TEXT BASED OPINION MINING ON TAX TWEETS USING DEEP LEARNING NEURAL NETWORK

Thesis submitted in Partial Fulfilment of the Degree of Master of Philosophy (M.Phil)

BY

SP.ANITHA

(18MPCSF001)

Department of Computer Science

Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 641043

July 2019

DECLARATION

I declare that the dissertation entitled **"Text Based Opinion Mining on Tax Tweets Using Deep Learning Neural Network"** submitted by me for the degree of Master of Philosophy (M.Phil.) is the record of work carried out by me during the period from August 2018 to July 2019 under the guidance of **Dr. (Mrs.) V. Radha, Professor & Head , Department of Computer Science,** Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, and has not Formed the basis for the award of any Degree, Diploma, Associateship ,Fellowship, Titles in this University or any other University or other similar institution of Higher Learning.

Signature of the Candidate

CERTIFICATE

This is to certify that the dissertation entitled **"Text Based Opinion Mining on Tax Tweets Using Deep Learning Neural Network"** submitted for the degree of Master of Philosophy (M.Phil.) by **Ms. SP.Anitha,** is the record of research work carried out by her during the period from **August 2018 to July 2019** under my guidance and supervision, and that this work has not formed the basis for the award of any Degree, Diploma, Associateship, Fellowship or other Titles in this University or any other University or institution of Higher Learning.

31/7/19

Signature of the Head of the Department

Dr. (Mrs) V. RADHA
Professor & Head
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education for Women
Coimbatore - 641 043

31/7/19

Signature of the Supervisor

Dr. (Mrs) V. RADHA
Professor & Head
Department of Computer Science
Avinashilingam Institute for Home Science
and Higher Education for Women
Coimbatore - 641 043

ACKNOWLEDGEMENT

I would like to express my sincere thanks to God Almighty, for his constant love and grace that he showered upon me.

I would like to express my deep sense of reverential gratitude and sincere thanks to **Shri, Dr.P. R. Krishnakumar, Chancellor,**Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for his support and encouragement during the course of my study.

I owe my great deal of gratitude to **Dr. PremavathyVijayan,Vice Chancellor**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for extending all resources that facilitated the conduct of the present work.

I express my gratitude to **Dr. S. Kowsalya, Registrar**, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing all facilities necessary for the work.

I also thankful to **Dr. K.UdayaChandrika, Dean,**School of Physical Sciences and Computational Sciences,Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for granting the facility required.

I wish to place my deep sense of gratitude to **Dr. V. Radha, Professor and Head, Department of Computer Science,**Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, for providing all the facilities required to complete my Research work.I heartily thank my esteemed project for imparting the tremendous assistance and well-timed support for triumph of my research work.

I would like to extend my hearty thanks to one and all who helped me directly or indirectly for successful completion of my research work.

Finally, I take pride to thank my beloved parents, my family members and my friends without whose support, encouragement and kind blessings I would not have succeeded in my Endeavour.

ABSTRACT

The tax plays a vital role in a country's economic contributions and growth. The improvements in tax schemes to raise budget of the state. Text mining can be used to understand the tax system's public opinion. This research initiates the fast development of social media and the dataset are taken from Twitter based on public opinion about the tax tweets. The dataset used is derived from Twitter as a source of data in processing tax tweets. The results of opinions in the form of public sentiment in part of services can be used as consideration to improve the quality of tax services. In this research, text mining is carried out with Deep Learning Neural Network (DNN) from the stages of text processing, feature selection and classification of the text, and also the results are compared with Support Vector Machine (SVM).Feature Selection was done using Information gain (IG) and classification were performed by using Recurrent Neural Network algorithm, and its performance metrics has been evaluated in terms of precision, recall, f-measure and accuracy compared with SVM and DNN.To reduce the issues in tax tweets based on the public opinion mining. Testing can be performed to evaluate the DNN performance using accuracy, precision, recall, and f-measure.

CONTENTS

LIST OF TABLE

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview of Taxes

The country's economy is heavily backed with the amount of state budget of the country. State budget cause taxes, non-tax state income (PNBP) and grant payments in both national and international. Tax payers money become a biggest contribution in government treasury, representing Eighty-five percent among all government budgets. The kinds of taxes often charged are Income Tax (PPh), Value-Added Tax (VAT), Sales Tax on Luxury Goods (PPnBM), Stamp Duty, Land and Building Tax (PBB) and other taxes. In terms of achieving a state budget goal throughout the coming year, various efforts and teamwork are focused on improving the taxation system from the military, general population, and the knowledgeable side. Behind the guidance of the Finance ministry, the strategies of the tax Directorate General keep creating attempts to improve the public consciousness about the tax.

Today's advancement in the World Wide Web (WWW) and social media had already made easy with human beings to share their thoughts. The public uses social media like Twitter, Facebook, Instagram, and path as an instrument to channel the opinions and circumstances around it. Furthermore, tax tweets are based on public sentiment which is useful for the tax plan policies. The public uses social media inspired by the growth of information. In this study, text mining is a text in which the information is generally acquired from papers in order to search for phrases that can express the content of a document to analyze inter-relationships and inter-document classes. To understand the sequence of short-time problems and strategies which occur in society, so that the more suitable approach is used to overcome the problems.

The different kinds of taxes are

- ✓ Income Tax (PPh)
- ✓ Value-Added Tax (VAT)
- ✓ Sales Tax on Luxury Goods (PPnBM)
- ✓ Stamp Duty
- ✓ Land and Building Tax (PBB) and other taxes.

1.2 About Taxes

A tax is based on people or organizations that differ to their corresponding revenue or gains and it is usually calculated as the result of a tax rate earned income. Tax levels might differ by the taxpayer's type or functionalities. The tax paid on businesses was generally referred to as company tax but is imposed at the fixed price.

The online and social networking growth has made it easy for people to express their views. The world is using social media such as Twitter, Facebook, Instagram, and Path as a tool to channel the views and situations around it. In addition, data in the form of public opinion impacts can be used as a tax policy plan. Text mining can be done by classification (classifier) or by examining the frequency and analyzing the sentiment.

Ahmad *et al.* (2017) launched a new technique for analyzing SVM output to detect polarity using various dataset and their main goal was to detect polarity. Their primary focus was on increasing SVM's efficiency. They suggested a framework task composed of several stages such as Preprocessing and Classification to fulfill their objective.

Some of the Twitter Reviews are

- ✓ Amazon products
- ✓ Movie reviews
- ✓ Famous celebrities
- ✓ Social media
- ✓ Airlines
- ✓ Countries
- ✓ Mobile phones etc

1.3 About Text mining with Sentiment Analysis

Text Mining is used to evaluate the post and remarks from internet sites like Wikipedia on Facebook by HimanshiAgrawal *et al.* (2016). The proposal was made on non-profit, entertainment websites, public pages and concluded that the spam activity cannot be totally removed from the internet.

Aldo Hernandez *et al.* (2016) introduce a Twitter content sentiment analysis technique for predicting future internet attacks. The process is focused on collecting tweets from two user sets on a daily basis; individuals who use the platform as a way of expressing opinions on appropriate problems and those who use it to show the content identified with web security threats. Regular information is transformed into data that can be statistically broken down to predict if an attack is feasible. In reaction to a worldwide case, they finally completed by researching the aggregate feeling of users and organizations of hacking activists.

Chandrasekhar Rangu *et al.* (2017) discusses about text mining referred to as text data processing, as the technique to gather non-trivial patterns and attention-grabbing or information

from text files. It utilizes algorithms to remodel (unstructured) free stream text into information that will be analyzed using applied mathematics, language process and machine learning techniques. Text mining is emerging technology that allows companies to understand their clients well and make it easier for them to find the customer wishes. They attempt to obtain some useful data from asset reviews that will be used to improve engineering characteristics of assets and help to improve the quality of service and client experience.

Xing Fang *et al.* (2015) defines the extraction of subjective data, consisting of opinion includes at least one positive or negative terms. Such phrases were tokenized and segregates into words based on the components and the various tags are used. The tokens and results of sentiment are data taken from the real dataset this is called as feature vector.

Duwairi *et al.* (2018) stated that the assessment of sentiments determines the polarity of the specified text using either a machine learning strategy or using a lexicon-based method. The classifiers applied to the datasets were Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), where SVM gave the lowest accuracy and KNN gave the lowest recall. The data sets were also tested with 10-fold cross validation. They showed that the accuracy obtained by SVM is 75.25 %, was the highest accuracy and that the recall received by KNN is 69.04 % was the highest recall. Therefore, to achieve better classification results, larger data sets were required and semi-supervised crowd sourcing was considered to be followed.

AshishShukla *et al.* (2017) addresses the user opinions on the item and its services through multiple sources. The assessment is performed by the sentiment analysis scheme using k means and the Naive Bayes algorithm which uses runtime and decreases complexity. It is easy

to extend the same scheme to view the other product reviews and the rating system is shown in figure. 1.1

Figure.1.1 Rating system

Some of the main tasks in the evaluation of sentiment are

- ✓ The person's remarks on a product are termed as positive in one scenario and negative in another.
- ✓ Some human beings are not in the same manner of expressing an opinion. Many reviews should have positive and negative remarks that can be managed somewhat by analyzing phrases one after the other
- ✓ Sometime people give bad tweets about the product which provides poor evaluation.
- ✓ Manual techniques can sometimes manage the issue of sentiment assessment.

PranaliBorele *et al.* (2016) were used Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms to deliver their greatest outcomes in sentiment analysis, but none of them can solve all the difficulties. Researchers find out SVM has elevated precision compared to various algorithms, it has some constraints and they used ANN, which results in improving ANN with fuzzy logic.

Websites for social networking, such as Facebook, are rich in textbooks that allow users to generate different text content in the form of comments, wall posts, social media, and blogs. A huge quantity of information is accessible via the Web due to the invisible use of social network in latest year. Using text mining methods on social networking websites can show important outcomes in relation to interaction behavior between person to person. In addition, text mining methods in combination with social networks can be used in large-scale systems to find general view on any particular topic, patterns of human thinking, and group identification Aggarwal *et al.* (2011).

Anta *et al*. (2013) provided a comparative study with classification functions of distinct approaches to sentiment analysis of Spanish tweets. Different studies were conducted to classify Spanish tweets according to emotions and subjects. The use of stemmers and lemmatizers, word forms, n-grams, negations, connection processing, valence shifters, search engines, unique Twitter semantics (Hashtags) and various classification techniques were assessed as a thorough and comprehensive research. These results showed that classical techniques could be used to analyze and classify the Spanish text. The best precision seen was 58% for subjects and 42% for classification of sentiments.

Ji Fang *et al*. (2011) were use two classic strategies to sentiment analysis: lexicon look up and machine learning and the lexicon of positive and negative phrases begins. Such language and background sensitivities of sentiment expressions are not captured by current sentiment lexicons. They suggested scheme to introduce an alternative technique that includes sentiment lexicons as initial assumption with approaches to machine learning such as SVM to enhance the precision of sentiment analysis.

Rajeswari *et al*. (2017) used Naive Bayes classifier and KNN classifier to produce a categorized text. The focus was on precision and efficiency with student datasets using Rapid Miner. Compared to KNN, Naive Bayes showed a better precision of 66.67 percent, demonstrating a 38.89 percent precision that was also analyzed.

Arun *et al*. (2017) depicted demonetization sentiment analysis on tweets. They first accessed the information and then converted it as an input dataset into text files. After removing the stop words, sentiment analysis was conducted, the polarity of the phrases was determined and the tweets were classified as positive and negative. A new technique for demonetization sentiment analysis was suggested and data cleaning, sentiment scores, bigrams, polarity and graphical methods were used for this process.

Discussions on data mining methods used to find specific characteristics across products are researched by Negar Hariri *et al.* (2013). New incremental diffusive algorithm is used to extract characteristics from internet product pages and then uses association rule mining and neighboring k-nearest. Machine learning technique is used for features during the domain analysis.

From the literature survey it was found that many methods and techniques are used by the researchers for text mining with sentiment analysis based on public opinions. The important literaturerelated methods for text mining are listed below,

- Support Vector Machine (SVM)
- Naive Bayes (NB)
- K-Nearest Neighbor (KNN)
- Natural Language Processing Technique (NLP)

- Lexicon method
- Hybrid Technique

1.4 Overview of Text Mining with Sentiment Analysis based on Tax Tweets

Text mining deals directly with developed processes for collecting useful information and data review. It has certain methods for analyzing sentiments, classifying and extracting entities that are used to obtain data and knowledge from the hidden text. Text mining techniques are broadly used in studies, evaluation of social media information, virus scanning, business strategies, client care, etc. The process of deriving high-quality information from text is text mining, also referred to as text data mining, which is roughly equivalent to text analytics. Opinions are central to nearly every human activity as they are main creative of our behaviors. Every time we have to make a decision, we want to know the opinions of others. In the real world, public opinion about tax is important. In the past, he/she asked opinion with families and friends. To know the public opinion about the tax, sentiment analysis is used and Deep learning neural network (DNN) is used to get higher accuracy.

1.5 Phase of Text Mining with Sentiment Analysis Based on Tax Tweets

a) Text Processing

Text Processing is the main phase of processing technique and at this point, the research design creates the input text for the process. Classification is based on sentiment and tax, the structure of words acquired from social media remarks is regulated and their requirements are shown in the below Table 1.1

REQUIREMENTS	EXPLANATION
1. Crawling	Take comments using Twitter
2. Case Folding	Change the words on tweets into lower case and by cleaning the hash tag or symbol
3. Convert Emoticons	Converts emoticons into words
4. Tokenizing	Break a comment into snippet
5. Filtering	Delete any irrelevant words
6. Stemming	Reduce every word to get the word base

Table.1.1 Text Processing Stages

b) Feature Selection

Feature selection is the important phase in text mining with sentiment analysis based on tax tweets. It improves the classification task effectively. Features that really discriminate between groups will improve the performance of identification. If any discriminating function is not included in the classification, then the classification accuracy will be reduced.

Some of the Feature selection methods are :

- Wrapper method
- Embedded method
- Filter method
- Conditional mutual information
- Joint mutual information

- **Information Gain (IG)**

Information gain (IG) measures how much "information" a feature gives us about the class. Features that perfectly partition should give maximal information. Unrelated features should give no information and it also measures the reduction in entropy. The value of Entropy is information contained in several distributions such as the distribution of class P(c). Therefore, the Information Gain of some features f measures the value of Entropy P (c) that changes after observed f using the following equation 1

$$\mathbf{IG\ (w) = -\sum_{c=0} P\,(c) log\ P(c) + \sum_{w=(0,1)} P\,(w) \sum_{c=0} P\,(c/w) log\ P(c/w) \quad \dots \quad (1)}$$

Equation 1. Information Gain

c) Classification

The method of classification is to classify the characteristics and provides the result needed. Neural network classifier, SVM, KNN etc, are the most frequently used classifiers. In this research classification was done by SVM and Recurrent Neural Network algorithm.

d) Performance Evaluation

Text classification rules were usually assessed to use performance measures from the retrieved of data. Common metrics for evaluating text categorization include precision, recall, f-measure and accuracy.

1.6 About Deep Learning Algorithms

Machine learning algorithms have been divided in two, they are supervised learning and unsupervised learning. Supervised learning algorithms are being further divided into

classification and regression and unsupervised has been further divided as clustering algorithms. Classification algorithms have been divided into neural networks algorithm, nearest neighbor algorithm, Support Vector Machine (SVM) etc., Figure.1.2 show various machine learning algorithms.

Wang *et al.* (2012) showed that various solutions of machine learning algorithms demonstrate significant variability in their results for text classification. They also proved that bigram showed steady improvements in assignments and Naive Bayes (NB) was more preferable in sentiment assignments than SVM for a small portion. Also, an alternative SVM demonstrates constantly better on datasets and has shown solutions to NB and SVM. It was found from their paper that Multinomial Naive Bayes (MNB) was more in favor of analyzing the sentiment. They came to the conclusion that NB-SVM was a powerful operator and that Bernoulli Naive Bayes (BNB) was worse than MNB

Tiwari *et al.* (2017) proved content-based strategy using sentiment analysis for internet audits, film ratings, etc. Supervised machine learning approaches grouped their views. Three distinct calculations of machine learning were regarded i.e. for findings. SVM, ME, NB and these findings were based on parameters of accuracy and f-measure. In their paper, various machine learning methods were suggested to classify the film reviews of rotten tomatoes dataset using n-gram technique. They also found that their output acquired better precision.

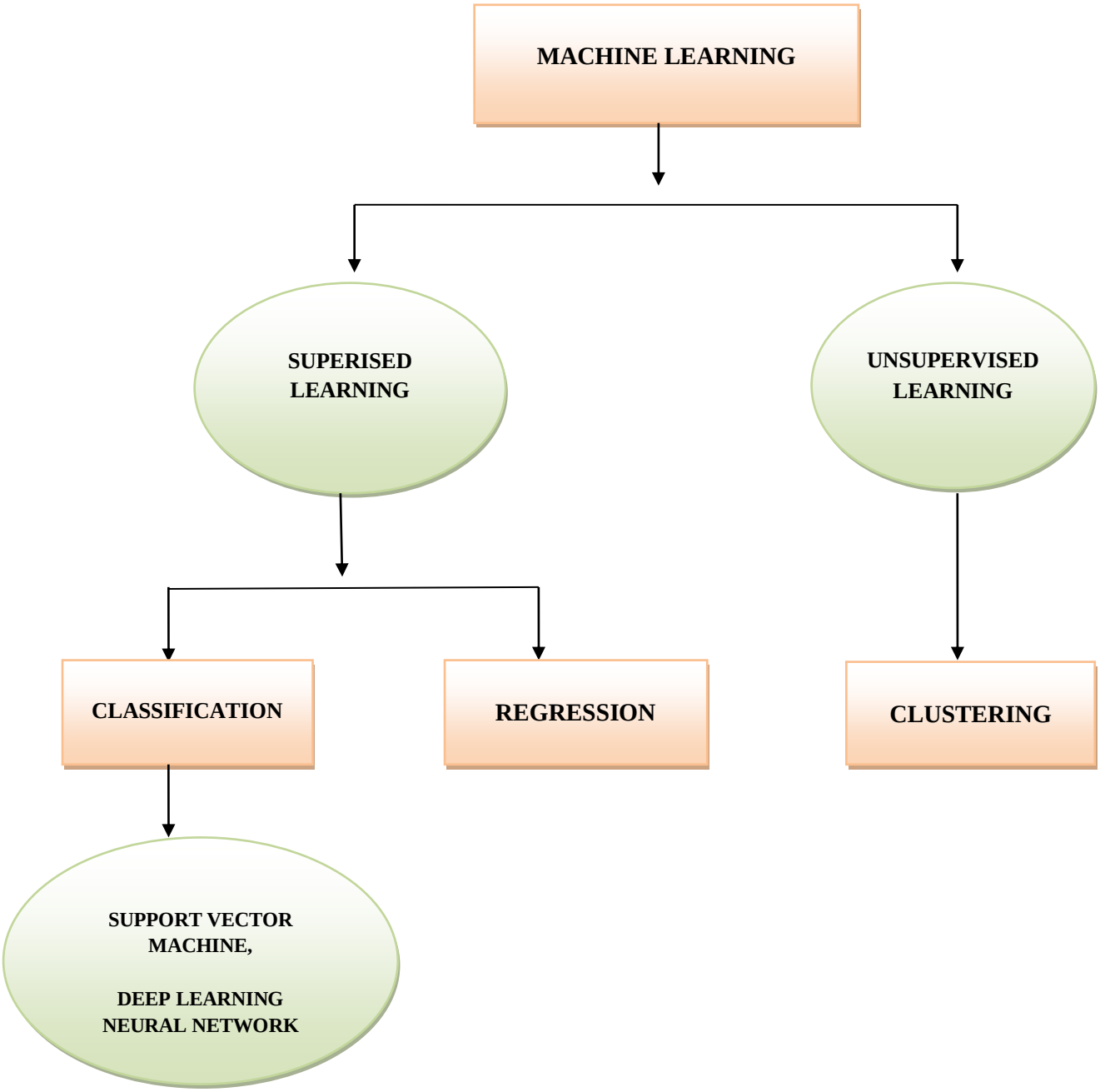


Figure.1.2 Machine learning Algorithms

Tripathy *et al.* (2016) stated that reviews and blog datasets acquired from social networking sites were unsystematic and needed to be classified as significant. Using monitored machine learning techniques, were able to categorized as positive, negative and neutral. They launched four distinct machine learning algorithms to classify feelings. Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient Descent (SGD) and Support Vector Machine

(SVM) based on precision, recall, F-measure. Their paper helped to classify film reviews using supervised machine learning algorithms that were further implemented using n-gram method to the IMDB dataset. They found that the precision reduces in n-gram strategy as the value of n rises. They also encountered certain constraints on further study as tiny size of twitter posts, reviews or remarks including punctuation symbols and phrases like "greatttt, fineee" because they have no adequate significance. So, after removing the stop words to select the best function, a fresh word list was ready for classification. Hybrid machine learning methods have also been regarded for greater precision.

1.7 Recurrent Neural Network

RNN are quite strong since they are combined with two characteristics:

- Distributed hidden state that allows them to store a lot of information about the past efficiently.
- Non-linear dynamics that allows them to update their hidden state in complicated ways.

Recurrent Neural Network (RNN) can calculate anything that the system can calculate with enough neurons and time.

M.D. Bah *et al.* (2018) discusses about the Deep neural network techniques has shown significant improvements over current approaches to machine learning in different fields, primarily computer vision and speech recognition. There are also two kinds of deep learning, Neural Networks they are Recurrent Neural Network (RNN) and Convolution Neural Network (CNN).

1.7.1 Forward Propagation

Algorithm that takes the neural network and the initial input and pushes the input through the network. Each hidden layer accepts the input data, processes it as per the activation function and

passes to the successive layer. In order to generate some output, the input data should be fed in the forward direction only. The data should not flow in reverse direction during output generation otherwise it would form a cycle and the output could never be generated. Such network configurations are known as feed-forward network. The feed-forward network helps in forward propagation.

1.7.2 Back Propagation

Back Propagation method is the primary goal to adapt the weights to reduce an error function and this method is used to minimize the error function hence it is based on the technique of gradient.

1. Takes output from the neural network
   - ✓ Compares it write the actual output
   - ✓ Calculates deviation from the actual output
2. Takes error from the layer
   - ✓ Back propagation calculates error associated with each unit from the preceding layer
   - ✓ Error calculated from each unit is used to calculate partial derivatives
3. Use partial derivatives with gradient descent to minimize cost function

Leo et al. (2017) suggested that fine tuning should be a strategy frequently found in deep learning and this is used to enhance the efficiency of stacked auto-encoders. Because the back propagation algorithm is based on the method of descent gradient and an arbitrary number of layers can be implemented. Back propagation algorithm can be used on arbitrary depth stacked auto encoders. They have embraced the weights of the links in their research work to achieve

minimal distinction between the network output and the required output. The algorithm is very easy and the neural network output is assessed against the required output. The connection between layers will be altered and the process will be performed until the error is sufficiently low or the outcomes are unsatisfactory.

1.8 Motivation

The motivation is to take up text mining based on tax tweets of public opinions. In the current scenario we do not know the public opinion about the taxes. In order to improve the accuracy, it was aimed to classify the polarities based on the tax tweets.

1.9 Problem Statement

To analyze the sentiments of Tax Tweets, where the main focus is to classify the polarities.

1.10 Research Ojectives

- ✓ To improve the accuracy of the algorithm.
- ✓ To compare the performance of SVM and DNN

1.11 Chapter Summary

In this chapter, a brief introduction on text minig based on income tax comments using Deep Learning Neural Network (DNN) have been included. The Recurrent Neural Network algorithm used in this research work has also been explained in this section and the literature reviews were included in the following chapter.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Introduction

This section examines some of the significant literature on text mining and algorithms that are used for text mining are also reviewed.

2.2 Review of Text Mining

AkshayAmolik *et al.* (2016) come out with a technique for analyzing Twitter sentiment to review films using machine learning methods with the assistance of social media from text, they concentrated on identifying and justifying human sentiments. The main disadvantage of their research work was the improper application of the algorithm, resulting in less accuracy. Human emotions were evaluated using text-based algorithms based on machine learning. They used actual data set and 75 percent of SVM and 65 percent of Naive Bayesian were found to be accurate.

AnanthiSheshasaayee *et al.* (2017) compared the classification of algorithms in Text mining. Three algorithms are compared with the accuracy of Naive Bayes, Random Forest, and Support Vector Machine. They noted outcomes on the 88 percent accuracy graph of Random Forest, 90 percent accuracy of Naive Bayes and 97 percent of SVM among these were the highest outcomes generated by SVM compared to other algorithms. The research was conducted by gathering positive and negative sample data set.

Hsin-Ying Wu *et al.* (2014) suggested a method for posting and commenting on Facebook. Their goal was to meet the needs of customers online, and customers can send

feedback directly about a particular product. Their first phase was to gather posts on Facebook and then process them to segment the known and unknown term by Chinese Knowledge and Information Processing (CKIP). They divided the user words, sentences, Facebook posts and conditions. The restriction of their work is to concentrate primarily on segmenting the customer's known and unknown phrases.

Suge Wang *et al.* (2009) suggested a technique for selecting features based on the discriminating proportion of fishermen for classifying text sentiments. Their focus was on identifying both positive and negative reviews. They used positive (thumbs up) or negative (thumbs down) to automate the text sentiments. Using Support Vector Machine (SVM), they proposed four types of feature selection method. They also take three types of measures, such as Precision, Recall and F1 value, using information gathered between January 2006 and March 2007. The development indicates to be 578 positive reviews and 428 negative reviews of Chinese text reviews 1006 and about 11 types of car products.

Neethu M *et al.* (2013) used machine learning methods to analyze public opinion on electronic products using sentiment analysis on Twitter. The data sets were gathered from April 2013 to May 2013 with the assistance of tweets. The data were evaluated using sentiment analysis in which positive or negative remarks were automatically annotated. Preprocessing was performed to prevent tweeting slang words and incorrect spelling.

ChetashriBhadane *et al.* (2015) discussed lexical and machine learning methods for identifying customer opinion mining and detecting mood as well as finding the precision of polarity and classification elements. Using SVM, they achieved a product review accuracy of 78 percent.

Devika *et al.* (2016) conducted a comparative research in sentiment analysis on multiple methods. It was noted that in terms of results, the following approaches such as Machine Learning, Rule Based and Lexicon Based were different from each other. Machine learning seems to be a better precision performance among these methods. Ongoing debates on these methods were conducted in order to obtain better outcomes for assessment of sentiments.

The Gephi tool used by NadeemAkhtar *et al*. (2013) is open source network and graph analysis software for studying social networks. From April 2009 to 2011, they gathered the data set via Facebook. They concentrate on the social network sub-graph containing high-degree nodes on Facebook. The sub-graph file has been transformed to Comma-Separated Values (CSV) format in their research. The amount of friendships is greater, and then they discovered that among these high-degree nodes there is little imminent relationship. Friendships were high, but the Facebook graph showed less of their interactions. The efficiency in Social Network Analysis (SNA) has also been enhanced with the assistance of the Gephi tool.

Shuhufta Fatima *et al.* (2017) developed a technique to categorize text files, by partitioning the files in a proportion of 60-40 and 80-20, they examined the records. As a result, 80 training papers and 20 test papers increased the amount of precision. They concentrated primarily on the current time consumed.

Hassan Saif *et al*. (2015) made a suggestion for Twitter's sentiment analysis on contextual semantics. SentiCircles ideas were implemented to detect Twitter sentiments to demonstrate the outcomes for the identification of sentiments at the entity level and at the tweet level. Obama-McCain Debate (OMD), Standford Sentiment Gold Standard (STS-Gold), Health Care Reform (HCR) datasets have been used for assessment.

Balakrishnan*et al*. (2012) talks about Twitter's micro blog information. In order to enhance precision efficiency, they implemented Synthetic Minority Over-Sampling Technique (SMOTE) and analyzed Twitter text processing. Using Random Forest, SVM and KNN are the algorithm classification with different datasets has been attempted. The level of precision risen to 80 percent compared to the 70 percent precision of Naive Bayes.

Dhanalakshmi *et al.* (2016) worked with supervised learning algorithms on student feedback opinion mining. Using the rapid miner tool, student feedback was found. Positive and negative remarks were categorized and the KNN algorithm was found to be the best in precision where the Naive Bayes algorithm is the best in performance.

Rajeswari *et al.* (2017) used Naive Bayes classifier and KNN classifier to produce a categorized text. The focus was on precision and efficiency with student datasets using Rapid Miner. Compared to KNN, Naive Bayes showed a better precision of 66.67 percent, demonstrating a 38.89 precision that was also analyzed and the graph were shown in the figure.2.1

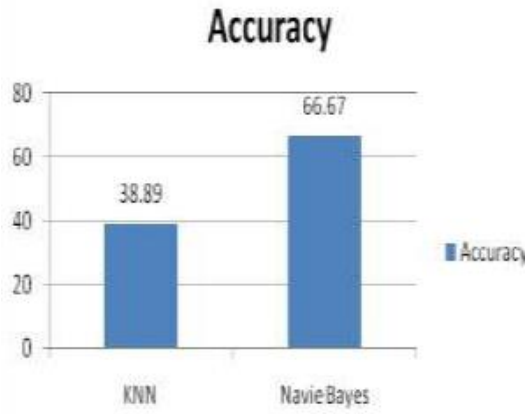


Figure.2.1 Accuracy of KNN and Naive Bayes

Using Twitter with sentiment analysis, Yang *et al.* (2015) attempted to accomplish better output. Naive Bayes classifier used to get high-level precision and also tried to get more data from chi-square. Their primary goal was to achieve high performance precision.

HimankGupta*et al.* (2018) suggested to identify spam tweets on the Twitter platform. Different machine learning algorithms such as Random Forest, Gradient Boosting, Neural Network and Support Vector Machine (SVM) have been used to improve precision.

Detailed assessment of textual emotions was performed and categorized by DianaLupan *et al.* (2012) as fear, rage, surprise, happiness, sadness and disgust. They aimed at capturing the feelings person while reading news. Processing methods for natural language are used to concentrate primarily on sentiments.

Text Mining to evaluate the post and remarks from internet sites like Wikipedia on Facebook by HimanshiAgrawal *et al.* (2016). The proposal was made on non-profit, entertainment websites, public pages and concluded that the spam activity cannot be totally removed from the internet.

Websites for social networking, such as Facebook, are rich in textbooks that allow users to generate different text content in the form of comments, wall posts, social media, and blogs. A huge quantity of information is accessible via the Web due to the invisible use of social network in latest year. Using text mining methods on social networking websites can show important outcomes in relation to interaction behavior between person to person. In addition, text mining methods in combination with social networks can be used in large-scale systems to find general view on any particular topic, patterns of human thinking, and group identification Aggarwal *et al.* (2011).

Liu B*et al.* (2008) has initially implemented sentiment analysis is also known as opinion mining, and subjectivity assessment is the method of determining the attitude or polarity of views or reviews published by individuals to rate products or services. Analysis of sentiment can be applied to any textual type of views such as blogs, reviews and microblogs. Microblogs are those small text messages that cannot enter 149 characters, such as tweets. These microblogs are simpler to analyze than other types of opinion.

ChintanDedhia *et al.* (2017) suggested a model set for the assessment of twitter sentiments. They aimed at improving SVM and Adaboost's efficiency, which was also used to increase efficiency.

Bollen *et al*. (2011) evaluated the random sample feeling of Twitter information, discovering that Dow Jones Industrial Average (DJIA) prices were associated with Twitter 2–3 days previously with 87.6 percent precision.

Kim Hammar *et al.* (2015) evaluated Instagram from the Social Media platform and used qualified information to analyze Instagram messages composed of pictures and texts. Tools for processing of natural language (NPL) are used to classify official text and post into distinct divisions.

HeniSulistiani *et al.* (2017) developed a technique for predicting customer loyalty. They produced a comprehensive feature selection method with comparative analysis. Their goal was to improve the forecast accuracy, efficiency and outcomes. The prediction of customer loyalty based on Random Forest method with comparison of precision.

Vijayarani *et al.* (2016) tried to fix issues by pre-processing methods on Text mining. They found that text mining study would be very helpful in analyzing the content of social media.

Po-Wei Liang *et al.* (2013) analyzed emotion messages by suggesting architecture to create the micro blog post automatically. They discovered messages of positive and negative view. It was also found that the methods of machine learning were successful.

Xia *et al*. (2011) discusses about the sentiment classification ensemble structure that used two kinds of function sets and three basic classifiers. Two kinds of feature sets are produced using Word and Part-of-Speech information. The chosen base classifiers are Naive Bayes, Maximum Entropy and Support Vector Machines. Dissimilar ensemble techniques such as Weighted Combination, Fixed Combination and Meta classifier combination were implemented for classification of sentiments.

Pak *et al*. (2010) developed a twitter corpus using Twitter API to automatically collect tweets and annotate those using emoticons. They constructed a sentiment classifier based on the Naive Bayes multinomial classifier that utilizes N-gram and POS tags as characteristics using that corpus. There is a likelihood of error in this technique since the emotions of tweets in the training set are marked exclusively on the basis of emoticons polarity. The training set is also less effective because it includes only emoticon tweets.

Wan *et al.* (2015) suggested the Airline domain classification scheme based on Ensemble Twitter. Naive Bayes, SVM, Random Forest, Bayesian Network and C4.5 Decision Tree were used as an ensemble for the 5 base classifier algorithm. C4.5 Decision Trees out performs other algorithms in the assignment of classification of sentiments.

The Lexicon-based strategy is the easiest way to classify feelings Cambria *et al.* (2013), since it calculates the sum of the amount of positive sentiment words and the negative sentiment phrases in the text file to evaluate the feeling of the text file. This strategy is poor when it comes to negative.

Tripathy *et al.* (2016) stated that reviews and blog datasets acquired from social networking sites were unsystematic and needed to be classified as significant. Using monitored machine learning techniques, they could be categorized as positive, negative and neutral. They launched four distinct machine learning algorithms to classify feelings. NB (Naive Bayes), ME (maximum entropy), SGD (stochastic gradient descent) and SVM (vector supporting machine) based on precision, recall, F-measure. Their paper helped to classify film reviews using supervised machine learning algorithms that were further implemented using n-gram method to the IMDB dataset. They found that the classification precision reduces in the n-gram strategy as the value of n rises. They also encountered certain constraints on further study as tiny size of twitter posts, reviews or remarks including punctuation symbols and phrases like "greatttt, fineee" because they have no adequate significance. So, after removing the stop words to select the best function, a fresh word list was ready for classification. Hybrid machine learning methods have also been regarded for greater precision

Li *et al.* (2010) used sentiment assessment and text mining methods to study hotspot and forecast online forums. First of all, an algorithm was developed to check the polarity of sentiment for each piece of text. The algorithm was then combined with k-means clustering and support vector machine (SVM) to create an unsupervised text mining strategy. The described text mining method was used to group forums into different clusters whose center represents a hotspot forum in the present period of time. The data sets were taken from the SINA sports

forum. Experimental findings showed that with k-means clustering, SVM forecasting is obtaining highly coherent outcomes. SVM forecasting's top 10 hotspot forums resemble 80 percent of k-means clustering outcomes. The same results were achieved for the top 4 hotspot forums of the year by both SVM and k-means. They developed an algorithm that automatically analyzes a text's feeling polarity, using text values. First, the design of algorithms could be enhanced to produce a more precise feeling calculation. Algorithms other than SVM, or variants of SVM, could also be added for supervised learning. Second, they integrated extraction of the subject. Third, as their significant future work, a practical system in the form of a website portal was requested.

Duwairi *et al.* (2018) stated that the assessment of sentiments determines the polarity of the specified text using either a machine learning strategy or using a lexicon-based method. The classifiers applied to the datasets were Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), where SVM gave the lowest accuracy and KNN gave the lowest recall. The data sets were also tested with 10-fold cross validation. They showed that the accuracy obtained by SVM is 75.25, was the highest accuracy and that the recall received by KNN is 69.04 was the highest recall. Therefore, to achieve better results in classification, larger data sets were needed and crowd sourcing was considered to be followed by semi-supervised learning.

Kouloumpis *et al.* (2011) showed the usefulness of speech characteristics and previous lexical resources used in micro-blogging to detect twitter message sentiments. The scientists found that microblogging characteristics were more helpful than POS characteristics from the current sentiment lexicon. They also found that the training data will be of less advantage, if they include micro-blogging characteristics.

Mudinas *et al.* (2012) found that concept level sentiment analysis (psenti) scheme was better than pure learning-based system and pure lexicon-based system due to more accuracy in well-structured and polarity classification , readable outcomes. They verified on experimentation that the hybrid method was better than sentistrength. They found that the psenti scheme was highly accurate than a pure lexicon-based scheme. It also delivered well-structured, readable outcomes and increased resistance to text style writing. They also found that it works better than sentistrength psenti scheme. In brief, the hybrid method suggested was able to combine a thoroughly constructed lexicon with a strong supervised learning algorithm.

Arun *et al.* (2017) depicted demonetization sentiment analysis on tweets. They first accessed the information and then converted it as an input dataset into text files. After removing the stop words, sentiment analysis was conducted, the polarity of the phrases was determined and the tweets were classified as positive and negative. A new technique for demonetization sentiment analysis was suggested and data cleaning, sentiment scores, bigrams, polarity and graphical methods were used for this process.

Anta *et al.* (2013) provided a comparative study with classification functions of distinct approaches to sentiment analysis of Spanish tweets. Different studies were conducted to classify Spanish tweets according to emotions and subjects. The use of stemmers and lemmatizers, word forms, ngrams, negations, connection processing, valence shifters, search engines, unique Twitter semantics (Hashtags) and various classification techniques were assessed as a thorough and comprehensive research. These results showed that classical techniques could be used to analyze and classify the Spanish text. The best precision seen was 58% for subjects and 42% for classification of sentiments.

Tiwari *et al*. (2017) proved content-based strategy using sentiment analysis for internet audits, film ratings, etc. Supervised machine learning approaches grouped these reviews. Three distinct calculations of machine learning were regarded i.e. for findings. SVM, ME, NB and these findings were based on parameters of accuracy and f-measure. In their paper, various machine learning methods were suggested to classify the film reviews of rotten tomatoes dataset using n gram technique. They also found that their output acquired better precision when compared with other study works.

Analysis of sentiments was used to process natural language with the assistance of which text document polarity was detected by Mesnil *et al.* (2014).Initially only positive and negative feelings were discriminated against, i.e. the issue of binary classification. This issue was matched by various machine learning methods. They used the IMDB dataset to obtain readily reproducible results. Researchers might also be able to combine their advances that would be helpful for further progress. A straightforward and effective technique for analyzing sentiments has been suggested. They had entered three distinct baseline models conceptually: First focused on language models, the second based on successive sentence models and the last based on BOW (Bag of Words) fast reweighing.

The multiple areas of text files were evaluated by Saleh *et al.* (2011) either by SVM or by certain weighting systems. The main aim was to verify the collections and to do so they suggested a distinct collection that would be a useful source for detecting opinion mining. They found that SVM was the best way to manage classification of sentiments. For other assignments, they wished to evaluate the review response and explore Senti-word net-like outermost experiences.

Basha *et al.* (2017) described that an exponential factor was also quickly increasing due to the popularity of e-commerce product reviews for a product. Other people's views would play an significant part in making a choice among various options where time and money were valuable. Now, as part of their studies, most organizations had opinion mining. Nearly every company was also affected by social media pages and blogs that led to sentimental analysis by these businesses. They used fuzzy rule-based (FRBS) schemes with models in their paper, namely: Mamdani, and Takagi Sugeno Kang (TSK) using FRBS package in R. They also contrasted these models with other techniques of classification. Experiments were conducted on the proposed algorithm to calculate the product's emotions and opinions, and the R package was also demonstrated.

Alomari *et al.* (2017) described Arabic tweets as a good chance for opinion mining studies but were reduced owing to a shortage of funds for sentiment analysis or difficulties in the assessment of Arabic language text. They included Arabic twitter corpus that either marked the tweets as positive or negative and they examined the tweets using machine learning. Their primary goal was to examine the Arabic sentiment analysis approach to machine learning. First they farmed Arabic tweets corpus. Then, using separate n-grams with distinct weighting systems and implementing stemming methods, they compared the two machine learning algorithms (SVM and NB).

Lin *et al.* (2009) recognized subjective data using tools and a novel probabilistic modeling structure called a sentiment / topic model that detects feeling and subject from text together. They found that the JST model was completely different from other approaches to machine learning. They suggested this model on the film dataset to classify the polarity of sentiments and enhance the precision of the classification of sentiments. A joint sentiment / topic

(JST) model was suggested in their paper that could be used to depict document level sentiment classification and simultaneously extract combination of subjects from text. On the other side, sentiment classification methods were based on supervised learning, whereas the suggested JST model was completely unattended, resulting in more flexibility and could effectively be merged with other applications. When the scores were contrasted with current supervised methods, they discovered that this model provided competitive efficiency in the classification of document level sentiment. They included bigrams and trigrams in the model.

Wang *et al*. (2012) showed that various solutions of machine learning algorithms demonstrate significant variability in their results for text classification. They also proved that bigram showed steady improvements in assignments and Naive Bayes (NB) was more preferable in sentiment assignments than SVM for a small portion. Also, an alternative SVM demonstrates constantly better on datasets and has shown solutions to NB and SVM. It was found from their paper that Multinomial Naive Bayes (MNB) was more in favor of analyzing the sentiment. They came to the conclusion that NB-SVM was a powerful operator and that Bernoulli Naive Bayes (BNB) was worse than MNB.

NG *et al*. (2006) tested two issues: (1) defining whether the document was a review or not, and (2) classifying the polarity as positive or negative. It has also been shown that the identification of the analysis could be carried out more accurately using only unigrams as characteristics. The role of four kinds of easy linguistic sources in a polarity classification scheme was then examined. The task of analyzing the sentiment of document level was examined. Polarity classification and review identification, two issues were examined in the evaluation of document level sentiment.

Turney *et al.* (2002) describes a unsupervised learning algorithm in order to classify a reviews.They worked on Sentiment Analysis at the document level.The precision for car reviews ranged from 84 percent.

Purver *et al*. (2012) worked primarily on emoticons and hashtags in their studies. In their role, emoticons and hashtags were categorized with 6 basic emotions. Classifications are done with linear kernel and unigram characteristics used by Support Vector Machines (SVM) application. Three individual experiments were conducted in their paper. Experiment 1: detection of emotions, experiment 2: discrimination of emotions and experiment 3: labeling manually. These studies are based on hashtags and emoticons. They got 50-80 percent precision.

Pang *et al.* (2002) used three methods of machine learning to analyze sentiment. They took a general view of the issue of classifying documents and determined whether it was positive or negative. They worked with movie reviews for their experiments. Three standard classifiers were experimented they are: Maximum Entropy classification, Naïve Bayes classification and Support Vector Machine.

Agarawal *et al.* (2011) developed a model for classification functions: a binary task of classifying sentiment into positive, negative and neutral. Support Vector Machines (SVM) was used for all their results.

Hima *et al.* (2016) introduces a novel fuzzy clustering model to analyze twitter sentiments. Then a comparative assessment is produced with the clustering methods of partitioning, such as Expectation Maximization algorithms and K-Means based on metrics. They suggested method is evaluated to be effective in performing results of high-quality twitter sentiments, according to the experimental evaluation.

Koweika *et al.* (2013) discusses about the presented Social Media Analysis. They covered the social media for expressing one's opinion. They predict gender and age based on the phrases and slang used by people's opinion.

ArtiBuche *et al.* (2013) discussed the Opinion Mining and analysis: A Survey. It obviously stated that the Sentiment analysis is a kind of Natural Language Processing that is used to monitor products and web brands. As a result, it has become a requirement for businesses to collect information from different sources such as micro-blogs, review site blogs, and thus determine whether they are regarded positively or negatively using component of speech tagging.

Jose *et al.* (2013) presented an Analysis for Social Media sentiment. The social media coverage has become one of the largest websites for expressing one's opinion. They also said about the positive or negative tweets. The forecast of gender and age can also be made on the basis of the phrases and slang used by human views.

Sentiment Analysis was introduced on Twitter by PreslavNakov *et al.* (2013). Researcher stated that Twitter prevents data on who follows who. Inside tweets, retweets and tags provide data on the discussion. Using contextual phrase-level polarity, views and reviews gathered from Twitter and SMS are categorized into sentiment. The sentiments categorized in three respects such as positive, negative or objective

JishaManjaly *et al.* (2013) suggested a Sentiment Analysis scheme for Subject Identification based on Twitter. It is commonly used for participatory data sharing and cooperation on social media such as blogs, Twitter and Facebook. Then opinions are categorized as positive or negative and neutral, further split into emotional states like sad, happy and angry.

A survey on Opinion Mining and Sentiment Polarity Classification was submitted by Sindhu C *et al.*(2013).Sentiment analysis relates to optimization algorithms for evaluating opinions from different sources, such as blog posts, remarks on forums, product reviews, policies or any social networking sites or tweets. The process of choosing the opinions and avoiding the specific phrase are based on Subjectivity Detection, which is then pre-processed by tokenizing, stopping filtering and stopping phrases. In a sentence they expressed their opinion is classified into 6 emotions: disgust, happiness, anger, fear, sadness, and surprise.

Scheme for the evaluation, acquisition and visualization of Twitter data was presented by Andrei Sechelea *et al.* (2016).Twitter posts are collected and kept in a distributed cluster, and in MapReduce structure the information is analyzed. They also develop a visualization method that allows a geographical location to follow the density of twitter activity.

Chandrasekhar Rangu *et al.* (2017) discusses about text mining referred to as text data processing, as the technique to gather non-trivial patterns and attention-grabbing or information from text files. It utilizes algorithms to remodel (unstructured) free stream text into information that will be analyzed using applied mathematics, language process and machine learning techniques. Text mining is emerging technology that allows companies to understand their clients well and make it easier for them to redefine customer wishes. They attempt to obtain some useful data from asset reviews that will be used to improve engineering characteristics of assets and help to improve the quality of service and client experience.

Bhadane *et al*. (2013) discusses about the sentiment analysis research focuses on the different methods used to classify a given set of natural language text according to the opinions i.e. whether the general opinion is negative or positive. They used a number of methods for

classification and product review polarity identification using Machine Learning (ML) in combination with field-specific lexicons. Their experimental results demonstrate that approximately 78 percent precision has been attained by the suggested methods and that they are successful in the activities.

Rincy Jose *et al.* (2015) describes sentiment analysis schemes using the web reviews and social media information bag-of-words approach to mining the sentiments. Instead of looking at the whole sentence / paragraph for evaluation, the bag-of-words strategy checks only single words and their characteristics as vectors. It may mislead the algorithm of classification, particularly when used for issues such as classification of sentiments. The machine learning algorithms such as Maximum Entropy, Naive Bayes, Support Vector Machine (SVM) etc, are commonly used to fix issues with classification. Experiments show that the semantic-based characteristic vector with ensemble classifies the bag-of-words strategy by 3-5 percent with a single machine learning classifier.

Aldo Hernández *et al*. (2016) introduce a Twitter content sentiment analysis technique for predicting future internet attacks. The process is focused on collecting tweets from two user sets on a daily basis; individuals who use the platform as a way of expressing opinions on appropriate problems and those who use it to show the content identified with web security threats. Regular information is transformed into data that can be statistically broken down to predict if an attack is feasible. In reaction to a worldwide case, they finally completed by researching the aggregate feeling of users and organizations of hacking activists.

kirtihuda *et al.* (2017) they discusses about the sentiment analysis techniques. The data is given as input in the first step, and the twitter data is the input data it can be either in excel sheet

or it can be extracted in real time. During the pre-processing stage, the data provided as input is pre-processed in which the information is tokenized and the phrases stop from the information are removed. The classification method was introduced to the feature extraction information for the sentiment analysis in the last phase. SVM classifier is used for data analysis in this role. The execution time of the proposed and existing algorithm, are shown in figure 2.2. It is evaluated that the improved algorithm based on patterns tasks less time for execution.

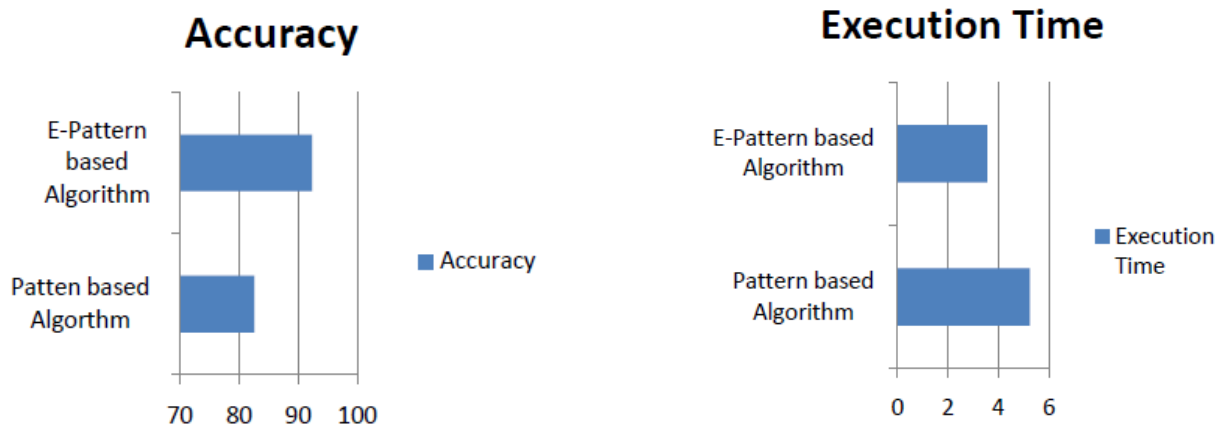


Figure.2.2 Accuracy Comparison and the Execution Time

Anurag P. Jain *et al.* (2015) presented a paper to examine the sentiments of consumers using data mining classifiers. It also compares the output of single classifiers for assessment of sentiments over the classifier ensemble. Acquired experimental result shows that elevated predictive precision is provided by the k-nearest neighbor classifier. Additionally, Random Forest provides excellent precision in prediction.

Xiaolong Wang *et al.* (2011) describe the investigation of illustrations that three kinds of data are helpful for automatically generating the general sentiment polarity for a specified hash tag over a certain period of time. They demonstrate that the efficiency can be significantly

improved by using an enhanced classification boosting environment in which they have used the literal significance of hash tags as semi-supervised data.

AndriusMudinas *et al.* (2012) describe the outcomes of sentiment analysis generated by their hybrid strategy in a favorable manner compared to the baselines of lexicon and learning only. Their psenti scheme accomplishes high precision, quite similar to the simple learning scheme and much greater than the pure lexicon-based scheme, for both sentiment polarity classification and sentiment strength detection. This strategy combines the strength as well as readability from a constructed lexicon with the best of two words, and the elevated precision from a strong supervised learning algorithm.

Dhanushka *et al.* (2013) discusses about the classification of sentiment applying sentiment classifier trained results in poor performance because words that occur in the train (source) domain might not appear in the test (target) domain. To overcome the feature mismatch problem in cross-domain sentiment classification proposed a cross-domain sentiment classifier using an automatically extracted sentiment sensitive thesaurus. They have done the comparisons against the SentiWordNet, a lexical resource for word polarity. They show that the created sentiment-sensitive thesaurus accurately captures words that express similar sentiments.

Shulong Tan *et al.* (2014) observed that previous research mainly focused on modeling and tracking public sentiment so, they moved one step further to interpret sentiment variations. They worked on twitter dataset. They observed that emerging topics (named foreground topics) within the sentiment variation periods are highly related to the actual reasons behind the variations. These foreground topics can help to interpret the sentiment variations .Based on this observation, they proposed a Latent Dirichlet Allocation (LDA) based model, Foreground and Background LDA (FB-LDA), to dig out foreground topics and filter out background topics. To further improve the

readability of the mined reasons, they ranked them with respect to their "popularity" within the variation period using Reason Candidate and Background LDA (RCB-LDA) method.

Goncalves *et al.* (2013) compare eight sentiment analysis methods in terms of coverage and agreement coverage measures the percentage of messages whose sentiment is identified; and agreement refers to the portion of identified sentiment that comply with ground truth. The major findings of are listed as follows:

- Coverage varies from 4% to 98% when sentiment analysis was applied to six real-world topics including Harry-Potter, Air France, H1N1, Susan Boyle, 2008 Olympic sand 2008 US-Election. Only a small fraction of texts could be accurately analyzed because of bias or insufficient amount of data.
- No sentiment analysis method achieved both high agreement and high coverage simultaneously. For example, the highest coverage appears in Emoticons method which is above 85%, however it also was very low agreement between 4% and 13%
- The agreement of existing sentiment analysis methods varies widely according to the predicted polarity from33% to 80%. That is, the same text could be interpreted differently if we change the analysis method
- Most sentiment analysis methods perform badly on negative texts.

Thelwalk *et al.* (2012) discusses about the performance of the unsupervised version of the algorithm SentiStrength is evaluated. The system combines several opinion lexicons of words and idioms. The system also incorporates a straight-forward method for detecting the scope of negation uses that does not invert the polarity of a word, but makes the word neutral. The authors conclude that their lexicon-based proposal is suitable for polarity classification in Twitter with

the supervised polarity classification system for English tweets is described. The system follows the same approach for defining negation that is, all subsequent words to a negative particle are considered as negated words.

Mejova *et al.* (2009) in his research work proposed that we can use presence of each character, frequency of occurrences of each character, word which is considered as negation etc. as features for creating feature vector. He also shows that we can effectively use unigram and bigram approaches to make feature vector in Sentiment analysis.

Domingos *et al*. (2009) suggested that Naive Bayes works well for dependent features for certain problem. This model is based on Bayesian algorithm. In this model, some efficient approaches are used for selecting feature, computation of weight and classification.

Celikyilmaz *et al.* (2010) developed one method as pronunciation based word clustering. This method normalizes noisy tweets. There are some words which have the same pronunciation but having different meanings. So, for eliminating this conflict, there is method mentioned above. In this mentioned method, words having same pronunciation are clustered and assigned common tokens.

Andrius *et al.* (2012) discusses about the concentrated on engineering students .Twitter posts to understand problems and glitches in their educational experiences. They first conducted a qualitative analysis on samples taken from about 25,000 tweets associated to engineering students' college life. They found engineering students encounter problems such as deficiency of social engagement, heavy study load, and sleep deficiency. Based on these results, we implemented a multi-label classification algorithm to categorize tweets reflecting students' problems. They then used the algorithm to train a detector of student problems from about

35,000 tweets issued at the geo-location of Purdue University. This work, presents an approach and results that show how casual social media data can provide insights into students' experiences.

2.3 Chapter Summary

This chapter presented a literature survey related to the research topic. From the literature study conducted, it can be understood that many researcher have spent time for various techniques and methods that are used for the sentiments of public opinion and some of the algorithms they used are Support Vector Machine (SVM), Machine Learning, Naive Bayes, Multinomial Naive Bayes (MNB) etc.

CHAPTER 3

RESEARCH METHODOLOGY

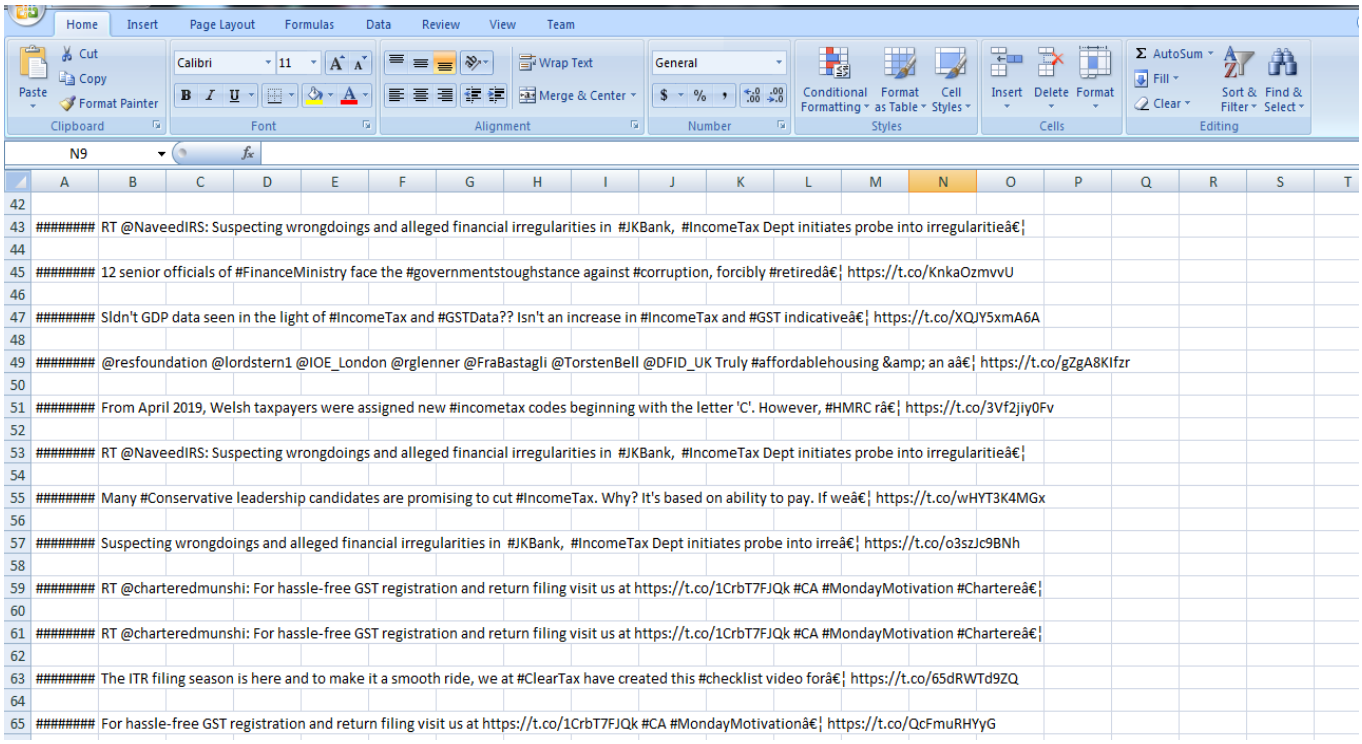
3.1 Introduction

In this research work, the processing of text mining using the DNN technique with feature selection classification is proposed. Feature selection is used to select the appropriate data to set the feature to improve DNN's efficiency of the classification and text mining is used to create a classification of positive, negative and neutral about the tax issues are focused on the public opinion from the tweets. In general, the stages of this research consist of defining tax problems, gathering social media information, text processing, feature selection, and classification using the DNN and testing stage to evaluate algorithm efficiency and show the performance of text mining. Identifying the issues is the initial step in recognizing the main issues, tax process, & transactions to assess which portion or emotional response of tax is intended for public awareness and the emotional response are based on the opinion of public tax tweets.

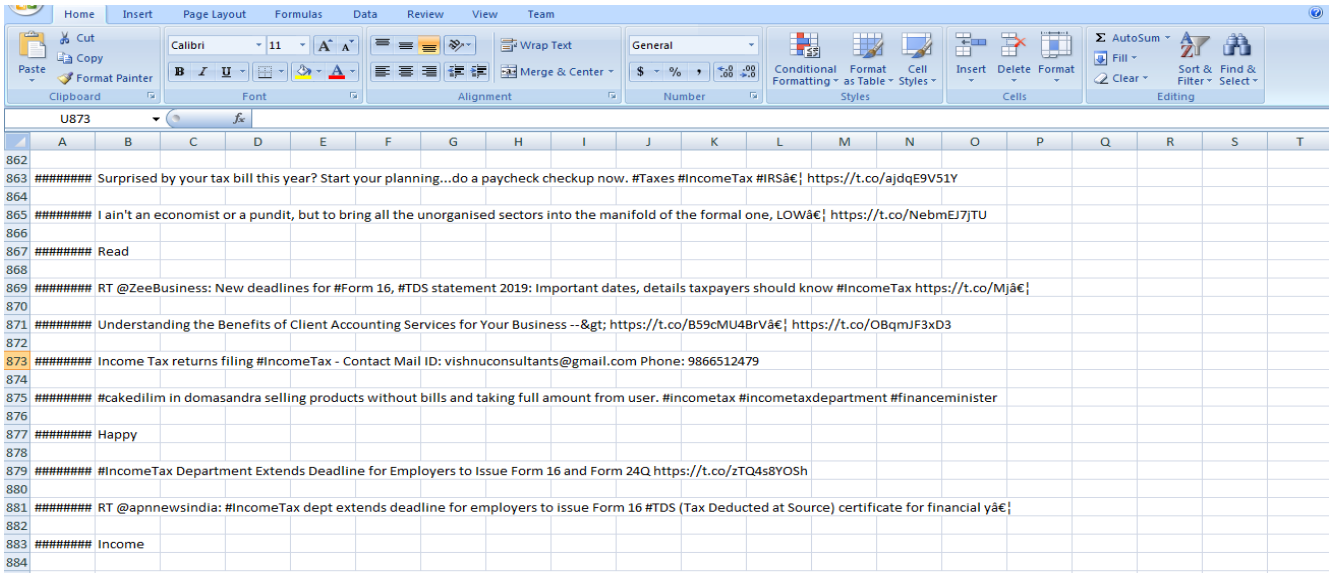
Text mining is a text analysis where data sources are usually obtained from documents with the aim of searching for words that can represent the contents of a document so that interrelationships and inter-document classes can be analyzed. Using this social media user encourage to host their opinion while leads to increase in unlimited textual information so there is a need for system to utilize the textual data with reducing the value of information. Text mining is a powerful tool that enables to analyze to explore the opinion quickly and efficiently. This research work analysis how text mining can be adopted for effective analysis of opinion for polarities.

Dataset used for this research work was taken from UCI repository ie. 2600 tax tweets were considered for this research work and the text processing techniques as well as the two classifiers were applied, and in order to classify the sentiment based on the public opinion and the sample datasets are shown in the figure 3.1.

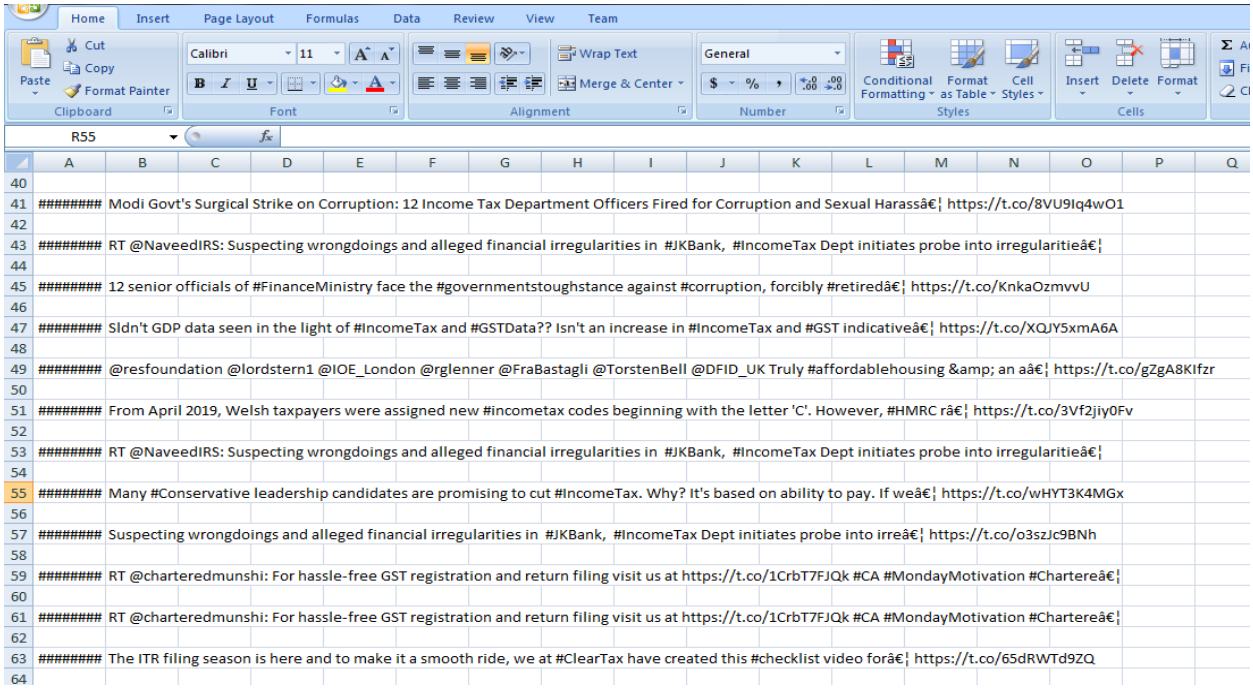
Sample Datasets



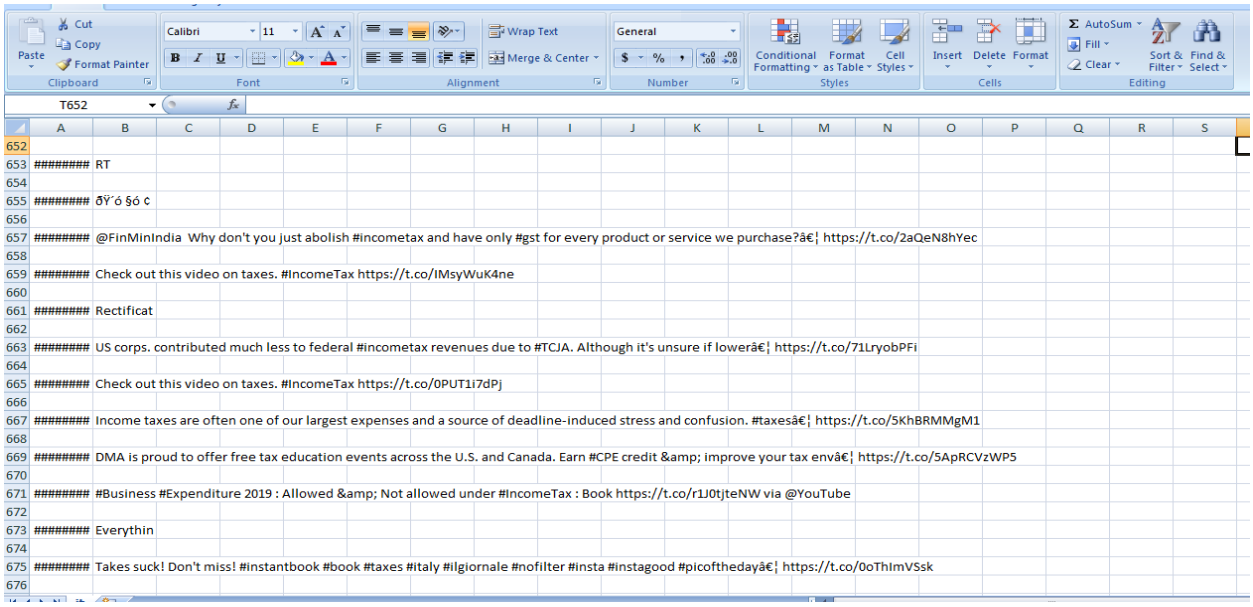
	A
42	
43	######## RT @NaveedIRS: Suspecting wrongdoings and alleged financial irregularities in #JKBank, #IncomeTax Dept initiates probe into irregularitieâ€¦
44	
45	######## 12 senior officials of #FinanceMinistry face the #governmentstoughstance against #corruption, forcibly #retiredâ€¦ https://t.co/KnkaOzmvvU
46	
47	######## SIdn't GDP data seen in the light of #IncomeTax and #GSTData?? Isn't an increase in #IncomeTax and #GST indicativeâ€¦ https://t.co/XQJY5xmA6A
48	
49	######## @resfoundation @lordstern1 @IOE_London @rglenner @FraBastagli @TorstenBell @DFID_UK Truly #affordablehousing & an aâ€¦ https://t.co/gZgA8KIfzr
50	
51	######## From April 2019, Welsh taxpayers were assigned new #incometax codes beginning with the letter 'C'. However, #HMRC râ€¦ https://t.co/3Vf2jiy0Fv
52	
53	######## RT @NaveedIRS: Suspecting wrongdoings and alleged financial irregularities in #JKBank, #IncomeTax Dept initiates probe into irregularitieâ€¦
54	
55	######## Many #Conservative leadership candidates are promising to cut #IncomeTax. Why? It's based on ability to pay. If weâ€¦ https://t.co/wHYT3K4MGx
56	
57	######## Suspecting wrongdoings and alleged financial irregularities in #JKBank, #IncomeTax Dept initiates probe into irreâ€¦ https://t.co/o3szJc9BNh
58	
59	######## RT @charteredmunshi: For hassle-free GST registration and return filing visit us at https://t.co/1CrbT7FJQk #CA #MondayMotivation #Chartereâ€¦
60	
61	######## RT @charteredmunshi: For hassle-free GST registration and return filing visit us at https://t.co/1CrbT7FJQk #CA #MondayMotivation #Chartereâ€¦
62	
63	######## The ITR filing season is here and to make it a smooth ride, we at #ClearTax have created this #checklist video forâ€¦ https://t.co/65dRWTd9ZQ
64	
65	######## For hassle-free GST registration and return filing visit us at https://t.co/1CrbT7FJQk #CA #MondayMotivationâ€¦ https://t.co/QcFmuRHYyG



	A
862	
863	######## Surprised by your tax bill this year? Start your planning...do a paycheck checkup now. #Taxes #IncomeTax #IRSâ€¦ https://t.co/ajdqE9V51Y
864	
865	######## I ain't an economist or a pundit, but to bring all the unorganised sectors into the manifold of the formal one, LOWâ€¦ https://t.co/NebmEJ7jTU
866	
867	######## Read
868	
869	######## RT @ZeeBusiness: New deadlines for #Form 16, #TDS statement 2019: Important dates, details taxpayers should know #IncomeTax https://t.co/Mjâ€¦
870	
871	######## Understanding the Benefits of Client Accounting Services for Your Business --> https://t.co/B59cMU4BrVâ€¦ https://t.co/OBqmJF3xD3
872	
873	######## Income Tax returns filing #IncomeTax - Contact Mail ID: vishnuconsultants@gmail.com Phone: 9866512479
874	
875	######## #cakedilim in domasandra selling products without bills and taking full amount from user. #incometax #incometaxdepartment #financeminister
876	
877	######## Happy
878	
879	######## #IncomeTax Department Extends Deadline for Employers to Issue Form 16 and Form 24Q https://t.co/zTQ4s8YOSh
880	
881	######## RT @apnnewsindia: #IncomeTax dept extends deadline for employers to issue Form 16 #TDS (Tax Deducted at Source) certificate for financial yâ€¦
882	
883	######## Income
884	



40		
41	########	Modi Govt's Surgical Strike on Corruption: 12 Income Tax Department Officers Fired for Corruption and Sexual Harassâ€¦ https://t.co/8VU9Iq4wO1
42		
43	########	RT @NaveedIRS: Suspecting wrongdoings and alleged financial irregularities in #JKBank, #IncomeTax Dept initiates probe into irregularitieâ€¦
44		
45	########	12 senior officials of #FinanceMinistry face the #governmentstoughstance against #corruption, forcibly #retiredâ€¦ https://t.co/KnkaOzmvvU
46		
47	########	Sldn't GDP data seen in the light of #IncomeTax and #GSTData?? Isn't an increase in #IncomeTax and #GST indicativeâ€¦ https://t.co/XQJY5xmA6A
48		
49	########	@resfoundation @lordstern1 @IOE_London @rglenner @FraBastagli @TorstenBell @DFID_UK Truly #affordablehousing & an aâ€¦ https://t.co/gZgA8KIfzr
50		
51	########	From April 2019, Welsh taxpayers were assigned new #incometax codes beginning with the letter 'C'. However, #HMRC râ€¦ https://t.co/3Vf2jiy0Fv
52		
53	########	RT @NaveedIRS: Suspecting wrongdoings and alleged financial irregularities in #JKBank, #IncomeTax Dept initiates probe into irregularitieâ€¦
54		
55	########	Many #Conservative leadership candidates are promising to cut #IncomeTax. Why? It's based on ability to pay. If weâ€¦ https://t.co/wHYT3K4MGx
56		
57	########	Suspecting wrongdoings and alleged financial irregularities in #JKBank, #IncomeTax Dept initiates probe into irreâ€¦ https://t.co/o3szJc9BNh
58		
59	########	RT @charteredmunshi: For hassle-free GST registration and return filing visit us at https://t.co/1CrbT7FJQk #CA #MondayMotivation #Chartereâ€¦
60		
61	########	RT @charteredmunshi: For hassle-free GST registration and return filing visit us at https://t.co/1CrbT7FJQk #CA #MondayMotivation #Chartereâ€¦
62		
63	########	The ITR filing season is here and to make it a smooth ride, we at #ClearTax have created this #checklist video forâ€¦ https://t.co/65dRWTd9ZQ
64		

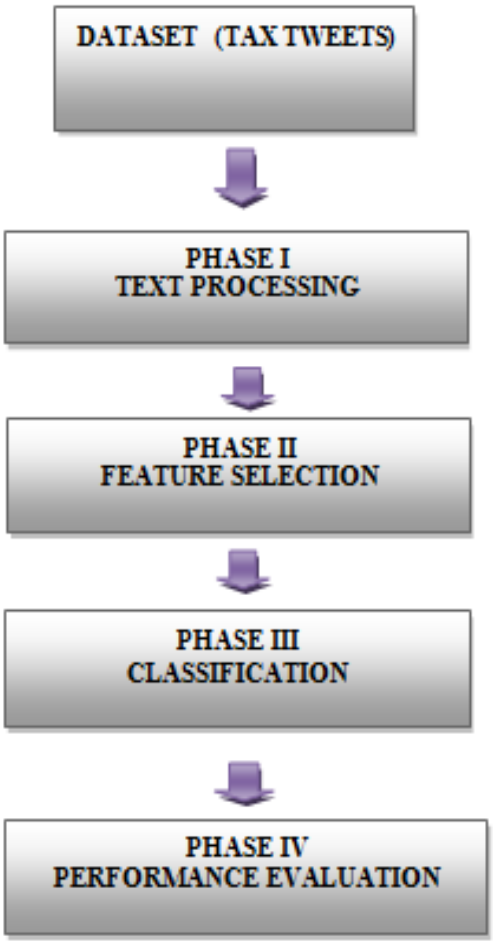


652		
653	########	RT
654		
655	########	ðŸ´ó §ó ¢
656		
657	########	@FinMinIndia Why don't you just abolish #incometax and have only #gst for every product or service we purchase?â€¦ https://t.co/2aQeN8hYec
658		
659	########	Check out this video on taxes. #IncomeTax https://t.co/IMsyWuK4ne
660		
661	########	Rectificat
662		
663	########	US corps. contributed much less to federal #incometax revenues due to #TCJA. Although it's unsure if lowerâ€¦ https://t.co/71LryobPFi
664		
665	########	Check out this video on taxes. #IncomeTax https://t.co/0PUT1i7dPj
666		
667	########	Income taxes are often one of our largest expenses and a source of deadline-induced stress and confusion. #taxesâ€¦ https://t.co/5KhBRMMgM1
668		
669	########	DMA is proud to offer free tax education events across the U.S. and Canada. Earn #CPE credit & improve your tax envâ€¦ https://t.co/5ApRCVzWP5
670		
671	########	#Business #Expenditure 2019 : Allowed & Not allowed under #IncomeTax : Book https://t.co/r1J0tjteNW via @YouTube
672		
673	########	Everythin
674		
675	########	Takes suck! Don't miss! #instantbook #book #taxes #italy #ilgiornale #nofilter #insta #instagood #picofthedayâ€¦ https://t.co/0oThImVSsk
676		

Figure 3.1. Sample Datasets

3.2 Methodology

A research methodology or involves specific techniques that are adopted in research process to collect, assemble and evaluate data. It defines those tools that are used to gather relevant information in a specific research study. Surveys, questionnaires and interviews are the common tools of research. The various phases of the research design is presented in figure 3.2 and are discussed below.



.

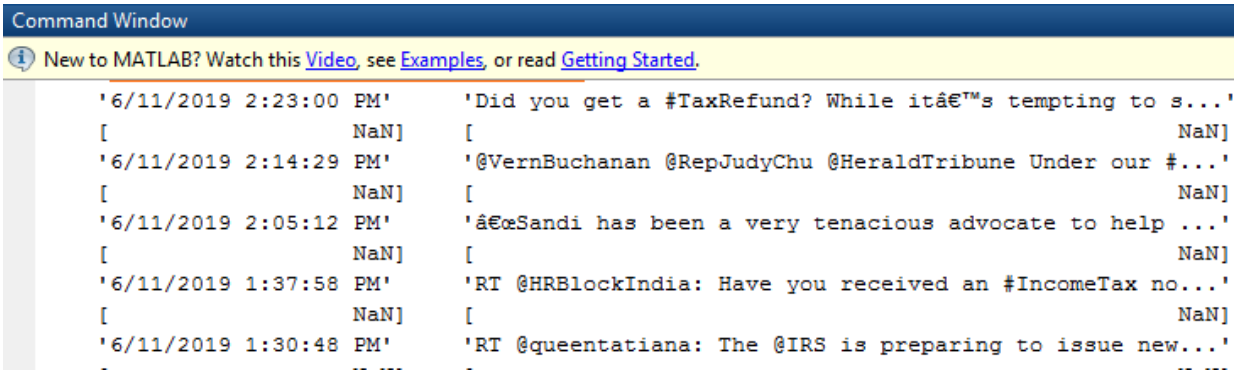
Figure.3.2 Research Design

3.3 Phase 1 : Text Processing

Text processing is the main phase of processing technique for text and at this point, and the research design creates the input text for the process. To support the classification based on sentiment and tax tweets, as well as the structure of words acquired from social media remarks is regulated and the steps involved are shown in the figure.3.5.

a) Case Folding : Case Folding is described as a method applied to a series of tweets characters in which their upper case equivalents replace those recognized as non-uppercase and their results was listed below in Figure 3.3.

Original Dataset



Case folding

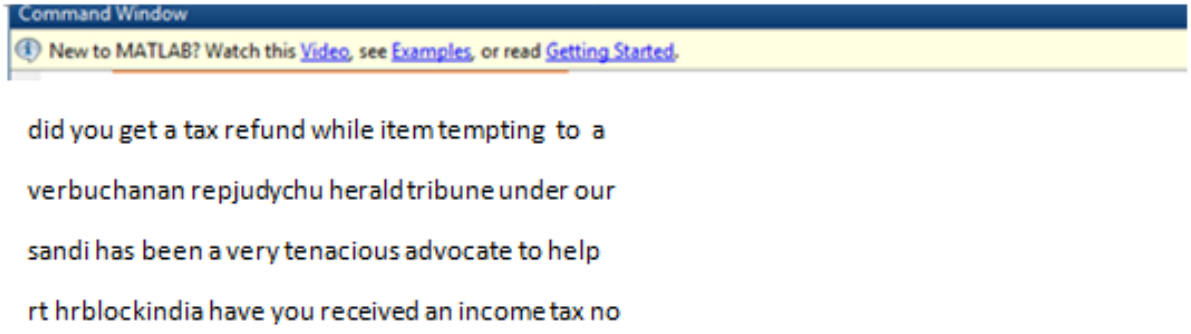


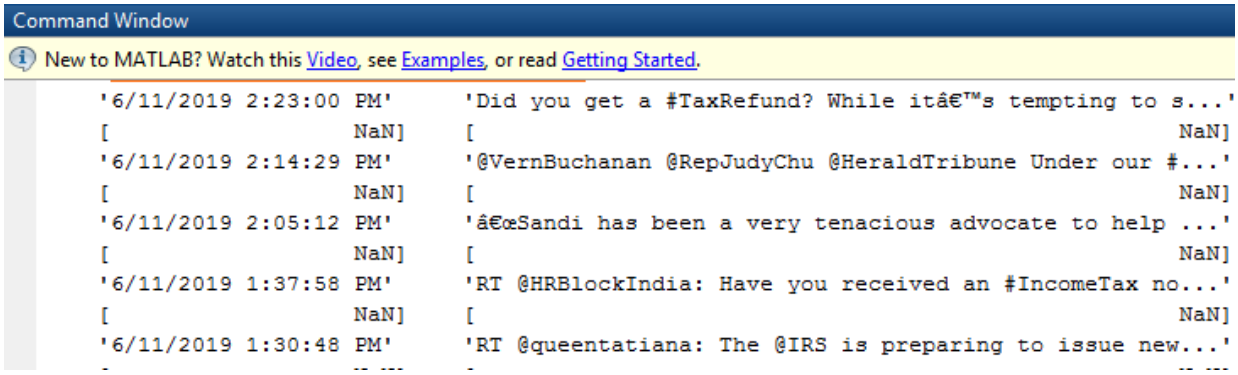
Figure. 3.3 Case Folding

b) Tokenizing: Tokenizing is the method of replacing the vulnerable tweets with distinctive symbols of identification that maintain all the important text information without compromising the data.

c) Stemming : Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words. Stemming is important in Natural understanding (NLU) and Natural Language Processing (NLP).

d) Filtering: Filtering is the process of deleting the irrelevant tweet text which means it is not related to the subject and their results was shown in the figure 3.4

Original Dataset



Filtering

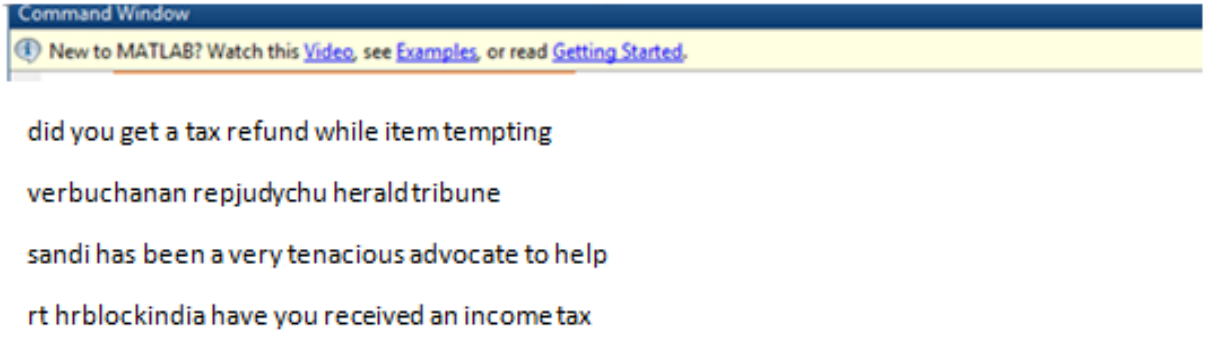


Figure. 3.4 Filtering

e) Convert Emoticons: Convert Emoticons will automatically converts the smiley faces into the text.

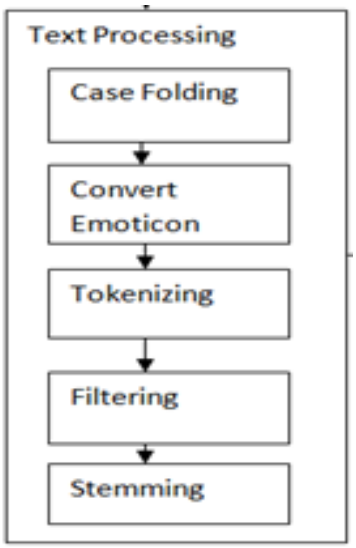


Figure. 3.5 Text Processing

3.4 Phase 2 : Feature Selection

Feature selection is the main phase in the text mining with sentiment analysis based on tax. It improves the classification task effectively. Features that really discriminate between groups will improve the performance of identification. If any discriminating function is not included in the classification, then the classification accuracy will be reduced and the overall methodology is presented in following figure.3.6.

3.5 Phase 3 : Classification

In this research work classification is the third phase; SVM and Recurrent Neural Network algorithm have been used for classification. Support Vector Machine (SVM) and Recurrent Neural Network (RNN) were also compared based on the accuracy.

Liu *et al.* (2012) discusses about the Sentiment classification methods that are usually split into three main categories called the Lexicon Based Approach (LB), Machine Learning

Approach (ML), and the Hybrid Approach. The Machine Learning Approach (ML) utilizes language characteristics and well-known ML algorithms. The Lexicon-based method is guided by a lexicon of view that is nothing more than a set of pre-compiled terms of opinion.

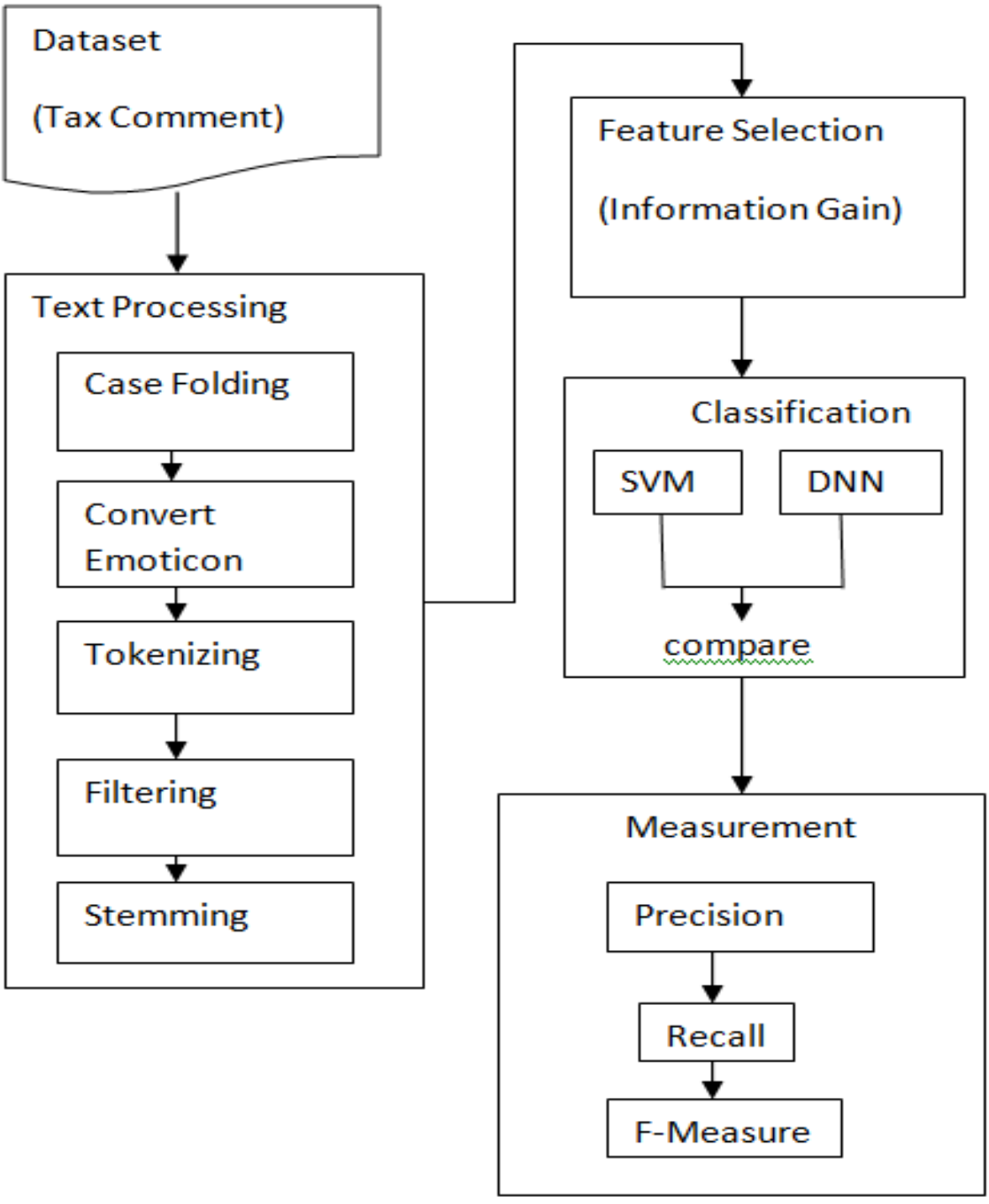


Figure .3.6 Overall Methodology

3.5.1 Support Vector Machine (SVM)

In machine learning, Support Vector Machines is supervised learning with related learning algorithms that analyze the information used to evaluate classification.

Abdelwahab *et al*. (2015) discusses about the evaluation of sentiments about such a specific subject or product, text processing is efficiently used to understand the reviews or suggestions of different public view. Research on text processing techniques for sentiment analysis is repeatedly performed to improve the efficiency of text mining in acquiring the suitable views as in research that assess the impact of information size training using SVM. The study indicates that, by combining SVM and Naive Bayes using AND-type fusion, the transition in training set does not significantly impact the rate of classification precision using SVM or Naïve Bayes, but indicates enhanced accuracy and SVM F-Score and the graph were shown in the figure.3.7.

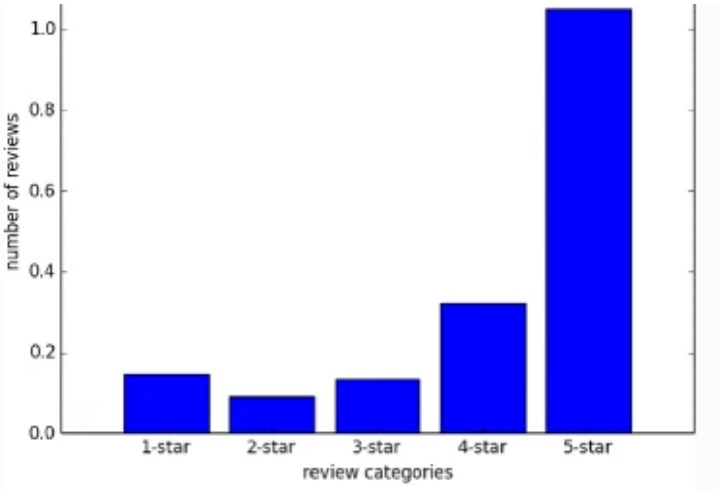


Figure. 3.7 Product Review

Sheshasaayee *et al.* (2017) in their studies they have used the data mining to discover the beneficial data from the World Wide Web (WWW) repository, while using text, video

pictures, and audio on extracting unstructured data, it is split into Content Mining, Usage Mining and Structure Mining. The study compared the accuracy of Random Forest, Naive Bayes, and SVM algorithms using the Machine Learning strategy.

3.5.2 Recurrent Neural Network (RNN) for Classification

Recurrent Neural Networks (RNN) is a popular architecture of deep learning for analyzing sentiments in sentence and it is widely applied for classification of texts. The main advantage of RNN is that they can be used to model variable-length temporal sequences, providing additional flexibility in the analysis of reviews of distinct lengths and the comparison of SVM and DNN is depicted in the below figure.3.8

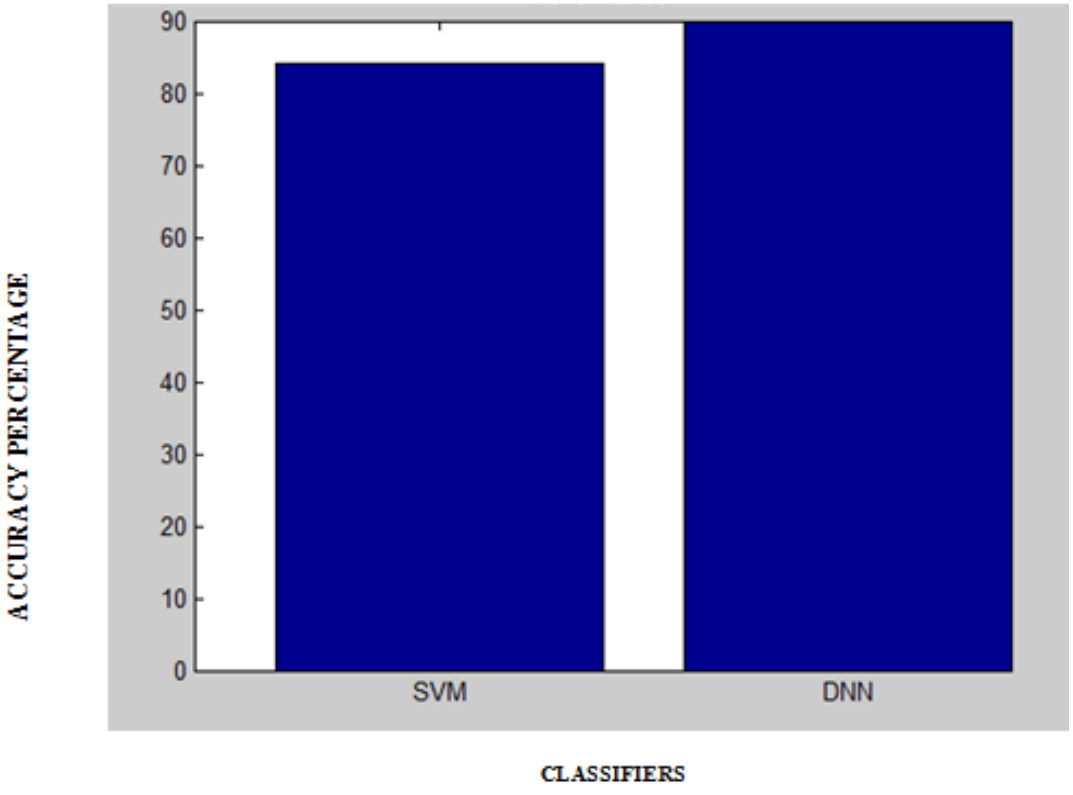


Figure.3.8 Comparison of SVM and DNN

3.6 Phase 4 : Performance Evaluation

Text classification rules were usually assessed to use performance measures from the retrieved data. Common metrics for evaluating text categorization include precision, recall, and accuracy and their flow is shown in the figure .3.9

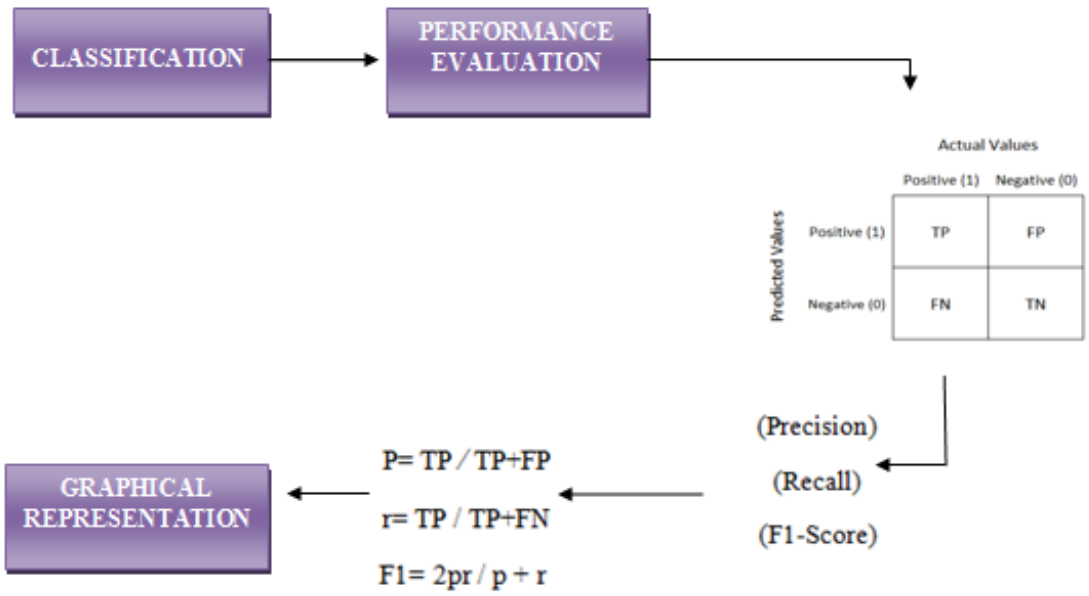


Figure .3.9 Performance Evaluation

3.6.1. Performance Metrics

The performance metrics were calculated by using True positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

True Positive is an outcome where the model correctly predicts the positive class. True negative is an outcome where the model incorrectly predicts the positive class. False positive is

an outcome where the model incorrectly predicts the positive class. False negative is an outcome where the model incorrectly predicts the negative class.

- ✓ Precision is calculated using the equation 3.1

  Equation 3.1 TP / TP + FP

- ✓ Recall is performed using equation 3.2

  Equation 3.2 TP / TP + FN

- ✓ F-measure is calculated using the equation 3.3

  Equation 3.3 2pr / p + r

3.7 Chapter Summary

In this chapter the dataset and the research methodology have been discussed. The phases included in the research work such as text processing, feature selection, classification, and performance evaluation with the help of Support Vector Machine (SVM) and Deep Learning Neural Network (DNN) are also been discussed. The result obtained in this research work is discussed in the next Chapter 4.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

In this research work, text processing with techniques such as crawling, case folding, tokenizing, convert emoticons, filtering and stemming. Text mining is used for classification of the positive, negative and neutral about the tax issues that are focused on the public opinion. In general, the stages of this research consist of defining tax problems, gathering social media information, text processing, feature selection, and classification using the DNN and testing stage to evaluate algorithm efficiency and show the performance of text mining. Feature Selection was done using Information gain (IG) and classification were performed by using Recurrent Neural Network algorithm, and its performance metrics has been evaluated in terms of precision, recall, f-measure and accuracy compared with SVM and DNN.

In this research work, based on public opinion (tweets) the system is used to classify the polarity. Datasets are taken from Twitter and the comparison of an algorithm is represented in the figure. An epoch is a hyperparameter defined before a model is trained. One epoch is when a whole dataset is passed through the neural network only once, both forward and backward. One epoch is too large to simultaneously feed the computer. So, we're dividing it into a few smaller lots. We use more than 1 epoch because it is not enough to pass the whole dataset through a neural network and we need to pass the complete dataset to the same neural network with multiple times. Mean Square Error (MSE) is used to calculate network performance. The outputs are shown in the Fig 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, and 4.8.

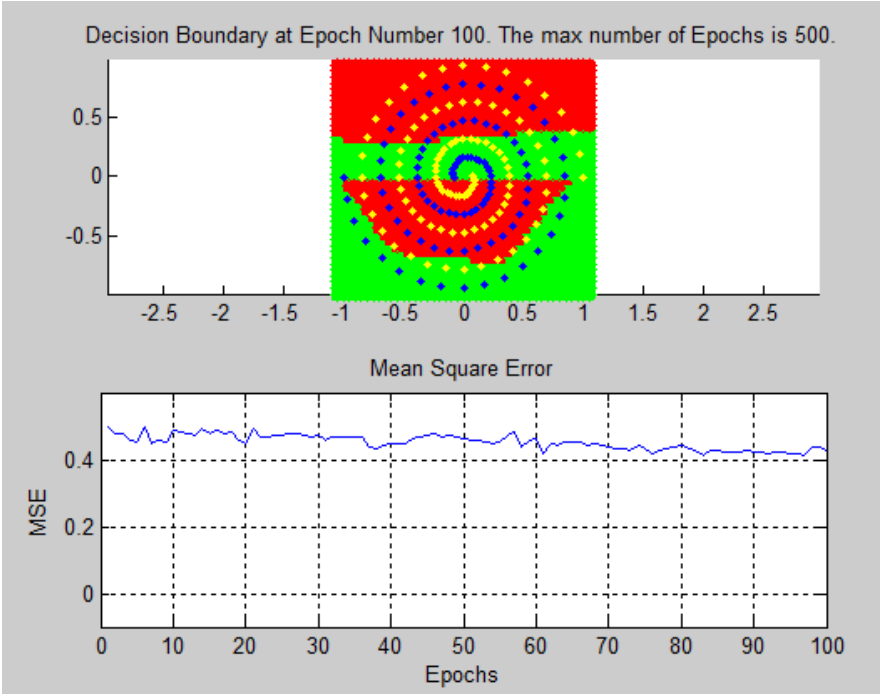


Figure.4.1 Result of Epoch Number is 100

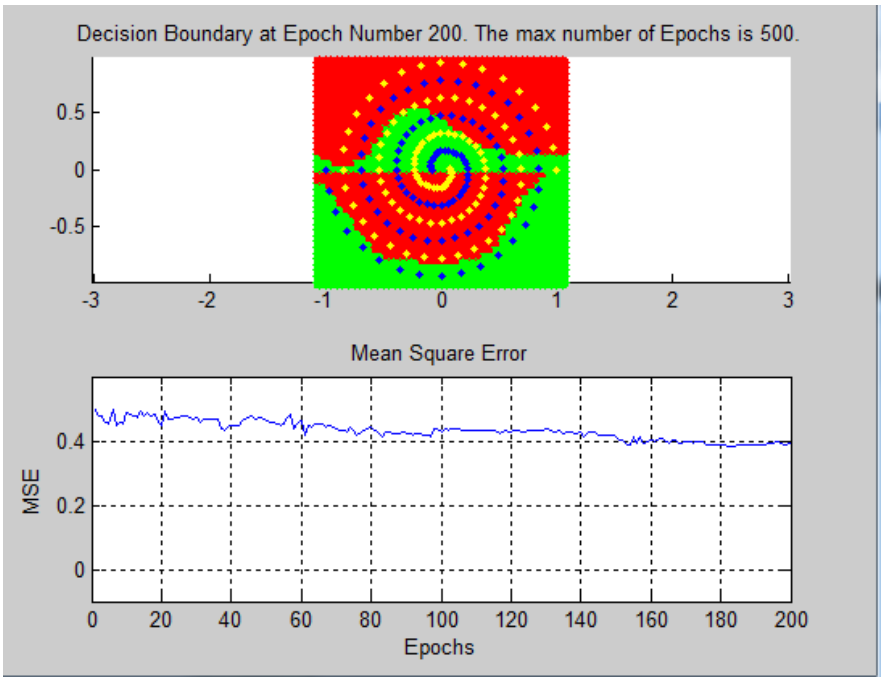


Figure.4.2 Result of Epoch Number is200

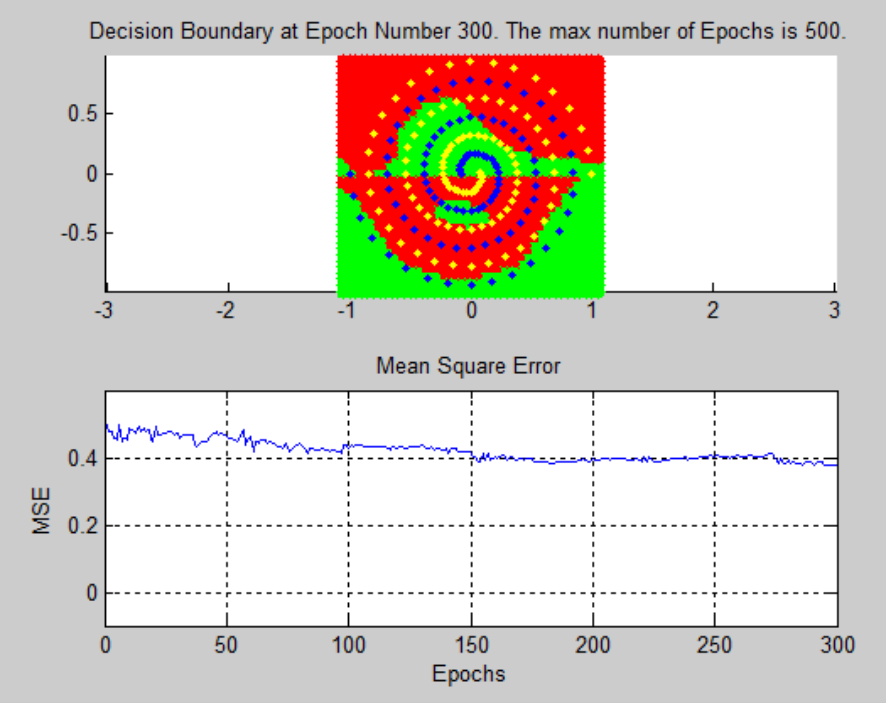


Figure.4.3 Result of Epoch Number is 300

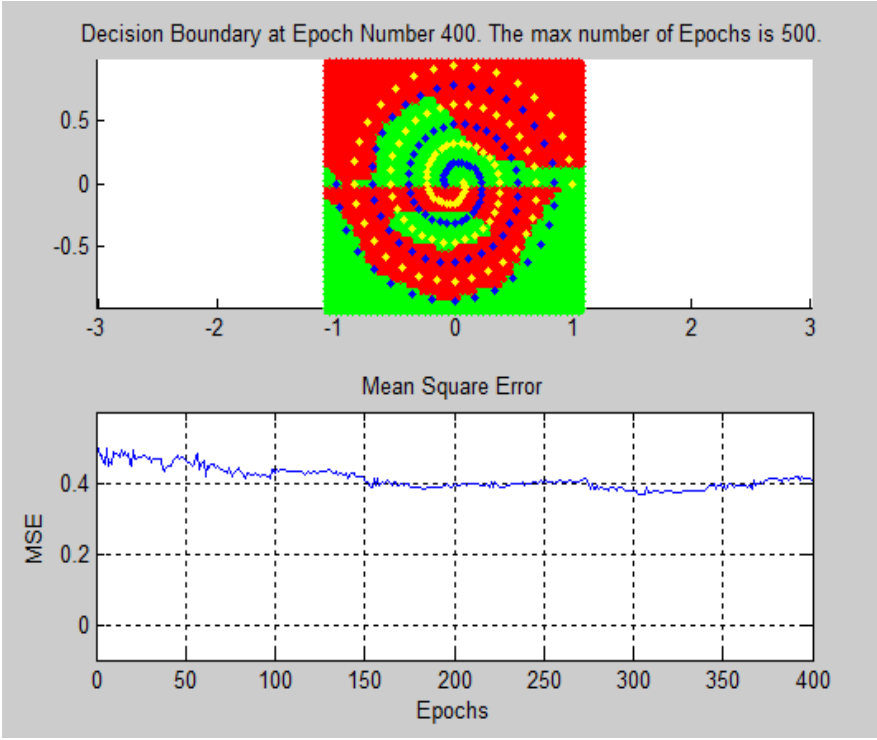


Figure.4.4 Results of Epoch Number is 400

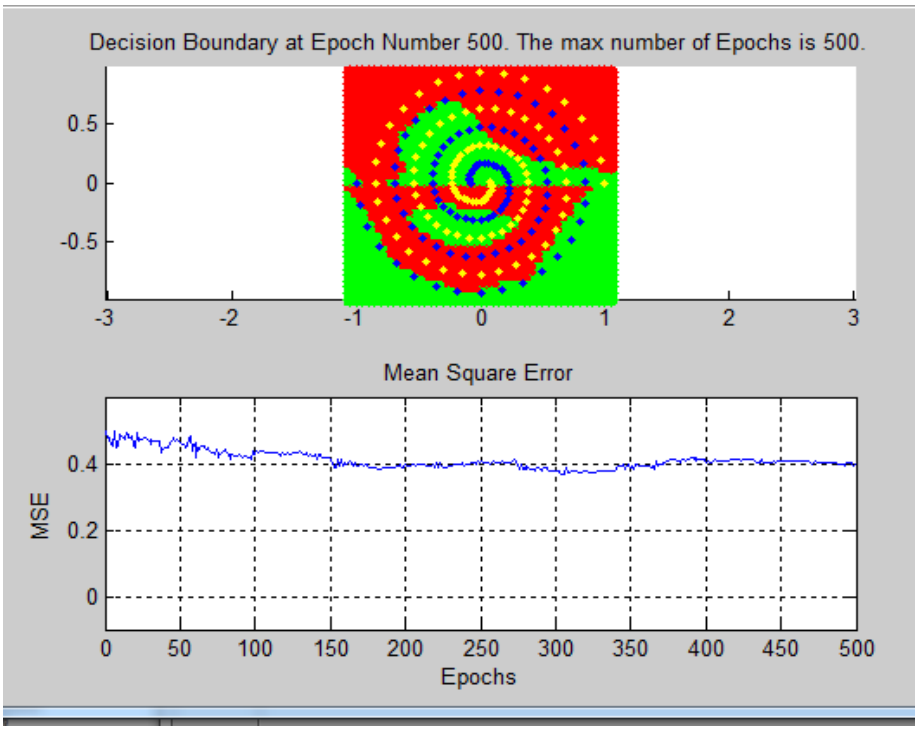
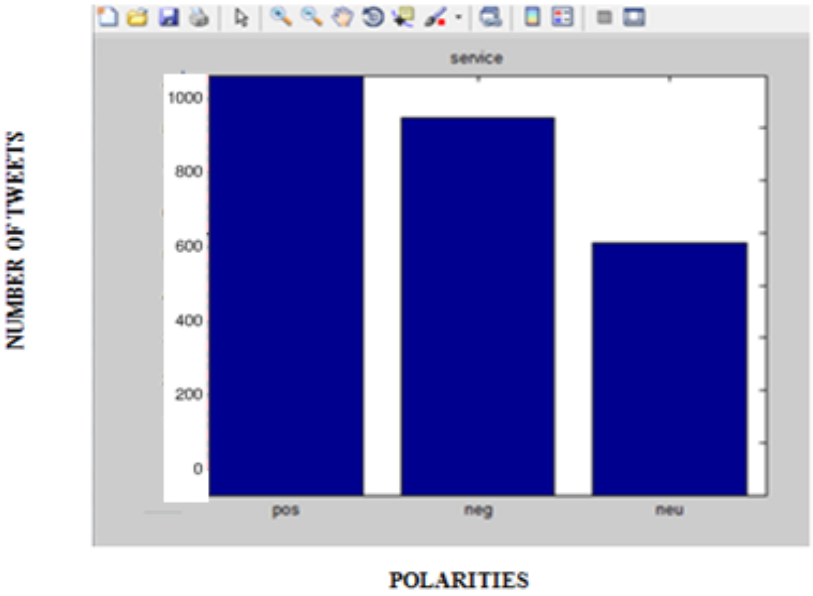


Figure.4.5 Result of Epoch Number is 500

METHODS / METRICS	SV M	DNN
PRECISION	89.70 %	90.30 %
RECALL	70.00 %	82.10 %
F-MEASURE	80.10 %	89.70 %

Figure.4.6 Performance Metrics of SVM and DNN

From this table it is evident that DNN performs well when compared with SVM.



Number of Tweets: 2600

Positive : 1150

Negative : 850

Neutral : 600

Figure.4.7 Polarities on Tweets

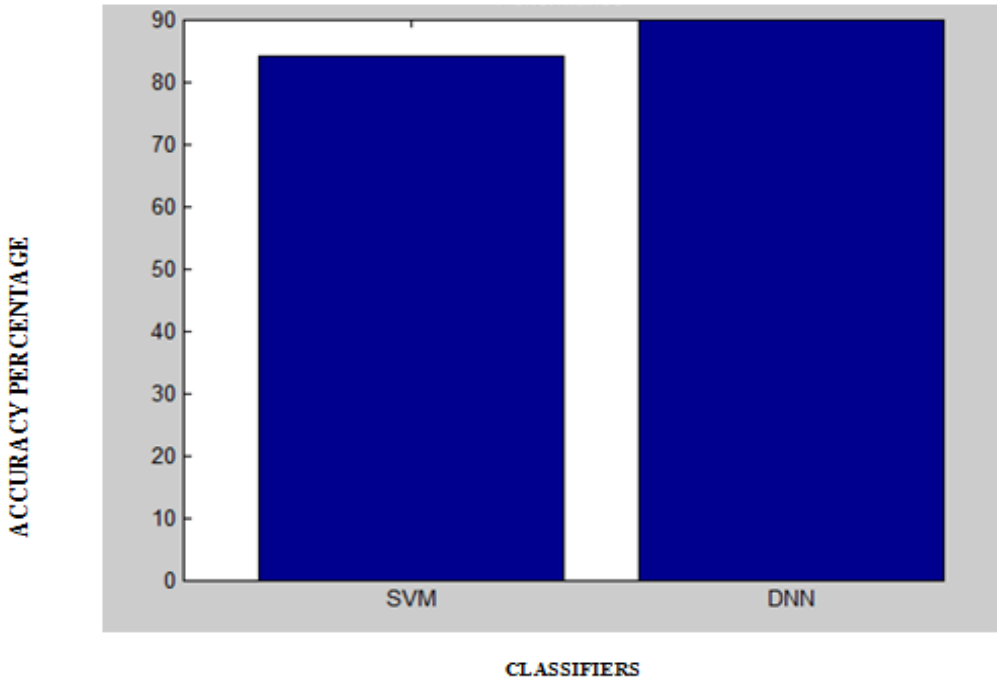


Figure.4.8 Accuracy of SVM and DNN

4.2 Chapter Summary

The text processing has been carried out by using crawling, case folding, convert emoticons, tokenizing, filtering and stemming, and the classification of text was done using Recurrent Neural Network algorithm**.** The performance metrics such as precision, recall, F-measure and accuracy were also discussed in this chapter**.** The accuracy of the Recurrent Neural Network algorithm was 90% when compared with SVM 82%.

CHAPTER 5

CONCLUSION AND FUTURE WORK

In this research work, text mining is achieved through the various steps like text processing, feature extraction and classification using DNN. The dataset used for this research work are tax tweets from Twitter, and the results of classification are based on positive, negative and neutral sentiments. The results achieved by the proposed method are 90%, 82%, and 89% respectively for precision,recall and f-measure. The assessment of tax services based on public opinion in tax cases. As a result, DNN accomplished 90% accuracy compared to SVM 82% accuracy.

FUTURE RESEARCH WORK

- ✓ The research work can be further improved by using another classification algorithm namely Relevance Vector Machine (RVM).
- ✓ Apply this research work to other social media applications like LinkedIn, Facebook, Instagram etc.

BIBLIOGRAPHY

A. Ratnaparkhi, (1996) A maximum entropy model for part-of speech tagging. In Proceedings of the EMNLP Conference, Pp. 133–142.

A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, (2013) Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization, Procedia Engineering, Vol. 53, Pp. 453–462.

Ahmad, Munir, & ShabibAftab, (2016) Analyzing the Performance of SVM for Polarity Detection with Different Datasets. International Journal Modern Education and Computer Science, Pp.29-36.

Al Emran, M.,& Shaalan, K. (2014) A Survey of Intelligent Language Tutoring Systems. In Advances in Computing, Communications and Informatics ICACCI, International Conference on Pp. 393-399 IEEE.

Al-Emran, M. N. H. (2014). Investigating Students and Faculty members Attitudes Towards the Use of Mobile Learning in Higher Educational Environments at the Gulf Region, Pp.195-202.

Al-Emran, M., & Malik, S. I. (2016) The Impact of Google Apps at Work: Higher Educational Perspective. International Journal of Interactive Mobile Technologies, Pp.85-88.

Al-Emran, M., & Shaalan, K. (2015). Learners and educators attitudes towards mobile learning in higher education: State of the art. In Advances in Computing, Communications and Informatics (ICACCI),IEEE. International Conference on Pp. 907-913.

Al-Emran, M., Zaza, S., & Shaalan, K. (2015). Parsing modern standard Arabic using Treebank resources. In Information and Communication Technology Research ICTRC, International Conference on Pp. 80-83 IEEE.

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016) A multi-stage method for content classification and opinion mining on weblog comments. Annals of Operations Research, Pp.197-213.

Alomari, K. M., ElSherif, H. M., &Shaalan, K. (2017) Arabic Tweets Sentimental Analysis Using Machine Learning. In International Conference on Industrial, Engineering and other Applications of Applied Intelligent Systems , Springer, Cham, Pp. 602-610.

Amolik, A.,Jivane N.,Bhandari M., Venkatesan M.(2016) Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques. International Journal of Engineering and Technology (IJET), Vol 7, Pp.2038-2044.

Andrea EsuliBaccianella, Stefano and FabrizioSebastiani.(2010) SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of LREC, Vol 10, Pp. 2200-2204.

AndriusMudinas, Dell Zhang, Mark Levene. (2012) Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis, ACM, Pp.1-8.

Anta, A. F., Chiroque, L. N., Morere, P., & Santos, A. (2013) Sentiment analysis and topic detection of Spanish tweets: A comparative study of NLP techniques. Vol.50, Pp. 45-52.

Arun, K., Srinagesh, A., & Ramesh, M. (2017) Twitter Sentiment Analysis on Demonetization tweets in India Using R language. International Journal of Computer Engineering in Research Trends, Pp. 252-258.

B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera. (2012) Opinion mining and sentiment analysis on a twitter data stream, in Advances in ICT for emerging regions, International Conference on. IEEE, Pp. 182–188.

B. Liu. (2007) Web Data Mining: Exploring Hyperlinks, Contents and Usage Data. Springer.

B. Pang and L. Lee. (2004) A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the ACL Conference.Pp.1-8.

B. Pang, L. Lee et al, (2008) Opinion mining and sentiment analysis, Foundations and Trends® in Information Retrieval, Vol. 2, Pp. 1–135.

B. Pang, L. Lee, and S. Vaithyanathan. (2002) Thumbs up? sentiment classification using machine learning techniques. In Proceedings of the ACL EMNLP Conference, Pp. 79–86.

Bai, R. Padman. (2006) Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text, Lecture Notes in Computer Science, Vol. 3932, Pp. 167–187.

Basha, S. M., Zhenning, Y., Rajput, D. S., Iyengar, N., & Caytiles, D. R. (2017) Weighted Fuzzy Rule Based SentimentPrediction Analysis on Tweets. International Journal of Gridand Distributed Computing, Pp. 41-54.

Berkhin, P. (2002) Survey of clustering and data mining techniques, Technical Report, Accrue Software, Inc. Pp.1-15

Berry Michael, W. (2004) Automatic Discovery of Similar Words. Survey of Text Mining: Clustering, Classification and Retrieval", Springer Verlag, New York, Pp. 24-43.

Berry, M. J. A. and Linoff, G. (1997) Data Mining Techniques: For Marketing, Sale, and Customer Support, John Wiley & Sons Inc, New York, NY, Pp. 55.

Bissattini C., Christodoulou K.(2013) Web sentiment analysis for revealing public opinions, trends and making good financial decisions. Journal of Advanced Research in Computer Science and Software Engineering.Pp.33-37.

Buche A., Chandak M.B., Zadgaonkar A.(2013) Opinion mining and analysis: a survey. International Journal on Natural Language Computing IJNLC, Pp.39-48.

C. Fellbaum,(1998) Wordnet: An electronic lexical database (language, speech, and communication), Pp.1-5.

C. Whissell. The dictionary of affect in language. Emotion: Theory, Research, and Experience, Pp. 113–131.

C.Bhadane, H.Dalal, and H.Doshi. (2015) Sentiment analysis: Measuring opinions, ProcediaComput.Sci., Vol.45, Pp. 808-814.

C.V. Trappey, T.T. Hsio, T.C. Chang, M.H. Che, W.J. Chiu. (2005) Consumer Driven Game Design, Proceedings of DiGRA Conference: Changing Views – Worlds in Play, Pp.265-270.

Carlo Strapparava and Alessandro Valitutti. (2004) Word net affect: an affective extension of wordnet. In LREC, vol. 4, Pp.1083-1086.

Cataldi M.,Ballatore A., Ilaria T. (2013) Good location, terrible food: Detecting feature sentiment in user-generated reviews. International Journal of Social Network Analysis and Mining (SNAM), Pp.1-16.

CataldoMusto, Giovanni Semeraro, Marco Polignano. (2014) A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts .8th International Workshop on Information Filtering and Retrieval Pisa Italy, Pp.1-10.

Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. (2007) Emotion classification using web blog corpora. In WI 07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society, Pp.275-278.

Chen, B., Tai, P.C., Harrison, R., and Yi, P. (2005) Novel hybrid hierarchical-K-means clustering method (H-K-means) for microarray analysis, Computational Systems Bioinformatics Conference, Aug. 8- 11, Standford CA, USA, Pp. 138-146.

Chenghua Lin, Yulan He, Richard Everson, Member, IEEE, and Stefan Ru¨ger. (2012) Weakly Supervised Joint Sentiment-Topic Detection from Text, IEEE trans. on knowledge and data engineering, Vol. 24, Pp.1134-1145.

Cho, K. et al. (2014) Learning phrase representations using RNN encoder decoder for statistical machine translation. In Proc. Conference on Empirical Methods in Natural Language Processing , Pp.1724–1734.

Curator. S. (2013) Social media: a rich source of customer sentiment for you to mine. Whatech

Channel, Pp.33-37.

D. Pucci, M. Baroni, F. Cutugno and A. Lenci. (2009) Unsupervised lexical substitution with a word space model, Proceedings of EVALITA workshop, 11th Congress of Italian Association for Artificial Intelligence, Citeseer, Pp.1-6.

D. Tang, F. Wei, B. Qin, T. Liu, and M. Zhou. (2014) Coooolll: A deep learning system for twitter sentiment classification, in Proceedings of the 8th international workshop on semantic evaluation, Pp. 208–212.

Danushka Bollegala, David Weir, and John Carroll. (2013) Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus, IEEE trans. on knowledge and data engineering, Vol. 25, Pp. 1-14.

David Combe , Christine Largeron , Elod Egyed-Zsigmond and Mathias Gery.(2010) A Comparative study of social network analysis tool, International Workshop on Web Intelligence and Virtual Enterprises 2, Pp.1-13.

Dedhia, C and Ramteke, J. (2017) Ensemble model for Twitter Sentiment Analysis.International Conference on Inventive Systems and Control, Pp.17-25.

Dhanalakshmi V.,DhivyaBino, Saravanan A.M. (2016) Opinion mining from student feedback data using supervised learning algorithms, International Conference on Big data and Smart City, Pp.1-5.

Diana Lupan, MihaiDascalu, Stefan Trausan-Matu, Philippe Dessus. (2012) Analyzing Emotional States Induced by News Articles with Latent Semantic Analysis, International Conference, Pp.1-11.

Dongcheol,K.and Soon--Ho,K. (2012) Investor Sentiment from Internet Message Postings and Predictability of Stock Returns. Working paper, Korea University Business School, Seoul 136--701, Pp. 2--53.

E. Alpaydin.(2009) Introduction to machine learning MIT press.

E. Boiy, P. Hens, K. Deschacht and M. Moens. (2007) Automatic sentiment analysis in on-line text, 11th International Conference on Electronic Publishing, Vol. 349360, Pp.349-360.

Erik Cambria, Daniel Olsher, and DheerajRajagopal. (2014) SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. AAAI, Quebec City, Pp.1515-1521.

Evans, B. M., Kairam, S., &Pirolli, P. (2010). Do your friends make you smarter?: An analysis of social strategies in online information seeking. Information Processing & Management, Pp. 679-692.

F. Camastra, A. Ciaramella, A. Placitelli, and A. Staiano. (2015) Machine Learning-based Web Documents Categorization by Semantic Graphs, Research Gate Publisher, Pp.75-82.

F. G. Contratres, S. N. Alves-Souza, L. V. L. Filgueiras, and L. S. DeSouza. (2018) Sentiment analysis of social network data for cold-start relief in recommender systems, in World Conference on Information Systems and Technologies. Springer, Pp. 122–132.

FabrizioSebastiani. (2002) Machine Learning in Automated Text Categorization, ACM Computing Surveys, Vol. 34, Pp. 1–47.

Fatima,S.,&Srinivasu,B. (2017) Text Document Categorization using support Vector Machine. International Research Journal of Engineering and Technology (IRJET), ACM, Vol.4, Pp. 375-384.

Francesco Bonchi, Carlos Castillo, Aristides Gionis and Alejandro Jaimes. (2011) Social Network Analysis and Mining for Business Application, ACM Trans. Intell. Syst. Pp.37.

Fuji Ren,Ye Wu. (2013) Predicting User-Topic Opinions in Twitter with Social and Topical Context, IEEE Trans on affective computing, vol.4, Pp.412-424

G. A. Miller. Nouns in WordNet : A lexical inheritance system. Int. J. of Lexicography.

G. Gautam and D. Yadav. (2014) Sentiment analysis of twitter data using machine learning approaches and semantic analysis, in Contemporary computing (IC3), 2 seventh international conference on. IEEE, Pp. 437–442.

G. Vinodhini and R. Chandrasekaran (2012) Sentiment analysis and opinion mining: a survey, International Journal, Vol. 2, Pp. 282–292.

Gartner.Hype Cycle for social software, 2008.G00158239, 2008.

George A Miller. (1995) WordNet: a lexical database for English. Communications of the ACM, Pp.39-41.

Gephi, an open source graph visualization and manipulation software https://gephi.org/

GiorgosCheliotis. Social Network Analysis (SNA) including a tutorial on concepts and methods, Communications and New Media, National University of Singapore.

Grimes, S. (2008). Unstructured data and the 80 percent rule. Carabridge Bridge points.

H. Hamdan, P. Bellot, and F. Bechet. (2015) lsislif: Feature extraction and label weighting for sentiment analysis in twitter, in Proceedings of the 9th International Workshop on Semantic Evaluation , Pp. 568–573.

H.Saif, Y.He, M.Fernandez, and H.Alani.(2015) Contextual semantics for sentiment analysis of Twitter, Inf. Process.Manag, Vol.5, Pp. 5-1.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H.(2009) The WEKA Data Mining Software: An Update, ACM SIGKDD Explorations Newsletter, Vol. 11, Pp. 10-18.

Hatzivassiloglou, V., &McKeown, K.R. (1997) Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL. New Brunswick, NJ: ACL.Pp. 174-181.

Hayes, P. and Weinstein, S. (1990). Construe/tis: a system for content-based indexing of a

database of news stories. In Annual Conferenceon Innovative Applications of AI, Pp.1-16.

HimanshiAgrawal, RishabhKaushal. (2016) Analysis of Text Mining Techniques over Public Pages of Facebook ,International Conference on Advance Computing, Pp.9-14.

HirotoshiTaira and Masahiko Haruno.(1999) Feature selection in SVM text categorization. In Proceedings of the 16th National Conference on Artificial Intelligence / Eleventh Conference on Innovative Applications of Artificial Intelligence, Pp. 480–486.

Hochreiter, S. &Schmidhuber .(1997) Long short-term memory. Neural Computation. Vol.9, Pp. 1735–1780.

Hsin-Ying Wu, Kuan-Liang Liu. (2014) Charles Trappey for Understanding Customers Using Facebook Pages: Data Mining Users Feedback Using Text Analysis. International Conference on Computer Supported Cooperative Work in Design, Pp.346-350.

Hung, J. L., & Zhang, K. (2012).Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. Journal of Computing in Higher education, Pp. 1-17.

Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A. &Tziritas, N. (2015). A survey on text mining in social networks. The Knowledge Engineering Review,Pp.157-170.

J. A. Tomlin. (2003)A new paradigm for ranking pages on the world wide web. In Proceedings of the WWW Conference.

J. B. Lovins (2011) Development of a stemming algorithm, Mech. Translat.& Comp. Linguistics, Vol.11, Pp. 22–31.

J. Han, J. Pei, and M. Kamber. (2011) Data mining: concepts and techniques. Elsevier, Pp.1-740

J. L. Elman. (2010) Finding structure in time, Cognitive science, Vol.14, Pp. 179–211.

J. M. Wiebe. (2000) Learning subjective adjectives from corpora. In Proceedings of the AAAI Conference, Pp.1-6.

J. Scott.(1994) Social Network Analysis : A handbook. Newbury Park CA , Sage Publications, Pp.1-30.

J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. (2003) Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In Proceedings ofthe IEEE International Conference on Data Mining, Pp. 1-8.

JJagtap V.S., Pawar K. (2013) Analysis of different approaches to sentence-level sentiment classification. International Journal of Scientific Engineering and Technology, Pp. 164-170.

JanyceWiebe, Theresa Wilson, and Claire Cardie.(2005) Annotating expressions of opinions and emotions in language. Language resources and evaluation, Pp.165-210.

K. S. McCurley.(2002) Geospatial mapping and navigation of the web. In Proceedings of the WWW Conference, Pp.1-8.

Kamal Nigam, Andrew Mccallum, Sebastian Thrun, and Tom Mitchell.(2002) Text classification from labeled and unlabeled documents using em. Machine Learning, Pp.103–134.

Kasture, N. R., &Bhilare, P. B. (2015) An Approach for Sentiment analysis on social networking sites. In Computing Communication Control and Automation (ICCUBEA), International Conference on IEEE ,Pp. 390-395.

Kennedy and D. Inkpen,(2006) Sentiment classification of movie reviews using contextual valence shifters, Computational Intelligence, Vol. 22, Pp. 110–125.

Kim Hammar, ShathaJaradat, NimaDokoohaki. (2015) Text Mining of Instagram data, International Conference, Pp.1-9.

Korde, V.&Mahender, C.(2012) Text Classification and Classifiers: A Survey. International

Journal of Artificial Intelligence & Applications (IJAIA), Vol. 3, Pp. 85-99.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg!.Icwsm, Pp.538-541.

KoweikaA.,Gupta A.,Sondhi K.(2013) Sentiment analysis for social media. International Journal of Advanced Research in Computer Science and Software Engineering, Pp.1-12.

Kumar,A.andTeeja, M.S. (2012) Sentiment Analysis on Twitter. IJCSI International Journal of Computer Science Issues, Vol.9, No 3, Pp. 372--373.

Kwak, H., Lee, C., Park, H. &Moon, S. (2010) What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World Wide Web, Pp.591–600.

L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. (2011) Target-dependent twitter sentiment classification, in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, Pp. 151–160.

L. Rovinelli and C. Whissell. (1998) Emotion and style in 30- second television advertisements targeted at men, women, boys, and girls. Perceptual and Motor Skills, Pp.1048–1050.

L. Terveen,W. Hill, B. Amento, D.McDonald, and J. Creter. (1998) PHOAKS: A system for sharing recommendations. CACM, Pp.59–62.

Law, D., Gruss, R., & Abrahams, A. S. (2017) Automated defect discovery for dishwasher appliances from onlineconsumer reviews. Expert Systems with Applications, Pp.84-94.

Lee, K. (2003). Text Categorization with a Small Number of Labeled Training Examples.PhD Thesis, School of Information Technologies, University of Sydney, Pp.1-171.

LI Bing, Keith C.C. Chan, Carol OU. (2014) Public Sentiment Analysis in Twitter Data for

Prediction of A Company's Stock Price Movements, 11th International Conference on e-Business Engineering IEEE, Pp.1-16.

Li, J., & Khan, S. U. (2009).MobiSN: Semantics-based mobile ad hoc social network framework. In Global Telecommunications Conference, GLOBECOM . IEEE Pp. 1-6.

Li, N., & Wu, D. D. (2010) Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision support systems, Pp.354-368.

Lin, C., & He, Y. (2009).Joint sentiment/topic model for sentiment analysis. In Proceedings of the 18th ACMconference on Information and knowledge management Pp. 1-10.

Liu, X., Wang, M., &Huet, B. (2016) Event analysis in social multimedia: a survey. Frontiers of Computer Science, Pp.1-14.

M. A. Friedl and C. E. Brodley, Decision tree classification of land cover from remotely sensed data, Remote sensing of environment, Vol. 61, Pp. 399–409.

M. D. Bah, A. Hafiane, and R. Canals. (2018) Deep learning with unsupervised data labeling for weeds detection on uav images, Pp.1-22.

M. M. Mostafa. (2013) More than words: Social networks text mining for consumer brand sentiments, Expert Systems with Applications, Vol. 40, Pp. 4241–4251.

M. Neethu and R. Rajasree. (2013) Sentiment analysis in twitter using machine learning techniques in Computing, Communications and Networking Technologies (ICCCNT), Fourth International Conference on. IEEE, Pp. 1–5.

M. P. Marcus, B. Santorini, and M. A. Marcinkiewicz. (1993) Building a large annotated corpus of english: the penntree bank. Computational Linguistics, Pp.1-12.

M. Sokolova, N.Japkowicz, and S. Szpakowicz. (2006) Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, in Australasian joint conference on artificial intelligence. Springer, 2006, Pp. 1015–1021.

M.D.Devika, C.Sunitha, and A.Ganesh, Sentiment Analysis: A Comparative Study on

Different Approaches, in Procedia Computer Science, 2016, vol. 87, Pp. 44-49.

Manjaly J.S. (2013) Twitter based sentiment analysis for subject identification. International Journal of Advanced Research in Computer and Communication Engineering, Pp.1-7.

Mathieu Bastian and SebastienHeymann, Mathieu Jacomy.(2009) Gephi : An Open Source Software for Exploring and Manipulating Networks, International AAAI Conference on Weblogs and Social Media,Pp.1-12.

McAuley J, Pandey J, Leskovec J. (2015) Inferring networks of substitutable and complementary products, Knowledge Discovery and Data Mining,Pp.1-2.

McDonald, R., Hannan, K., Neylon, T., Wells, M., &Reynar,J. (2007) Structured models for fine-to-coarse sentiment analysis. In Proceedings of the 45th annual meeting of the association of computational linguistics, Pp. 432-439.

Mesnil, G., Mikolov, T., Ranzato, M. A., &Bengio, Y. (2014).Ensemble of generative and discriminative techniques forsentiment analysis of movie reviews, Pp. 1-12.

Mikolov T, Karafiat M, Burget L, Cernock L, Khudanpur S. (2010) Recurrent neural network based language model, In: Proceedings of Interspeech,Pp.1045-1048.

Minas Gjoka and MaciejKurant and Carter T. Butts and AthinaMarkopoulou.(2010) Walking in Facebook: A Case Study of Unbiased Sampling of OSNs, Proceedings of IEEE INFOCOM '10, San Diego, CA, Pp.1-14.

Mudinas, A., Zhang, D., &Levene, M. (2012) Combining lexicon and learning based approaches for concept-level sentiment analysis. In Proceedings of the first international workshop on issues of sentiment discovery and opinion mining ACM ,Pp.12-15.

Naaman, M. (2012) Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. Multimedia Tools and Applications, Pp. 9-34.

NadeemAkhtar, HiraJaved, GeetanjaliSengar (2013) for Analysis of facebook social network. International Conference on Computational Intelligence and Communication Networks,Pp.1-14.

Nakov P.,Kozareva Z., Ritter A.(2013) Sentiment analysis in Twitter. Second Joint Conference on Lexical and Computational Semantics, Atlanta, Georgia, Pp. 312-320.

Neethu M, S, RajasreeR.(2013) Sentiment analysis in Twitter using Machine Learning Techniques, 4th ICCCNT, Pp. 2038- 2044.

Ng, V., Dasgupta, S., &Arifin, S. M. (200) Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In Proceedings of the COLING/ACL on Main conference poster sessions, Pp.611-618.

Nguyen, D. Q., Nguyen, D. Q., Vu, T., & Pham, S. B. (2014) Sentiment classification on polarity reviews: an empirical study using rating-based features. In Proceedings of the 5thworkshop on computational approaches to subjectivity, sentiment and social media analysis, Pp. 128-135.

O'leary, Daniel E. (2011) The use of social media in the supply chain: survey and extensions."Intelligent Systems in Accounting, Finance and Management 18.2-3 Pp. 121-144.

P. D. Turney. (2002) Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In Proceedings of the ACL Conference, Pp.417–424.

P. Domingos and M. Pazzani. (1997) On the optimality of the simple bayesian classifier under zero-one loss, Machine Learning, Vol. 29, Pp. 2-3.

P. Subasic and A. Huettner. (2011) Affect analysis of text using fuzzy semantic typing.IEEE Trans. on Fuzzy Systems, SpecialIssue, Pp. 1-5.

Pang, Bo; Lee, Lillian (2005).Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. Proceedings of the Association for

Computational Linguistics (ACL), Pp. 115–124.

Pascanu, R., Mikolov, T. &Bengio, Y. (2013)On the difficulty of training recurrent neural networks. In Proc. 30th International Conference on Machine Learning , Pp.1310–1318.

Popescu, A. M., Etzioni, O.(2005) Extracting Product Features and Opinions from Reviews, In Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Pp. 339–346.

Po-Wei Liang, Bi-Ru Dai. (2013) Opinion Mining on Social Media, International Conference on Mobile Data Management, IEEE , Pp.91-96.

Prabowo, R., &Thelwall, M. (2009). Sentiment analysis: A combined approach. Journal of Informetrics, 3 (2), Pp.143-157.

PritamGundecha, Huan Liu. Mining Social Media: A Brief Introduction Tutorials in Operations Research ,INFORMS 2012

Qingliang Miao, Qiudan Li, Ruwei Dai. (2009) AMAZING: A sentiment mining and retrieval system, Expert Systems with Applications 36 ,Pp.7192–7198.

QiuM.,Yang L., Jiang J. (2013) Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization. Proceedings of NAACL-HLT, Atlanta, Georgia,Pp. 2-8.

R. M. Tong. (2010)An operational system for detecting and tracking opinions in on-line discussion. In SIGIR Workshop on Operational Text Classification, Pp. 21-30.

R.A. Feinberg, K. DeRuyter, C.V. Trappey, and T.Z. Lee. (2005) Consumer- Defined Service Quality in International Retailing, Total Quality Management, Vol. 6, Pp. 61-67.

Rahmani, A., Chen, A., Sarhan, A., Jida, J., Rifaie, M., &Alhajj, R. (2014). Social media analysis and summarization for opinion mining: a business case study. Social Network Analysis and Mining, 4(1), Pp.1-11.

Rajeswari R.P., Kavitha Juliet, Dr.Aradhana. (2017) Text Classification for Student Data Set using Naïve Bayes Classifier and KNN Classifier, International Journal of Computer Trends and Technology 43(1), Pp. 8-12.

Robinson, R., Goh, T. T., & Zhang, R. (2012) Textual factors in online product reviews: a foundation for a more influential approach to opinion mining. Electronic Commerce Research, 12(3), Pp. 301-330.

Rudy Prabowo, Mike Thelwall. (2009) Sentiment analysis: A combined approach. Journal of Informetrics3 ,Pp.143–157.

Rui Xia, ChengqingZong, Shoushan Li. (2011) Ensemble of feature sets and classification algorithms for sentiment classification, Information Sciences 181 Pp. 1138–1152.

Rui Xia, FengXu, ChengqingZong,Qianmu Li, Yong Qi, and Tao Li. (2015) Dual Sentiment Analysis: Considering Two Sides of One Review, IEEE Trans.on Knowledge and Data Engineering, Pp.1783-1787.

S. Asur and B. A. Huberman.(2010) Predicting the future with social media, in Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Vol.1. IEEE Computer Society, ,Pp. 492–499.

S. Morinaga, K. Yamanishi, K. Teteishi, and T. Fukushima. (2002) Mining product reputations on the web. In Proceedings of the ACM SIGKDD Conference,Pp.1-5.

S. Vijayarani, R.Janani.(2016) Text Mining :A survey, International Journal of Innovative Research in Computer and Communication Engineering, Pp.1-9.

S. Wasserman and K. Faust.(1994) Social Network Analysis. bridge Cambridge University Press, Pp.1-12.

Saleh, M. R., Martín-Valdivia, M. T., Montejo-Ráez, A., &Ureña-López, L. A. (2011).Experiments with SVM to classify opinions in different domains. Expert Systems with Applications, 38 (12), Pp. 14799-14804.

Salloum, S. A., Al-Emran, M., &Shaalan, K. (2016). A Survey of Lexical Functional Grammar in the Arabic Context.Int. J. Com. Net. Tech, 4(3),Pp.1-7.

Salloum, S. A., AlHamad, A. Q., Al-Emran, M., &Shaalan,K. (2018). A Survey of Arabic Text Mining. In Intelligent Natural Language Processing: Trends and Applications Springer, Cham, Pp.417-431.

Salvatore A. Catanese, Pasquale De Meo,, Emilio Ferrara, GiacomoFiumara, Alessandro Provetti. (2011) Crawling Facebook for Social Network Analysis purpose WIMS'11, Sogndal, Norway,Pp.25-27.

Savage N. (2011) Twitter as medium and message, Communications of the ACM Society, Issue: 3, Vol.54, Pp: 18-20.

Schuster, M. and Paliwal, K. K. (2017) Bidirectional recurrent neural networks. Signal Processing, IEEE Transactions on, 45(11), Pp.2673–268.

Sebastiani F. (2002) Machine Learning in Automated Text Categorization, ACM Computing Surveys, Pp. 1-47.

Sheshasaayee A. &Thailambal G. Comparison of Classification Algorithms in Text Mining. International Journal of Pure and Applied Mathematics 2017, Vol.116 ,Pp 425-433.

Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, JiajunBu, ,Chun Chen, and Xiaofei He. (2014) Interpreting the Public Sentiment Variations on Twitter, IEEE trans. on knowledge and data engineering, Vol. 26, Pp. 1158-1170.

Siwei Lai, LihengXu, Kang Liu, and Jun Zhao. (2015) Recurrent convolutional neural networks for text classification. In Twenty-Ninth AAAI Conference on Artificial Intelligence, Pp.2267-2273.

Socher R, Perelygin A, J. Wu, J. Chuang, C. Manning, A. Ng, C. Potts.(2013) Recursive Deep Models for Semantic Compositionality Over a Sentiment Tree bank In Proceedings of EMNLP, Pp.1631-1642.

Sorensen, L. (2009) User managed trust in social networking-Comparing Facebook, MySpace and Linkedin. In Wireless Communication, Vehicular Technology, Information Theory and Aerospace &Electronic Systems Technology, Wireless VITAE 1st International Conference on IEEE, Pp.427-431.

Suge Wang, Deyu Li, Yingjie Wei, Hongxia Li (2009) for A Feature Selection Method Based on Fisher's Discriminant Ratio for Text Sentiment Classification, Pp.88-97.

Sulistiani, H., &Tjahyanto.(2017) A. Comparative Analysis of Feature Selection Method to Predict Customer Loyalty. Journal of Engineering, Vol.3, Pp.310-314.

T. Denoeux. (2015) A k-nearest neighbor classification rule based on dempstershafer theory, IEEE transactions on systems, man, and cybernetics, vol. 25, Pp. 804–813.

T. Mullen and N. Collier.(2004) Sentiment analysis using support vector machines with diverse information sources. In Proceedings of the ACL Conference, Pp. 1-12.

T. Shaikh and D. Deshpande. (2016) A review on opinion mining and sentiment analysis, in IJCA Proceedings on National Conference on Recent Trends in Computer Science and Information Technology, Vol. 2, Pp. 6–9.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005).Introduction to Data Mining. Addison Wesley, 1 edition, Pp.1-169.

Thorsten Joachims. (2015) Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the European Conference on Machine Learning, Pp.137–142.

Timmaraju A, Khanna V. (2015) Sentiment analysis on movie reviews using recursive and recurrent neural network architectures Semantic Scholar; Pp.1-7.

Tiwari, P., Mishra, B. K., Kumar, S., & Kumar, V. (2017) Implementation of n-gram methodology for rotten tomatoes review dataset sentiment analysis. International Journal of Knowledge Discovery in Bioinformatics Pp. 30-41.

Tong, R.M. (2001)An operational system for detecting and tracking opinions in on-line discussions. Working Notes of the ACM SIGIR Workshop on Operational Text Classification . New York, NY: ACM, Pp.1-6.

Tripathy, A., Agrawal, A., &Rath, S. K. (2016) Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, Pp. 117-126.

Tsytsarau, M., &Palpanas, T. (2012) Survey on mining subjective data on the web. Data Mining and Knowledge Discovery, 24(3), Pp.478-514.

Tulankar S.,Athale R.,Bhujbal S (2013) Sentiment analysis of equities using data mining techniques and visualizing the trends. International Journal of Computer Science Issues, Pp.265-269.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre. (2008) Fast unfolding of communities in large networks, in Journal of Statistical echanics: Theory and Experiment , Pp. 454.

W. Sack.(2014) On the computation of point of view. In Proceedings of the AAAI Conference, Pp.1-12.

WalaaMedhat , Ahmed Hassan , HodaKorashy. (2014) Sentiment Analysis Algorithms and Applications: A survey ,AinShams Engineering Journal, Pp.7-11.

Wang, S., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Vol 2, Pp. 90-94.

Wang, Xingyou et al. (2016) Combination of Convolutional and Recurrent Neural Network for Sentiment Analysis of Short Texts. In COLING,Pp.1-12.

Wiebe, J.M. (2000) Learning subjective adjectives from corpora. Proceedings of the 17th National Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, Pp.12-16.

William, et al. (2018) Interaction and transformation on social media: the case of Twitter campaigns." Social Media+ Society4.1, Pp.1-12.

X. Ding, B. Liu, and P. S. Yu.(2008) A holistic lexicon-based approach to opinion mining, Proceedings of the Conference on Web Search and Web Data Mining (WSDM), Pp.1-9.

XiaohuiYu,Yang Liu, Jimmy Xiangji Huang, and Aijun An.(2012) Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain, IEEE Trans.On knowledge and data engineering, Vol. 24, Pp.1-5.

Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, Ming Zhang. (2011) Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach, ACM, CIKM'11, Pp. 24–28.

XinChen,MihaelaVorvoreanu, and Krishna Madhavan. (2014) Mining Social Media Data for Understanding Students' Learning Experiences IEEE trans. on learning technologies, Vol. 7, Pp.16-24.

Xing Fang and Justin Zhan.(2015) Sentiment analysis using product review data .Journal of Big Data, Pp.1-14.

Yang Liu, Xiangji Huang, AijunAn, Xiaohui Yu. (2007) ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs" SIGIR"07, Pp. 23–27.

Yang, Ang, et al. (2015) Enhanced Twitter Sentiment Analysis by using Feature Selection and Combination, Pp. 52-57.

Yang, L., Geng, X., & Liao, H. (2016) A web sentiment analysis method on fuzzy clustering for mobile social media users.EURASIP Journal on Wireless Communications and Networking, Pp. 1-13.

Younggue B.,Hongchul L.(2012) Sentiment analysis of Twitter audience: Measuring the positive or negative influence , Journal of the American Society for Information Science and Technology, Pp.2521-2535.

YulanHe. (2010) Learning sentiment classification model from labeled features. Proceedings of the 19th ACM, Conference on Information and Knowledge Management, Toronto, Ontario, Canada, Pp. 26-30.

Z.Hu, J.Hu, W.Ding, and X.Zheng,(2015) Review Sentiment Analysis Based on Deep Learning, IEEE 12th International Conference on e-Business Engineering, Pp. 87-94.

Zheng, L., Wang, H., &Gao, S. (2018) Sentimental feature selection for sentiment analysis of Chinese online reviews. International journal of machine learning and cybernetics, Pp. 75-84.

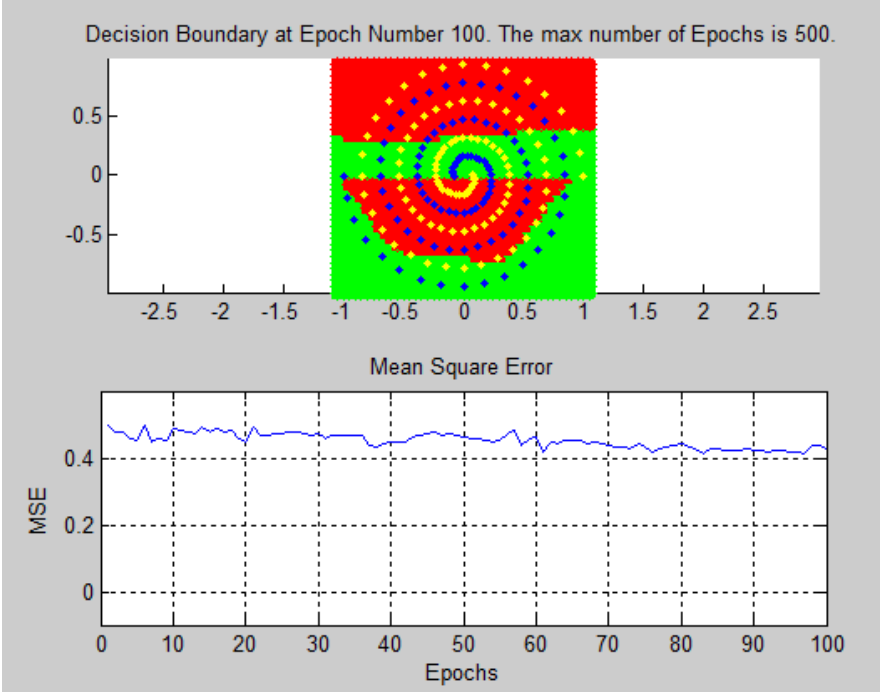
Zhou et al. (2016) Text Classification improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling, Pp.3485-3495.

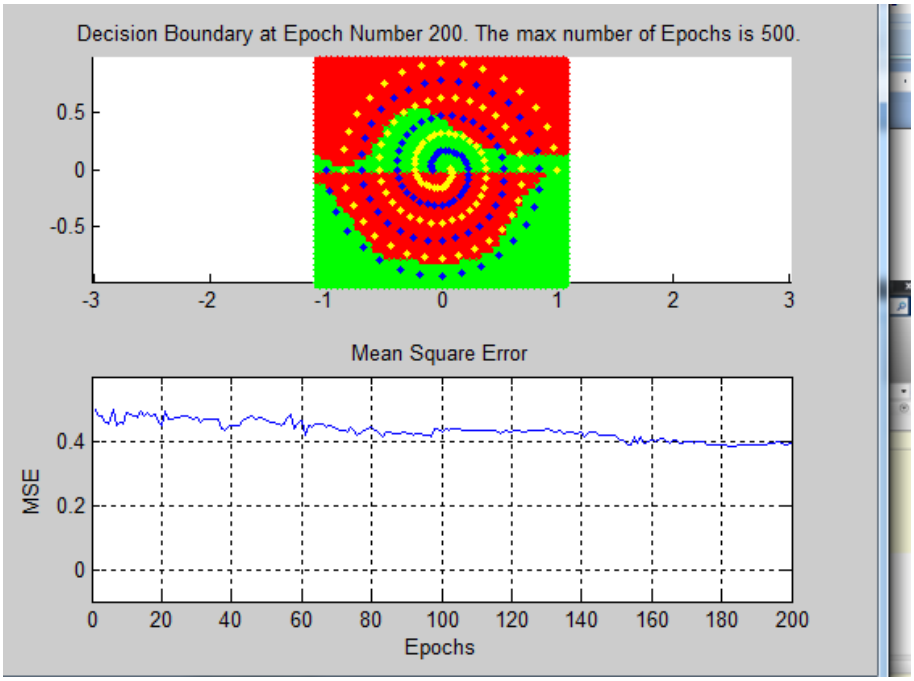
Zhou, L. and P. Chaovalit (2008) Ontology-supported Polarity Mining, Journal of the American Society for Information Science and Technology, Pp. 1-13.

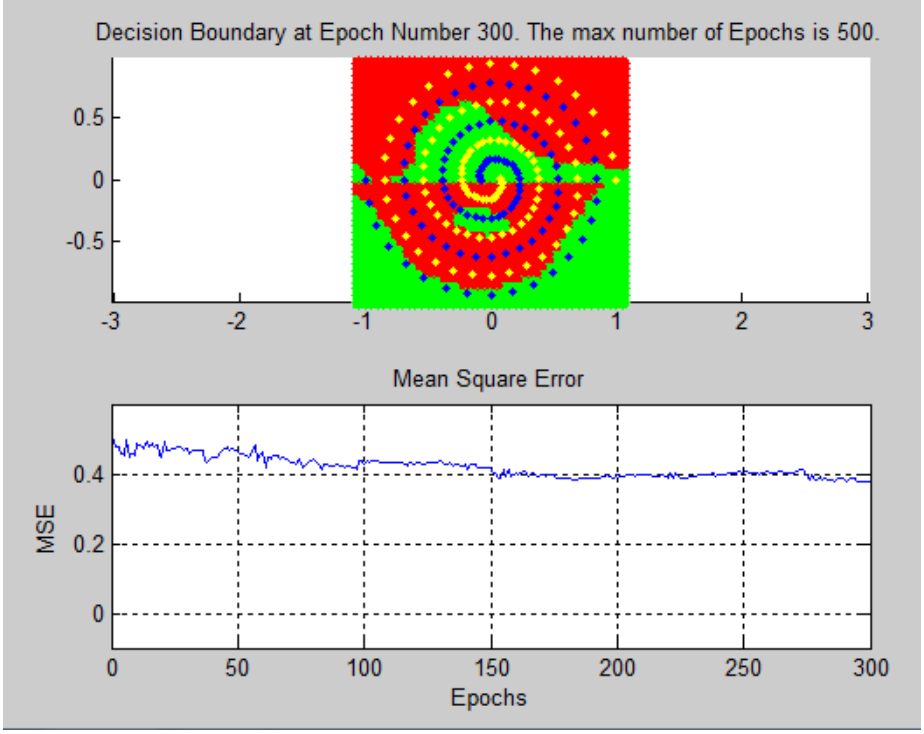
ZHU Jian, XU Chen, and WANG Han-shi.(2010) Sentiment classification using the theory of ANNs, The Journal of China Universities of Posts and Telecommunications, Pp. 58–62.

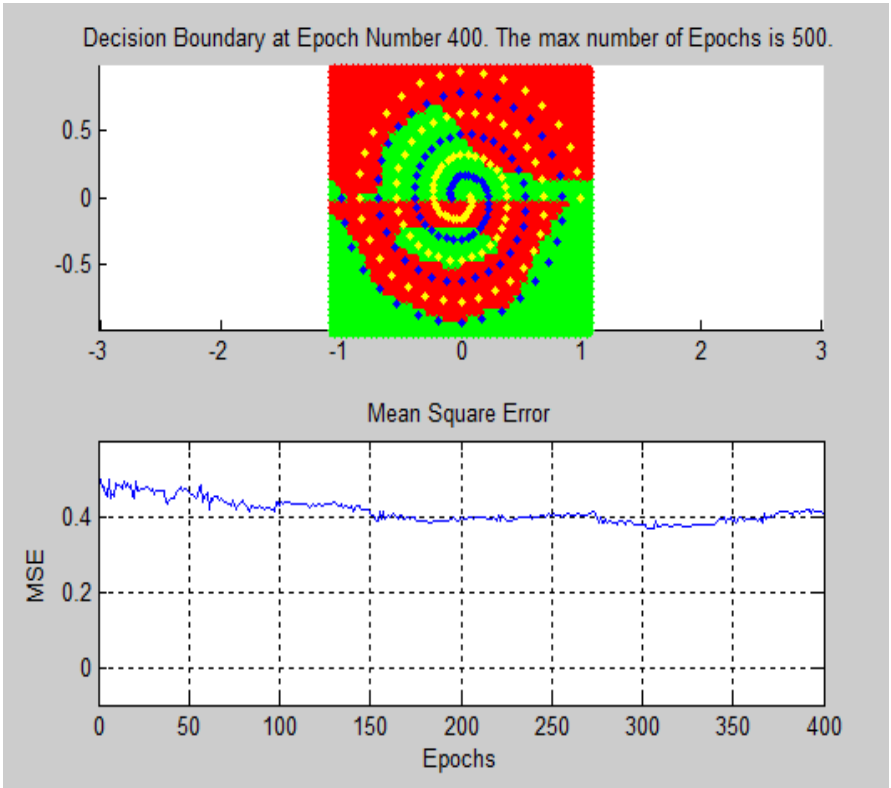
Zuell B., Preradovic N.M. (2013) Methods and usage of sentiment analysis in the context of the TV industry. International Journal of Recent Advances in Information Science, Pp.33-37.

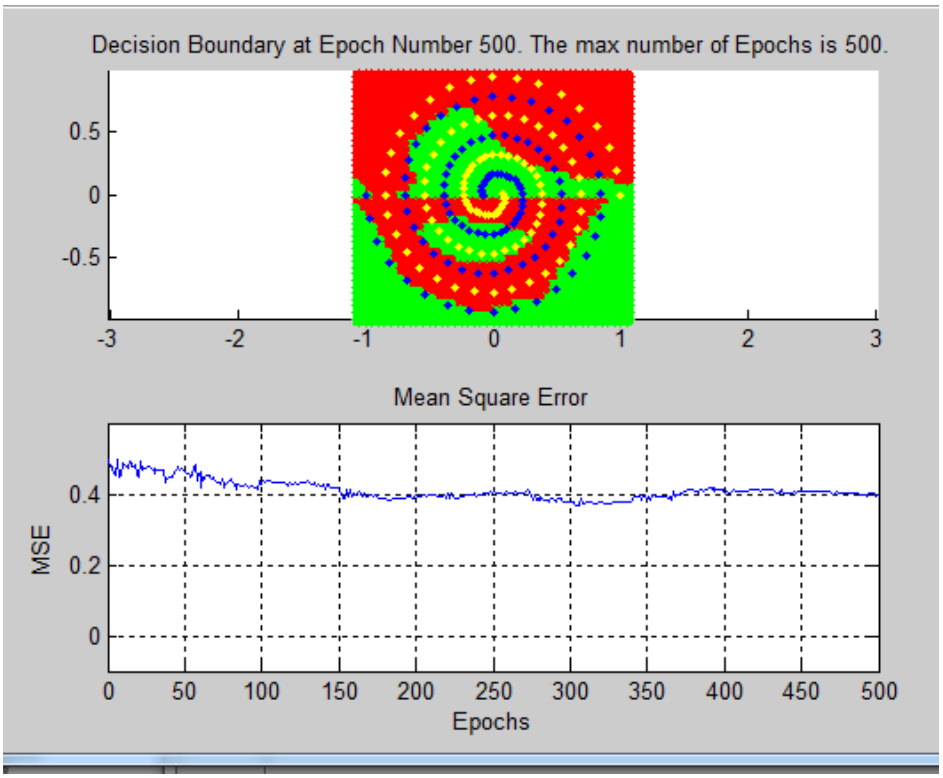
APPENDIX

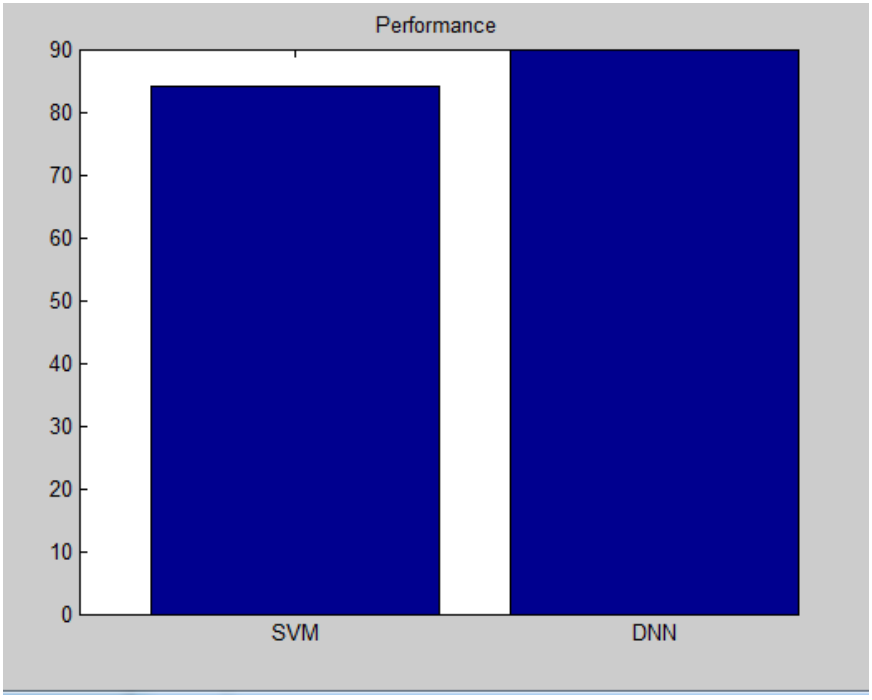

Decision Boundary at Epoch Number 100. The max number of Epochs is 500.
0.5
0
-0.5
-2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5
Mean Square Error
MSE
0.4
0.2
0
0 10 20 30 40 50 60 70 80 90 100
Epochs


Decision Boundary at Epoch Number 200. The max number of Epochs is 500.
0.5
0
-0.5
-3 -2 -1 0 1 2 3
Mean Square Error
MSE
0.4
0.2
0
0 20 40 60 80 100 120 140 160 180 200
Epochs


Decision Boundary at Epoch Number 300. The max number of Epochs is 500.
0.5
0
-0.5
-3
-2
-1
0
1
2
3
Mean Square Error
MSE
0.4
0.2
0
0
50
100
150
200
250
300
Epochs


Decision Boundary at Epoch Number 400. The max number of Epochs is 500.
0.5
0
-0.5
-3
-2
-1
0
1
2
3
Mean Square Error
MSE
0.4
0.2
0
0
50
100
150
200
250
300
350
400
Epochs


Decision Boundary at Epoch Number 500. The max number of Epochs is 500.
0.5
0
-0.5
-3
-2
-1
0
1
2
3
Mean Square Error
MSE
0.4
0.2
0
0
50
100
150
200
250
300
350
400
450
500
Epochs


Performance
90
80
70
60
50
40
30
20
10
0
SVM
DNN

PUBLICATIONS

1. Anitha SP, Dr. Radha V. (2019) Text Mining Based on Social Media- A Survey, International Journal of Scientific & Engineering Research, Vol. 10, Pp. 151-156. ISSN 2229 - 5518.
2. Anitha SP, Dr. Radha V. (2019) Opinion Mining Based on Tax Tweets using Deep Learning Neural Network (DNN) yet to publish.