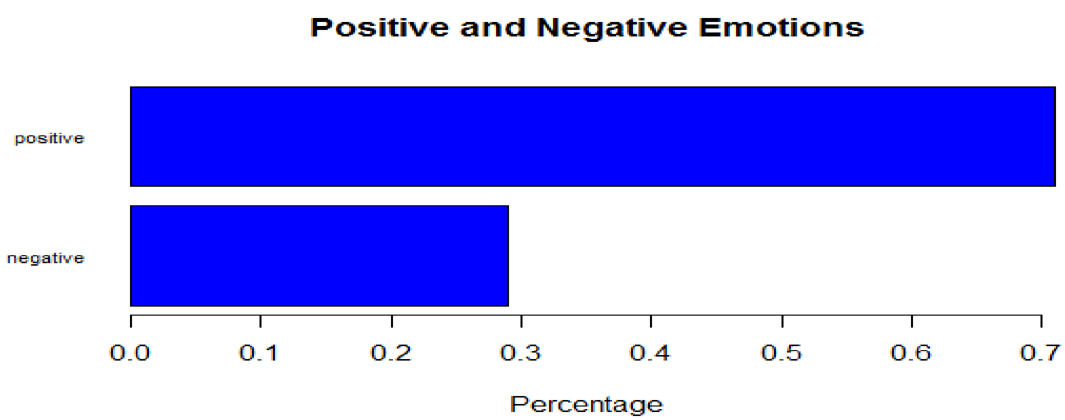


Figure 5.18: Percentage of Positive and Negative Emotion of Apple Reviews

5.3.2 Binary Classification

The class distribution of the classified data is given in Table 5.5. It can be observed that the positive class is highly dominant in the dataset where Positive class is more than six times the negative class.

Table 5.5: Class Distribution

Class	**Count**
Positive	352513
Negative	56371

Four techniques followed for data balancing are Undersampling, Oversampling, both Undersampling and Oversampling and SMOTE.

- **Undersampling:** The class distribution after performing undersampling on the data is given in Table 5.6. The class distribution is quite balanced now with not much difference.

Table 5.6: Class Distribution of Undersampled Data

Class	**Count**
Positive	43629
Negative	56371

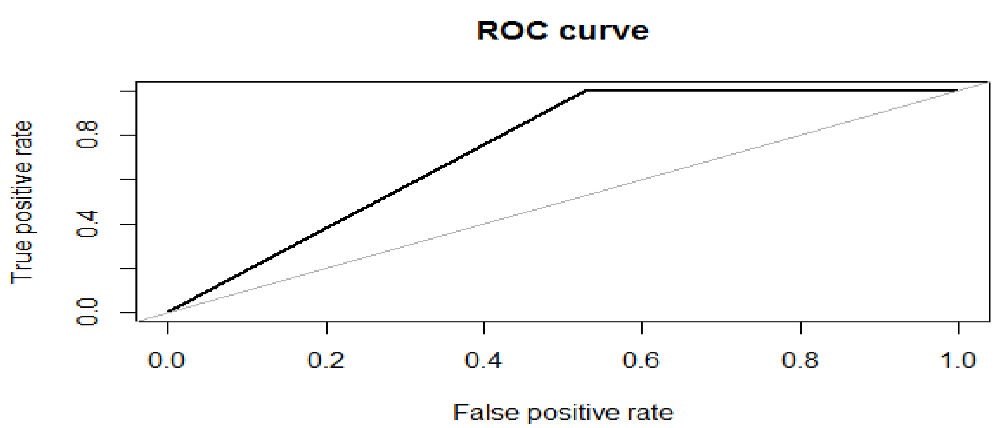


Figure 5.19: ROC Curve for Undersampled Data

- **Oversampling:** Class distribution of the oversampled data is shown in Table 5.7. Figure 5.20 showcases the ROC curve for the distribution. ROC curve for oversampled data shows better accuracy than the graph of undersampled data.

Table 5.7: Class Distribution of Oversampled Data

Class	**Count**
Positive	352513
Negative	247487

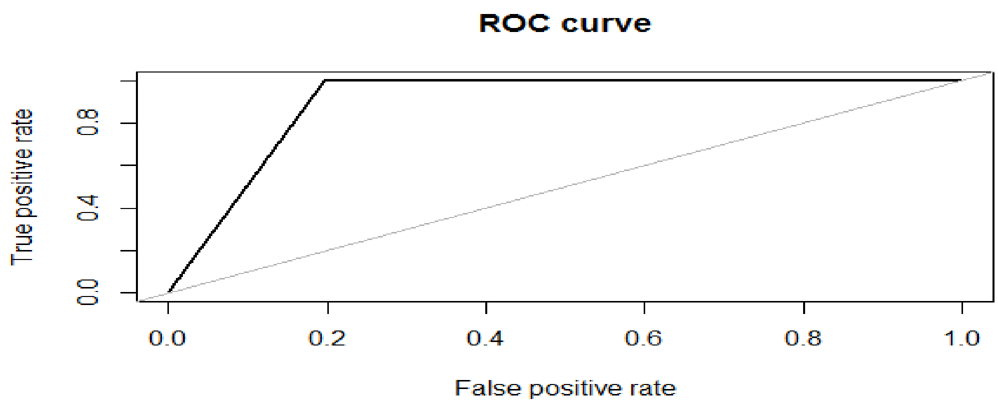


Figure 5.20: ROC Curve for Oversampled Data

- **Both undersampling and oversampling:** Table 5.8 depicts the positive and negative distribution when both undersampling and oversampling is applied together. The ROC curve as shown in Figure 5.21 demonstrates better accuracy than the previous two methods applied.

Table 5.8: Class Distribution of both Undersampled and Oversampled Data

Class	**Count**
Positive	50008
Negative	49992

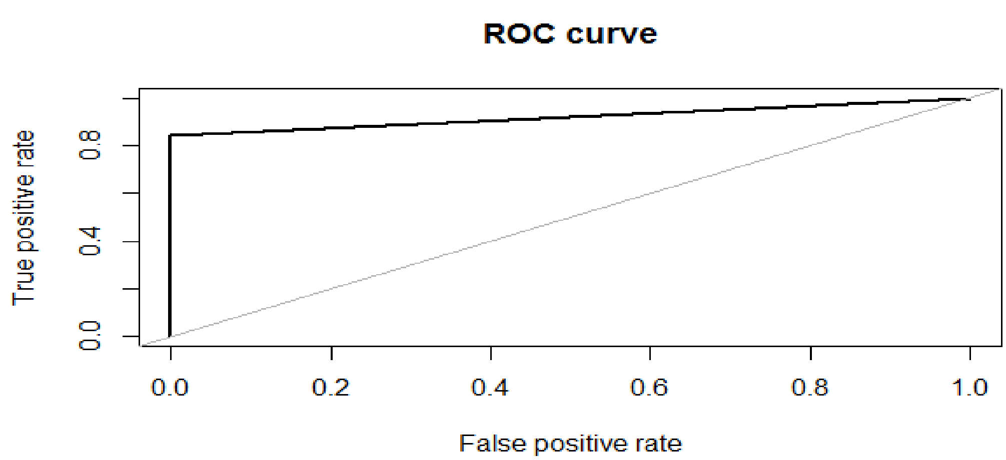


Figure 5.21: ROC Curve for both Undersampled and Oversampled Data

- **SMOTE:** Distribution of the class into positive and negative with its count is shown in Table 5.9. The ROC curve in Figure 5.22 exhibits that this technique is so far the best technique for balancing this particular dataset in comparison to the previous techniques applied.

Table 5.9: Class Distribution using SMOTE

Class	Count
Positive	204158
Negative	204726

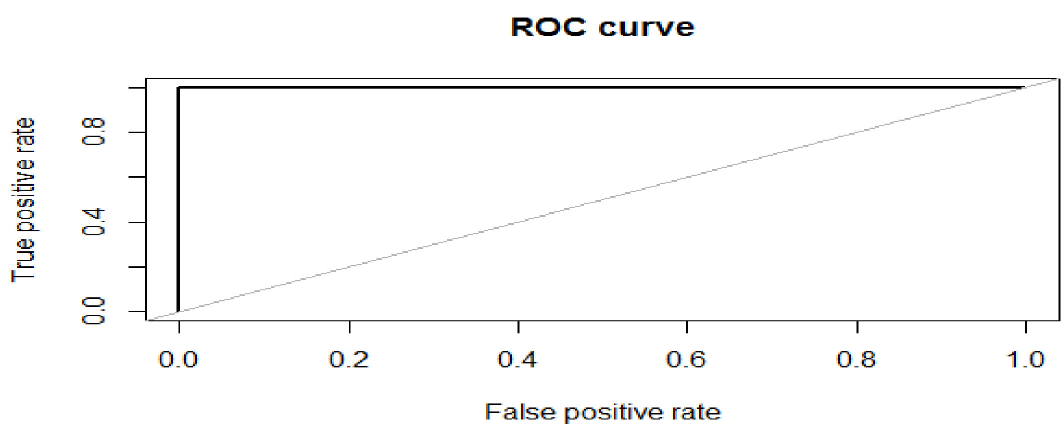


Figure 2.22: ROC Curve for Balanced Data using SMOTE

The accuracy of each of the techniques discussed is shown in Table 5.10. It can be observed that SMOTE technique has the highest accuracy of 99.8% among all.

Table 5.10: Accuracy of Data Balancing Techniques

Data Balancing Technique	Accuracy
Undersampling	73.6
Oversampling	90.2
Both	92.1
SMOTE	99.8

Table 5.11 shows the cross validation of the three models for ten runs. It is observed that the accuracy for all the three models varies in the range of ±10. The scatter plot in Figure 5.23 elucidates that the SVM model reaches the highest accuracy mark of 84.87% among all the models for a number of iterations. Naïve Bayes model has the lowest accuracy of 57.48% among the three models. The graph depicts that SVM model has the highest accuracy out of the three models and Naïve Bayes model has the least predictive accuracy. Table 5.12 shows the predictive accuracies of three classifiers along with recall accuracy. Recall accuracy of SVM is the highest and that of decision tree is the lowest.

Table 5.11: Cross Validation for Binary Classification

Runs	**Naïve Bayes**	**SVM**	**Decision Tree**
1	56.10	85.95	74.14
2	58.14	84.44	77.13
3	56.52	83.69	70.80
4	54.95	86.11	78.00
5	60.52	85.09	72.63
6	56.46	84.58	76.22
7	59.58	82.11	71.84
8	57.16	85.85	77.62
9	55.20	85.66	67.68
10	60.20	85.22	81.25

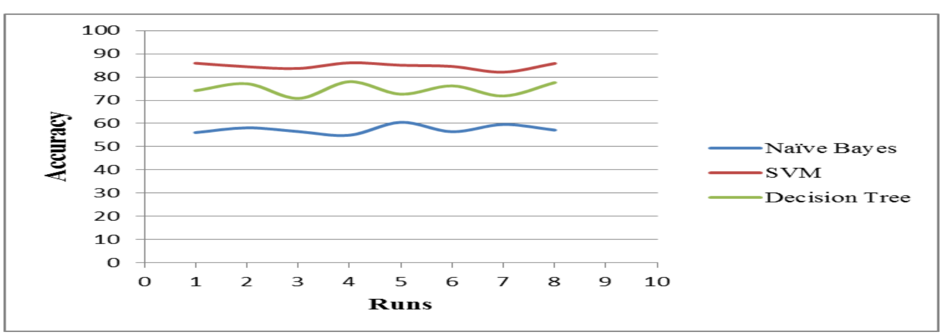


Figure 5.23: Cross Validation for Binary Data

Table 5.12: Predictive Accuracy of Models

Model Name	Predictive Accuracy	Recall
Naïve Bayes	57.48	40.21
SVM	84.87	66.70
Decision Tree	74.73	13.34

5.3.3 Multi-label classification

The three models are cross validated 10 times. Table 5.13 shows the cross validation of the three models for ten runs. It is observed that the accuracy of all the three models in all the iterations varies in the range of ±10. The scatter plot in Figure 5.24 elucidates that the SVM model reaches the highest accuracy mark of 87.48% among all the models for a number of iterations. Naïve Bayes model has the lowest accuracy of 35.70% among the three models. The graph clearly depicts that SVM model has the best accuracy out of the three models and Naïve Bayes model has the least predictive accuracy. Table 5.14 lists the predictive accuracies of the three classifiers.

Table 5.13: Cross Validation for Multi-label Classification

Runs	Naïve Bayes	SVM	Decision Tree
1	34.85	86.43	46.11
2	33.91	83.41	42.60
3	33.8	90.80	42.72
4	39	85.86	44.65
5	29	86.51	49.46
6	40.40	86.63	47.02
7	36.80	85.90	44.39
8	37.25	88.94	50.88
9	36.66	89.41	49.26
1	35.40	90.95	46.30

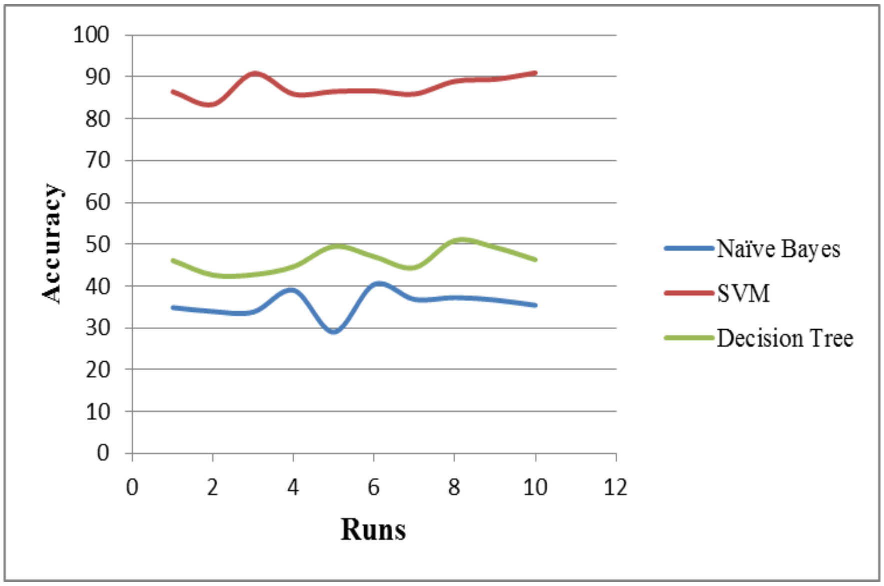


Figure 5.24: Cross Validation for Multiclass Data

Table 5.14: Predictive Accuracy of Models

Model Name	**Predictive Accuracy**
Naïve Bayes	35.70
SVM	87.48
Decision Tree	46.33

5.3.4 Comparison of Performance between Binary and Multi-label Distribution of Data

Table 5.15 exhibits the predictive accuracy of each of the three models both for binary and multiclass distribution of data. Figure 5.25 showcases the accuracy for the same in the form of bar graph. The predictive accuracy of SVM is the highest in both the cases and its accuracy improves when multiclass data is used to train the classifier. The predictive accuracy of Naïve Bayes is the lowest for both binary and multiclass data. However, the performance of Naïve Bayes model degrades when trained using multi-label data. It can also be seen that the efficiency of Decision Tree model decreases extensively for multiclass data.

Table 5.15: Predictive Accuracy of Binary and Multiclass Data

Model Name	**Binary Predictive Accuracy**	**Multiclass Predictive Accuracy**
Naïve Bayes	57.48	35.70
SVM	84.87	87.48
Decision Tree	74.73	46.33

5.3.5 Sentiment Classification using Deep Learning

Deep learning is performed using two methods: Vocabulary Based Vectorization and Feature Hashing.

- **Vocabulary Based Vectorization**

Sentiment classification is performed for unigram and bigram.

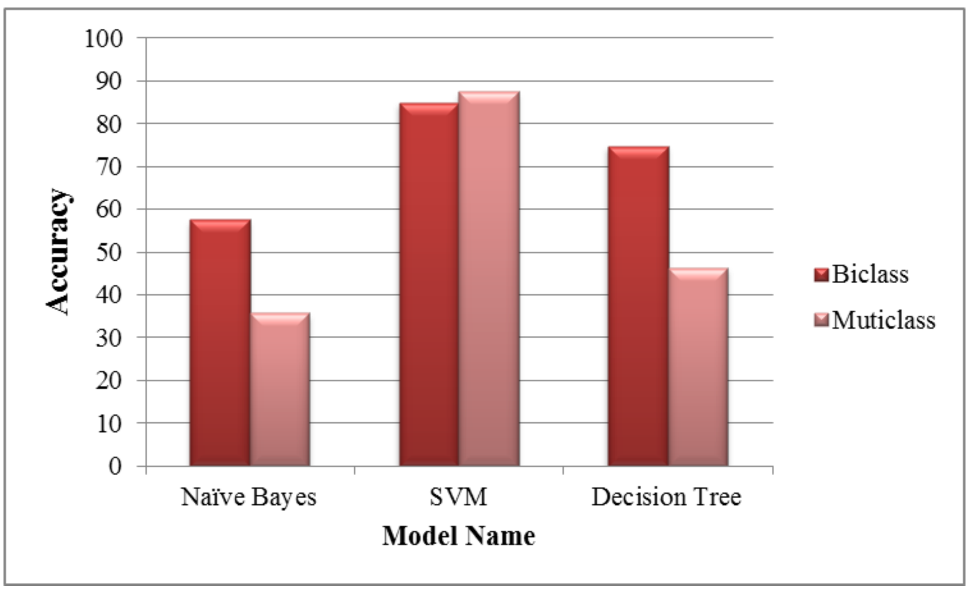


Figure 5.25: Binary and Multiclass Accuracy Comparison

- *Unigram*

Accuracy distribution of the 'glmnet' model is determined using 5-fold cross validation is performed. The maximum accuracy of the model is 98.64%. Figure 5.26 shows the distribution of accuracy of the model. The accuracy curve has a steep fall which implies that the accuracy varies considerably where the lowest mark is 52.94%.

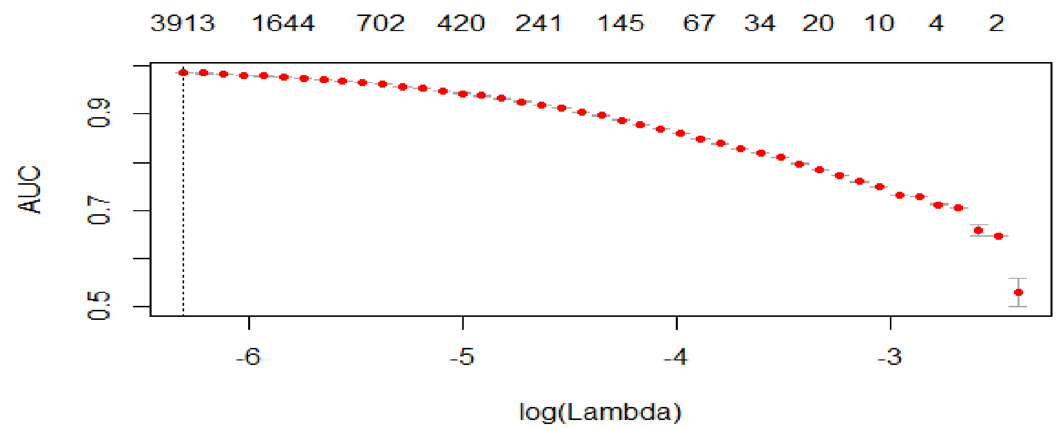


Figure 5.26: Distribution of Accuracy in Unigram

- *Bigram*

Accuracy distribution of the glmnet model is shown in Figure 5.27 when 5-fold cross validation is performed. The maximum accuracy of the model is 92.09%. The accuracy curve is not as steep as that of Unigram in Figure 5.26. Hence, the difference between accuracies is less.

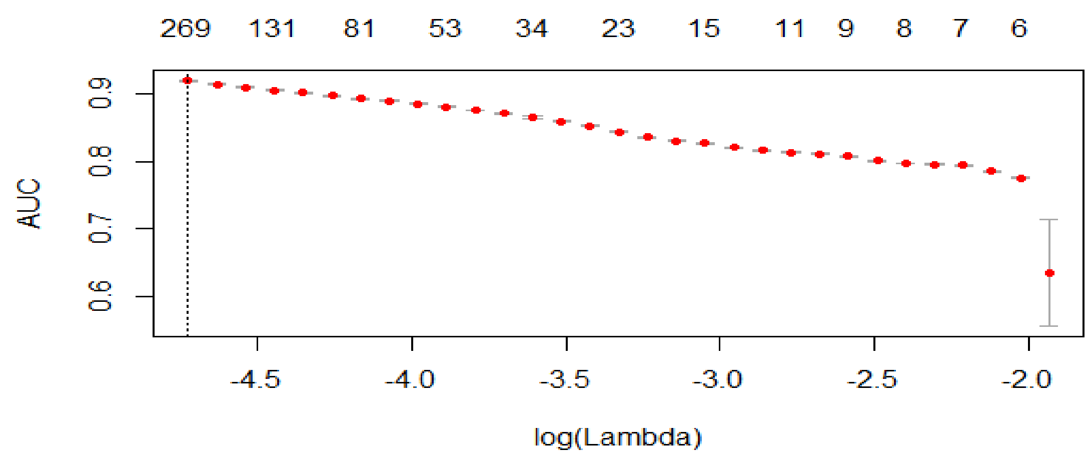


Figure 5.27: Distribution of Accuracy in Bigram

- **Feature Hashing**

Accuracy distribution of the 'glmnet' model is shown in Figure 5.28 when 5-fold cross validation is performed. The maximum accuracy of the model is 92.14%. Figure 5.28 shows the distribution of accuracy of the model. Feature Hashing's accuracy curve lies in between the Unigram and Bigram in terms of steepness.

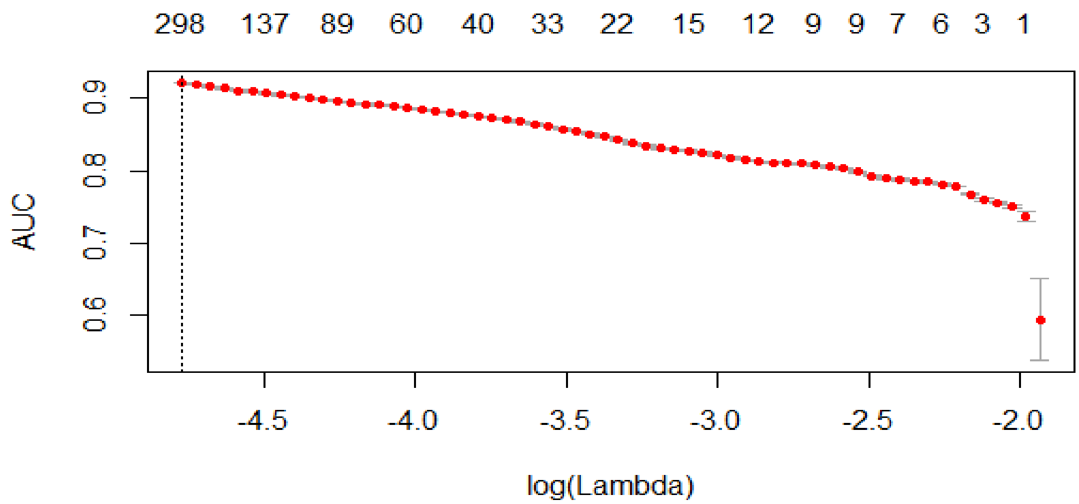


Figure 5.28: Distribution of Accuracy in Feature Hashing

Table 5.16 lists the accuracy of the three techniques used and the bar graph shown in Figure 5.29 plots their respective accuracies. It can be observed that Vocabulary-based categorization considering 1-gram outperforms the other two. The proposed model is then tested to check its performance when it is fed with unseen data. The confusion matrix in Table 5.17 demonstrates the actual and predicted classification of reviews using unigram vocabulary-based vectorization where 0 denotes positive class and 1 denotes negative class.

Table 5.16: Performance Comparison of Unigram, Bigram and Feature Hashing

Technique Used	**N-grams**	**Accuracy**
Vocabulary-based Vectorization	**Unigram**	98.81%
	Bigram	92.09%
Feature Hashing		92.14%

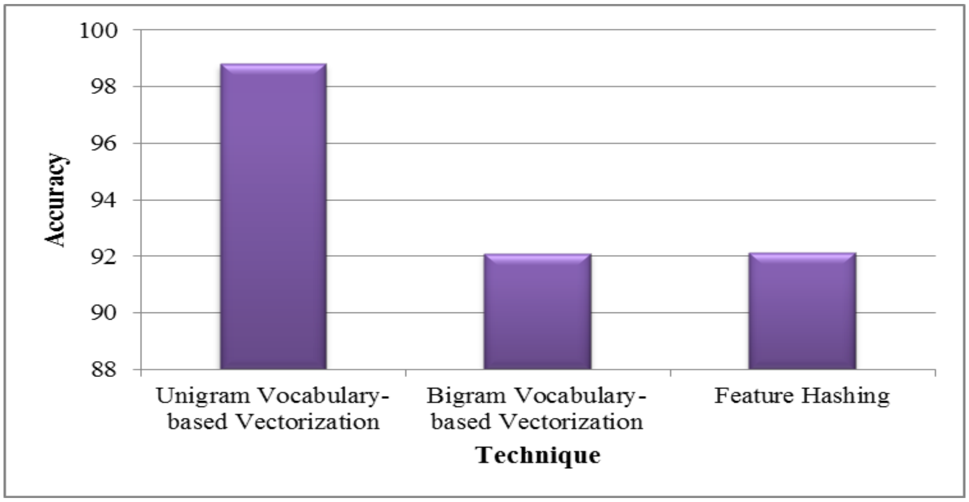


Figure 5.29: Performance Comparison of Deep Learning Techniques

Table 5.17: Confusion Matrix for Unigram Vocabulary-based Vectorization

Prediction	**0**	**1**
0	40042	2886
1	945	37904

Table 5.18 depicts the values of different evaluation metrics for the predicted model. It can be observed that the model has high accuracy in terms of positive and negative predicted values. It has high sensitivity and specificity; hence the rate at which positive and negative values are predicted correctly is quite efficient. The overall predictive accuracy of the model is very high. Therefore, the model can be considered very reliable and efficient.

Table 5.18: Evaluation Metrics for Unigram Vocabulary-based Vectorization

Evaluation Metric	**Percentage**
Positive Predicted Value	93.28%
Negative Predicted Value	97.57%
Sensitivity	97.69%
Specificity	92.62%
Predictive Accuracy	98.81%

5.3.6 Performance Comparison of Supervised Learning and Deep Learning in Sentiment classification

Table 5.19 lists the accuracies of the techniques employed for sentiment classification of the text. It can be observed that the three learning algorithms used are less accurate than the algorithms used in deep learning. Thus, deep learning is more efficient for binary classification of text. Figure 5.30 shows the bar plot of the accuracies of each of the techniques followed.

Table 5.19: Performance comparison of Supervised Learning and Deep Learning

Learning Technique	**Algorithm used**	**Accuracy**
Supervised Learning	Naïve Bayes	57.48%
Supervised Learning	SVM	84.87%
Supervised Learning	Decision Tree	74.73%
Deep Learning	Unigram Vocabulary-based Vectorization	98.81%
Deep Learning	Bigram Vocabulary-based Vectorization	92.09%
Deep Learning	Feature Hashing	92.14%

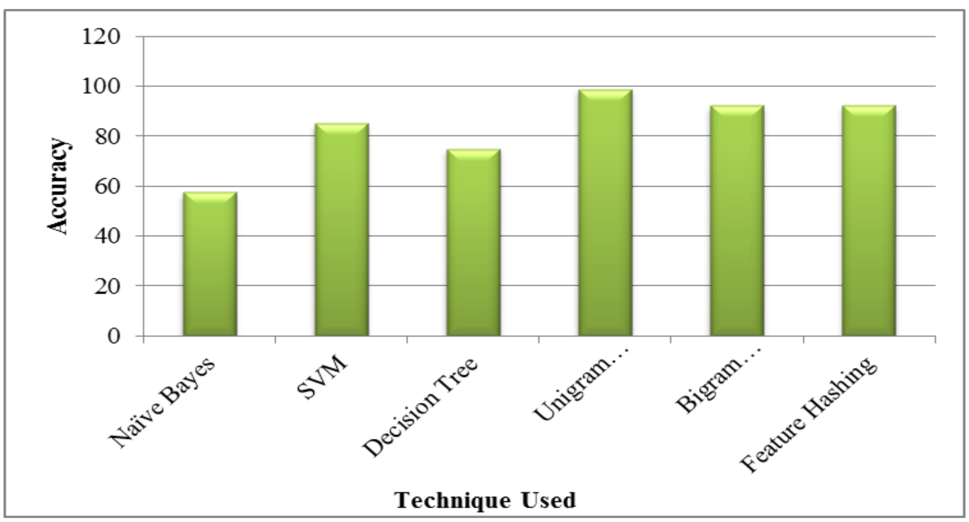


Figure 5.30: Accuracy of Sentiment Analysis Techniques Applied

5.3.7 Rule-based Extraction of Reviews

Reviews of top five products of apple have been chosen namely, Iphone 6, Iphone 6 plus, Iphone 7 and Iphone 7 plus for extracting the features of these products. Top five nouns, verbs, adjectives and adverbs present in the reviews in terms of their occurrence are shown in the Table 5.20.

Table 5.20: Top 5 Nouns, Verbs, Adjectives and Adverbs

S.No.	**Nouns**	**Verbs**	**Adjectives**	**Adverbs**
1.	Battery	Working	Great	Not
2.	Screen	Expected	Good	Really
3.	Camera	Looked	New	So
4.	iPhone	Added	Better	Well
5.	Touch	Compared	Bright	Fast

Table 5.21 showcases the features obtained from the nouns present in reviews along with their frequency. Total five features of relevance have been shortlisted.

Table 5.21: Features extracted from Nouns

Features	**Frequency**
Battery	108
Screen	74
Camera	73
Touch	68
Storage	8

Over 10,000,00 rules are formed out of which rules having features as consequent are chosen. The sample of rules generated corresponding to each feature is given below. Minimum support and minimum confidence specified are 0.05 and 0.8 respectively for the generation of rules.

- *Rules based on Battery*

Figure 5.31 illustrates the top 20 frequently occurring terms in reviews featuring battery. Words like ‘life’, ‘battery’, ‘good’ and ‘great’ have very high frequencies. It can be inferred from this frequency plot that iPhone’s battery is good. This can be checked from the rule database as well.

Rules based on Battery are shown in Table 5.22. It can be inferred that some of the battery of iphones ‘crash’, ‘drain fast’ and ‘heat up’. While for some users batteries worked absolutely fine with no issues. This can be inferred from the presence of positive words like ‘nice’, ‘amazing’ and ‘better’.

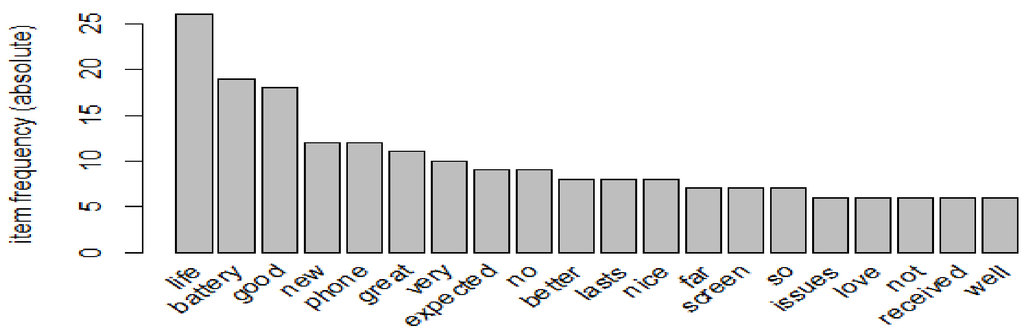


Figure 5.31: Frequent Terms in Battery Reviews

Table 5.22: Sample of Rules for Battery

Rules
{crashing} => {battery}
{drain} => {battery}
{heats} => {battery}
{works} => {battery}
{nice} => {battery}
{crashing, drain} => {battery}

{crashing, really} => {battery}
{crashing, update} => {battery}
{crashing, fast} => {battery}
{drain, fast} => {battery}
{come, heats} => {battery}
{heats, not} => {battery}
{battery, received} => {tests} {issues, no} => {tests}
{amazing, have} => {battery}
{better, expected, much, well, working} => {battery}

- *Rules based on Screen*

Figure 5.32 showcases the top 20 words occurring in reviews containing screen as feature with 'not' and 'flickering' occurring very frequently. This gives a subtle idea that maybe the screen has or doesn't have flickering problem. The rules obtained as shown in Table 5.23 provide the following characteristics of the screen of iPhone: Excellent screen, easy to read screen, unlocks in a fraction of second, bigger screen and a good read. Hence, the rules provide a positive feedback about the screen.

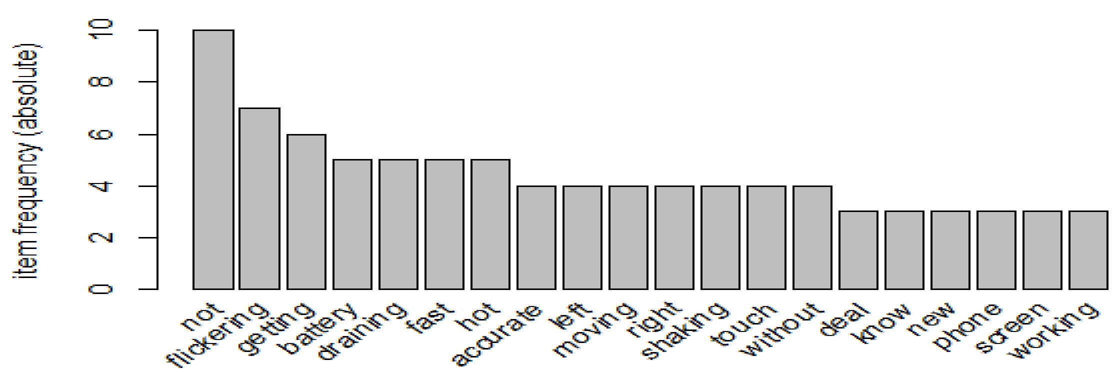


Figure 5.32: Frequent Terms in Screen Reviews

Table 5.23: Sample of Rules for Screen

Rules
{excellent} => {screen}
{read} => {screen}
{older} => {screen}
{easy} => {screen}
{fraction, unlocks} => {screen}
{fast, unlocks} => {screen}
{bigger, carrying} => {screen}
{bigger, excellent} => {screen}
{bigger, size} => {screen}
{carrying,excellent}=> {screen}
{love, read} => {screen}
{good, read} => {screen}
{easy, great} => {screen}
{easy, good} => {screen}
{fraction, second, unlocks} => {screen}

- *Rules based on Camera*

The frequently occurring words in reviews related to camera are ‘great’, ‘good’ and ‘camera’ as exhibited in Figure 5.33. Thus, it can be inferred that mainly positive words are associated with these reviews.

By going through the rules database of camera listed in Table 5.24, one can interpret that the camera is nice, working good, nice front camera, much better than before and is as expected. Hence, the camera quality of iPhone is good overall.

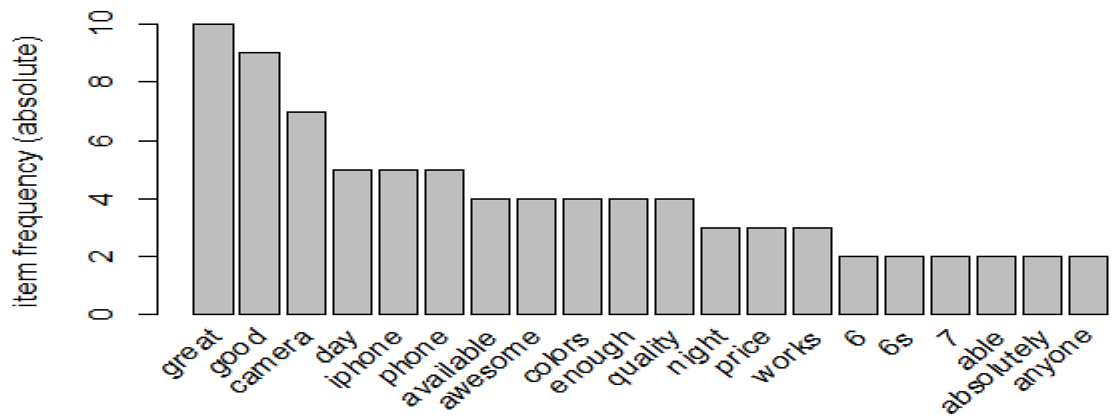


Figure 5.33: Frequent Terms in Camera Reviews

Table 5.24: Sample of Rules for Camera

Rules
{works} => {camera}
{nice} => {camera}
{description,well}=> {camera}
{description, good} => {camera}
{front, nice} => {camera}
{better, much} => {camera}
{love, much} => {camera}
{good, much} => {camera}
{thank, works} => {camera}
{card, expected, love, much, working} => {camera}

- *Rules based on Touch*

The frequency plot in Figure 5.34 depicts the high frequency words present in reviews related to touch. The association rules in Table 5.25 demonstrated that the touch id of an iPhone exhibit stopped working.

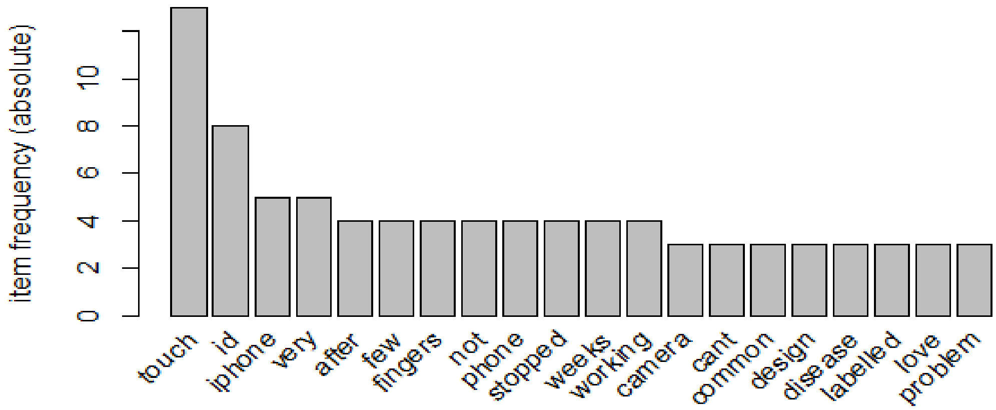


Figure 5.34: Frequent Terms in Touch Reviews

Table 5.25: Sample of Rules for Touch

Rules
{crack} => {touch}
{stopped} => {touch}
{after, stopped} => {touch}
{crack, does} => {touch}
{crack, work} => {touch}
{crack ,id} => {touch}
{button, work} => {touch}
{stopped, working} => {touch}
{after, stopped, working} => {touch}

- *Rules based on Storage*

The rule base related to storage showcased in Table 5.26 demonstrates that 128 GB storage of iPhone is great and as expected. Hence, high level of user satisfaction is there.

Table 5.26: Sample of Rules for Storage

Rules
{128gb} => {storage}
{128gb, expected} => {storage}
{128gb, great} => {storage}
{as, expected} => {storage}
{expected, great} => {storage}
{128gb,as,expected,great}=> {storage}

The feature based extraction of rule-base helps to evaluate the feature performance of mobile phones removing the need of going through the filtered reviews; hence saving a lot of time. Comparison of different mobile phones can be easily performed on the basis of some particular features, not only making the process of choosing a particular handset easier for consumers but also improvising the product the for product designers.

Chapter 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

Through this research, we intend to focus on understanding the relationship between different features and conduct sentiment analysis of the mobile reviews data set.

- The problem stated in Section 3 has been solved as elucidated in Sections 4 and 5. In binary classification, the distribution of positive and negative classes is not balanced, hence four techniques of data balancing are employed, namely, Oversampling, Undersampling, both Undersampling and Oversampling combined and SMOTE. SMOTE technique has been established as the most accurate technique when classification is performed.
- In binary classification, performance of SVM is the best among Naïve Bayes, Decision Tree and SVM. There is an improvement in the performance of SVM when multi-labeled data is fed into the classifier, on the other hand, performance of Naïve Bayes and Decision Tree drops in this case.
- Deep Learning is found to be the most accurate technique for sentiment classification when compared to SVM, Naïve Bayes and Decision Tree classifiers in which contextual polarity and negation effect of the text is considered. Deep learning using Vocabulary-based vectorization is more efficient than Feature Hashing.
- Feature-based extraction of reviews is performed to provide feedback to both the consumers and designers about the performance of the product. Rule-base for latest iPhones is made for representing different features like battery, camera, screen, touch and storage. By examining the rule-base, it can be said that the battery performance was good but there were instances of problems like battery draining, heating and crashing down. When it comes to screen, users were found

to be very satisfied with its size and display. Camera quality of iPhone is found to be good, while in some iPhones touch id crashed. Finally, it was found that users are satisfied with 128 GB of storage.

- Through statistical analysis, it can be concluded that Samsung, BLU and Apple are the three top ranked brands currently in the market. The most positive feedback is received for the brand Samsung. It was also observed that detailed reviews not necessarily attracted better rating and higher priced products didn't fetch delineated reviews. However, better rating was observed in case of high priced products depicting high levels of customer satisfaction and better quality of the products than the low-priced products.

6.2 Summary of Contributions

The proposed research work has performed data analysis and conducted classification of text.

- Sentiment classification of text performed using Lexicon-based approach, Supervised Learning and Deep Learning.
- Both binary and multi-label classification of text has been performed.
- Improved performance of SVM by using multi-labeled data to train the model.
- Contextual polarity and negation effect taken into consideration while classification using deep learning. Deep learning using vocabulary-based vectorization is the most efficient technique for sentiment classification out of the three techniques applied.
- Inference rules are created to generate a rule-base of different features of mobile phones to establish the performance of features pertaining to mobile phones.
- Helpful reviews fetched out on the basis of linguistic features and user rating.
- Data balancing performed to even out the distribution of classes in order to remove any biasness in the results of the classifier.

6.3 Future Scope

In future, this work can be extended to extract features on the basis of customer segmentation targeting people of different age groups. A person of age 20 will need large storage as he will be installing a large segment of applications related to games, education and music, etc. On the other hand, a person of age 70 will be using his phone for lesser tasks, so moderate storage will work for him. Hence, people of different segments have different requirements. The research work can be extended to detect sarcasm and extract the actual meaning of the text. Detecting spam reviews can also be a part of data pre-processing before classification. This can improve the efficiency of the classifiers. It can also be used to predict rating of a product from helpful reviews. This will provide users with reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. This extensibility in the research will be of great benefit to the industry and will make requirement gathering much less time-consuming, shrinking down the expenses incurred using surveys, questionnaires, interviews, market research and trends.

VIDEO PRESENTATION

Video Presentation Link: https://youtu.be/TbJbKactC0U

REFERENCES

[1] P. Russom, Big Data Analytics, 2011.

[2] S. M. Mudambi and D. Schuff, "What makes a helpful online review? A study of customer reviews on Amazon.com," *MIS Quarterly*, 2010.

[3] A. V. Team, "Introduction to Feature Selection methods with an example (or how to select the right variables?)," [Online]. Available: https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/. [Accessed 10 March 2017].

[4] B. Pang and L. Lee, "Seeing stars: Expoliting class relationships for sentiment categorization with respect to rating scales," in *Proceedings of the Association for Computer Language(ACL)*, 2005.

[5] B. Pang and L. Lee, "Subjectivity Detection and Opinion Identification," in *Opinion Mining and Sentiment Analysis*, Now Publishers Inc., 2008.

[6] B. Liu, "Sentiment Analysis and Subjectivity," in *Handbook of Natural Language Processing*, 2010.

[7] "Machine Learning: What it is and why it matters," SAS Institute Inc., [Online]. Available: https://www.sas.com/en_us/insights/analytics/machine-learning.html. [Accessed 12 February 2017].

[8] R. Bhonde, B. Binita, S. Ingulkar and A. Pande, "Sentiment Analysis Based on Dictionary Approach," *International Journal of Emerging Engineering Research and Technology* , vol. 3, no. 1, pp. 51-55, 2015.

[9] R. Feldman, "Sentiment Analysis Tutorial," [Online]. Available: http://ijcai13.org/files/tutorial_slides/tf4.pdf. [Accessed 20 March 2017].

[10] "DL4J DEEPLEARNING4J," [Online]. Available: https://deeplearning4j.org/doc2vec.html. [Accessed 27 April 2017].

[11] S. S. Mukku, N. Choudhary and R. Mamidi, "Enhanced Sentiment Classification of Telugu Text using ML Techniques.," *SAAIP@ IJCAI*, pp. 29-34, 2016.

[12] H. Shirani-Mehr, "Applications of Deep Learning to Sentiment Analysis," 2014. [Online]. Available: https://cs224d.stanford.edu/reports/Shirani-MehrH.pdf.

[13] S. Erevelles, N. Fukawa and L. Swayne, "Big Data Consumer Analytics and the Transformation of Marketing," *Journal of Business Research,* vol. 69, no. 2, pp. 897-904, 2016.

[14] G. George, M. Haas and P. A., "Big Data and Management," *Academy of Management Journal,* vol. 57, no. 2, pp. 321-326, 2014.

[15] S. C., Humanizing Big Data: Marketing at the meeting of data, social science and consumer insight, Kogan Page Publishers, 2015.

[16] M. Stone, Consumer insight: How to use data and market research to get closer, Kogan Page Publishers, 2004.

[17] N. Korfiatis, D. Rodríguez and M. Sicilia, "The Impact of Readability on the Usefulness of Online Product Reviews: A Case Study on an Online Bookstore," *Springer,* vol. 5288, pp. 423-432, 2008.

[18] P. J. Sher and S.-H. Lee, "Consumer skepticism and online reviews: An Elaboration Likelihood Model Perspective," *Social Behavior and Personality: An International Journal,* vol. 37, no. 1, pp. 137-143, 2009.

[19] P. Resnick and R. Zeckhauser, "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," in *The Economics of the Internet and E-commerce*, Emerald Group Publishing Limited, 2002, pp. 127-157.

[20] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: predicting the usefulness and impact of reviews," in *Proceedings of the ninth international conference on Electronic commerce,* ACM, 2007, pp. 303-310.

[21] S. M. Mudambi and D. Schuff, "What makes a helpful review? A study of customer reviews on Amazon. com," 2010.

[22] P.-Y. Chen, S. Dhanasobhon and M. D. Smith, All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon. com, 2008.

[23] L. Zhu, G. Yin and W. He, "Is this opinion leader's review useful? Peripheral cues for online review helpfulness," *Journal of Electronic Commerce Research,* vol. 15, no. 4, p. 267, 2014.

[24] K. K. a. H. K.-L. Kuan, P. Prasarnphanich and H.-Y. Lai, "What makes a review voted? An empirical investigation of review voting in online review systems," *Journal of the Association for Information Systems,* vol. 16, no. 1, p. 48, 2015.

[25] J. a. C. Y. Liu, C.-Y. Lin, Y. Huang and M. Zhou, "Low-Quality Product Review Detection in Opinion Summarization.," in *EMNLP-CoNLL*, 2007, pp. 334-342.

[26] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics," *Decision Support Systems, Elsevier,* vol. 81, pp. 30-40, 2016.

[27] D. Mazylin, Y. Dover and J. Chevalier, "Promotional reviews: An empirical investigation of online review manipulation.," *The American Economic Review,* vol. 104, no. 8, pp. 2421-2455, 2014.

[28] N. Hu, I. Bose, N. S. Koh and L. Liu, "Manipulation of online reviews: An analysis of ratings, readability, and sentiments," *Decision Support Systems,* vol. 52, no. 3, pp. 674-684, 2012.

[29] Y. a. J. J. Liu, P. Ji, J. A. Harding and R. Y. Fung, "Identifying helpful online reviews: a product designer's perspective," *Computer-Aided Design,* vol. 45, no. 2, pp. 180-194, 2013.

[30] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies,* vol. 5, no. 1, pp. 1-167, 2012.

[31] C.-S. Yang, C.-P. Wei and C. C. Yang, "Extracting customer knowledge from online consumer reviews: a collaborative-filtering-based opinion sentence identification approach," in *Proceedings of the 11th International Conference on Electronic Commerce*, 2009.

[32] C. C. Yang, X. Tang, Y. Wong and C.-P. Wei, "Understanding online consumer review opinions with sentiment analysis using machine learning," *Pacific Asia Journal of the Association for Information Systems,* vol. 2, no. 3, 2010.

[33] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 1997, pp. 174-181.

[34] T. Wilson and J. a. H. P. Wiebe, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of the conference on human language technology and empirical methods in natural language processing*, Association for Computational Linguistics, 2005, pp. 347-354.

[35] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 2004, p. 271.

[36] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining.," *LREc*, vol. 10, no. 2010, 2010.

[37] S. Tan and J. Zhang, "An empirical study of sentiment analysis for chinese documents," *Expert Systems with applications,* vol. 34, no. 4, pp. 2622-2629, 2008.

[38] A. Sharma and S. Dey, "A comparative study of feature selection and machine learning techniques for sentiment analysis}," in *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, ACM, 2012, pp. 1-7.

[39] Q. Ye, Z. Zhang and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert Systems with Applications,* vol. 36, no. 3, pp. 6527-6535, 2009.

[40] Z. Zhang, Q. Ye, Z. Zhang and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Systems with Applications,* vol. 38, no. 6, pp. 7674-7682, 2011.

[41] M. WAHYUDI and D. A. KRISTIYANTI, "SENTIMENT ANALYSIS OF SMARTPHONE PRODUCT REVIEW USING SUPPORT VECTOR MACHINE ALGORITHM-BASED PARTICLE SWARM OPTIMIZATION.," *Journal of Theoretical \& Applied Information Technology,* vol. 91, no. 1, 2016.

[42] D. N. Devi, C. K. Kumar and S. Prasad, "A feature based approach for sentiment analysis by using support vector machine," in *Advanced Computing (IACC), 2016 IEEE 6th International Conference on*, IEEE, 2016, pp. 3-8.

[43] V. Narayanan, I. Arora and A. Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," in *International Conference on Intelligent Data Engineering and Automated Learning*, Springer, 2013, pp. 194-201.

[44] H. Kang, S. J. Yoo and D. Han, "Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications,* vol. 39, no. 5, pp. 6000-6010, 2012.

[45] J. Deriu, A. Lucchi, V. De Luca, A. Severyn, S. Muller, M. Cieliebak, T. Hofmann and M. Jaggi, "Leveraging Large Amounts of Weakly Supervised Data for Multi-Language Sentiment Classification," in *Proceedings of the 26th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, 2017, pp. 1045-1052.

[46] M. d. P. Salas-Zarate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. A. Rodriguez-Garcia and R. Valencia-Garcia, "Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach," *Computational and mathematical methods in medicine,* vol. 2017, no. Hindawi Publishing Corporation, 2017.

[47] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2004, pp. 168-177.

[48] B. Liu, M. Hu and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proceedings of the 14th international conference on World Wide Web,* ACM, 2005, pp. 342-351.

[49] A.-M. Popescu, B. Nguyen and O. Etzioni, "OPINE: Extracting product features and opinions from reviews," in *Proceedings of HLT/EMNLP on interactive demonstrations*, Association for Computational Linguistics, 2005, pp. 32-33.

[50] C. C. Yang, X. Tang, Y. Wong and C.-P. Wei, "Understanding online consumer review opinions with sentiment analysis using machine learning," *Pacific Asia Journal of the Association for Information Systems,* vol. 2, no. 3, 2010.

[51] K. Zhang, R. Narayanan and A. N. Choudhary, "Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking.," *WOSN,* vol. 10, pp. 11-11, 2010.

[52] B. Xu, T.-J. Zhao, D.-Q. Zheng and S.-Y. Wang, "Product features mining based on conditional random fields model," in *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, IEEE, 2010, pp. 3353-3357.

[53] C.-S. Yang and H.-P. Shih, "A Rule-Based Approach For Effective Sentiment Analysis.," in *PACIS*, 2012, p. 181.

[54] K. Khan, B. B. Baharudin, A. Khan and others, "Automatic Extraction of Features and Opinion-Oriented Sentences from Customer Reviews," *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering,* vol. 4, no. 2, pp. 102-106, 2010.

[55] Y. Zhai, Y. Chen, X. Hu, P. Li and X. Wu, "Extracting Opinion Features in Sentiment Patterns," in *Information Networking and Automation (ICINA), 2010 International Conference on*, IEEE, 2010, pp. V1-115.

[56] C. C. Chen and Y.-D. Tseng, "Quality evaluation of product reviews using an information quality framework," *Decision Support Systems,* vol. 50, no. 4, pp. 755-768, 2011.

[57] D. Tang, B. Qin and T. Liu, "Aspect level sentiment classification with deep memory network," *arXiv preprint arXiv:1605.08900,* 2016.

[58] F. Luo, C. Li and Z. Cao, "Affective-feature-based sentiment analysis using SVM classifier," in *Computer Supported Cooperative Work in Design (CSCWD), 2016 IEEE 20th International Conference on,* IEEE, 2016, pp. 276-281.

[59] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts and others, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the conference on empirical methods in natural*

language processing (EMNLP), 2013, p. 1642.

[60] O. Abdelwahab and A. Elmaghraby, "UofL at SemEval-2016 Task 4: Multi domain word2vec for Twitter sentiment classification," *Proceedings of SemEval,* pp. 164-170, 2016.

[61] L.-C. Yu, J.-L. Wu, P.-C. Chang and H.-S. Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news," *Knowledge-Based Systems,* vol. 41, pp. 89-97, 2013.

[62] M. Hagenau, M. Liebmann and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems,* vol. 55, no. 3, pp. 685-697, 2013.

[63] T. Xu, Q. Peng and Y. Cheng, "Identifying the semantic orientation of terms using S-HAL for sentiment analysis," *Knowledge-Based Systems,* vol. 35, pp. 279-287, 2012.

[64] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," *Decision Support Systems,* vol. 53, no. 4, pp. 680-688, 2012.

[65] G. Qiu, X. He, F. Zhang, Y. Shi, J. Bu and C. Chen, "DASA: dissatisfaction-oriented advertising based on sentiment analysis," *Expert Systems with Applications,* vol. 37, no. 9, pp. 6182-6191, 2010.

[66] T.-K. Fan and C.-H. Chang, "Blogger-centric contextual advertising," *Expert systems with applications,* vol. 38, no. 3, pp. 1777-1788, 2011.

[67] M. Altini, "DEALING WITH IMBALANCED DATA: UNDERSAMPLING, OVERSAMPLING AND PROPER CROSS-VALIDATION," 2015. [Online]. Available:http://www.marcoaltini.com/blog/dealing-with-imbalanced-data-undersampling-oversampling-and-proper-cross-validation.

[68] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.

[69] K. P. Murphy, "Naive bayes classifier," *University of British Columbia,* 2006.

[70] J. Friedman, T. Hastie, N. Simon and R. Tibshirani, "Package 'glmnet'," 6 May 2016. [Online]. Available: ftp://debian.ustc.edu.cn/CRAN/web/packages/glmnet/glmnet.pdf.

[71] M. Schweinberger, "Part-of-Speech Tagging with R," 2016. [Online]. Available: http://www.marcoaltini.com/blog/dealing-with-imbalanced-data-undersampling-oversampling-and-proper-cross-validation.

[72] M. Michalke, "Using the koRpus package for text analysis," 2014.

[73] K. Hornik, "Package 'OpenNLP'," 18 February 2016. [Online]. Available: https://cran.r-project.org/web/packages/openNLP/openNLP.pdf.

[74] Y. Zao, "Association Rule Mining with R," October 7 2016. [Online]. Available: http://www.rdatamining.com/docs/association-rule-mining-with-r.

[75] M. Stone, Consumer insight: How to use data and market research to get closer, Kogan Page Publishers, 2004.

[76] K. M. Leung, "Naive bayesian classifier," *Polytechnic University Department of Computer Science/Finance and Risk Engineering,* 2007.

[77] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of machine learning research,* vol. 2, no. Nov, pp. 45-66, 2001.

[78] I. a. A. N. B. a. E. Z. Jenhani, "Decision trees as possibilistic classifiers," *International Journal of Approximate Reasoning,* vol. 48, no. 3, pp. 784-807, 2008.

[79] J. Pennington, R. Socher and C. D. Manning, "GloVe: Global Vectors for Word Representation," August 2014. [Online]. Available: https://nlp.stanford.edu/projects/glove/.

[80] A. V. C. Team, "Practical Guide to deal with Imbalanced Classification Problems in R," Analytics Vidhya, [Online]. Available: https://www.analyticsvidhya.com/blog/2016/03/practical-guide-deal-imbalanced-classification-problems/. [Accessed 12 April 2017].

[81] A. V. Team, "Understanding Support Vector Machine algorithm from examples," [Online]. Available: https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/. [Accessed 18 April 2017].

[82] J. Schneider, "Cross Validation," 7 Feb 1997. [Online]. Available: https://www.cs.cmu.edu/~schneide/tut5/node42.html. [Accessed 20 April 2017].

[83] A. Wasilewska, "APRIORI Algorithm," [Online]. Available: http://cse.iitkgp.ac.in/~bivasm/uc_notes/07apriori.pdf. [Accessed 2 May 2017].

LIST OF PUBLICATIONS

- Zeenia Singla, Sukhchandan Randhawa and Sushma Jain, “Sentiment Analysis of Customer Product Reviews Using Machine Learning”, International Conference on Intelligent Computing and Control (I2C2), IEEE, 2017. **[Accepted]**
- Zeenia Singla, Sukhchandan Randhawa and Sushma Jain, “Statistical and Sentiment Analysis of Consumer Product Reviews”, International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2017. **[Accepted]**
- Zeenia Singla, Sukhchandan Randhawa and Sushma Jain, “Deep Learning Techniques for Sentiment Classification of Text”. **[Communicated]**

PLAGIARISM REPORT