AN ENHANCED APPROACH IN LEXICON-BASED SENTIMENT ANALYSIS FOR SOCIAL ISSUES

YASIR MEHMOOD

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

AN ENHANCED APPROACH IN LEXICON-BASED SENTIMENT ANALYSIS FOR SOCIAL ISSUES

YASIR MEHMOOD

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Yasir Mehmood

Matric No: WGA150001

Name of Degree: Master of Computer Science

Title of Dissertation ("this Work"): An Enhanced Approach in Lexicon-Based Sentiment Analysis for Social Issues

Field of Study: Opinion Mining (Information System)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate"s Signature   Date:

Subscribed and solemnly declared before,

Witness"s Signature   Date:

Name: Dr. Vimala Balakrishnan

Designation: Senior Lecturer

AN ENHANCED APPROACH IN LEXICON-BASED SENTIMENT ANALYSIS FOR SOCIAL ISSUES

ABSTRACT

Millions of users use social media to publically share their opinion and sentiment on different aspects of life. In decision making or option selection process, it is very important to know what others are thinking. In the last more than one and half decades, sentiment analysis has transformed into a very attractive research area due to the extended need to extract opinion and sentiment from the huge opinionated text data. In this context, mostly research has been conducted on the product and services. Nevertheless, sentiment analysis of social issues is different than sentiment analysis of product and services. Also minimal literature is available for the sentiment analysis of social issues. The purpose of this research is to enhance the lexicon-based sentiment analysis for social issues. Two datasets of custom data collected randomly from tweets over the issue of illegal immigration were used in the experiment of proposed technique. Same datasets manually labeled by the industry experts were analyzed by ten online sentiment analysis tools to check the effectiveness of proposed solution by using benchmark evaluation metrics precision, recall, F measure and accuracy. The proposed enhanced approach not only outperformed with overall accuracy of 81.4 and 82.3 as compared to the highest accuracy of 72.9 and 74.6 among ten online tools for both datasets respectively, but also classified each class of positive, negative and neutral with highest F measure values.

Keywords: Sentiment analysis, lexicon-based, sentiment analysis tools, online sentiment analysis, Twitter, General inquirer

PENINGKATAN PENDEKATAN DALAM LEXICON BERASASKAN ANALISIS SENTIMEN UNTUK MASALAH SOSIAL

ABSTRAK

Berjuta-juta pengguna menggunakan media sosial untuk umum berkongsi pendapat dan sentimen mereka mengenai pelbagai aspek kehidupan. Dalam membuat keputusan atau proses pemilihan pilihan, ia adalah sangat penting untuk mengetahui apa yang orang lain fikirkan. Dalam lebih daripada satu setengah dekad yang lalu, analisis sentimen telah berubah menjadi kawasan penyelidikan yang sangat menarik kerana keperluan lanjutan untuk mengeluarkan pendapat dan sentimen daripada data teks keras kepala besar. Dalam konteks ini, kebanyakannya kajian telah dijalankan ke atas produk dan perkhidmatan. Walau bagaimanapun, analisis sentimen isu-isu sosial adalah berbeza daripada analisis sentimen produk dan perkhidmatan. Juga sastera yang minimum disediakan untuk analisis sentimen isu-isu sosial. Dua set data data peribadi dikumpul secara rawak daripada tweet mengenai isu pendatang tanpa izin telah digunakan dalam eksperimen teknik dicadangkan. Dataset sama dilabel secara manual oleh pakar-pakar industri telah dianalisis oleh sepuluh talian alat analisis sentimen untuk memeriksa keberkesanan penyelesaian yang dicadangkan dengan menggunakan metrik penilaian penanda aras „precision‟, „recall‟, „F measure‟ dan „accuracy‟. Pendekatan dipertingkatkan yang dicadangkan bukan sahaja mengatasi dengan ketepatan keseluruhan 81.4 dan 82.3 berbanding ketepatan tertinggi 72.9 dan 74.6 di kalangan sepuluh alat dalam talian untuk kedua-dua set data masing-masing, tetapi juga diklasifikasikan setiap kelas positif, negative dan neutral dengan nilai-nilai tertinggi „F measure‟.

Keywords: Sentiment analysis, lexicon-based, sentiment analysis tools, online sentiment analysis, Twitter, General inquirer

ACKNOWLEDGEMENTS

First and foremost Alhamdulillah, all the praises to Allah almighty for the countless blessings, strength and energy in completing this thesis despite good and bad time through these years. I am deeply grateful to my supervisor Dr. Vimala Balakrishnan for giving me a chance of doing Masters under her supervision. She was always available when I knock her door, providing intelligent guidance, support and motivation which enable me to achieve the goals of this study.

I would also like to express my gratitude to my grandmother, parents, siblings and extended family member for their support and continuous prays throughout my life and especially during this span of time. My special appreciation and acknowledgement for my beloved wife for all the cooperation and encouragement to move forward regardless of difficulties faced all over this time of research. I also want to sincerely thank all the people who contributed in any way to the completion of this thesis. Today I am here due to all of these wonderful people in my life and I would not have been able to accomplish this without their endless support, love and motivation.

Last but not the least, I would like to thank the University of Malaya for the financial support under International Graduate Research Assistant Scheme (iGRAS).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

API	:	Application Programming Interface
COSMOS	:	Cardiff Online Social Media Observatory
KNN	:	K-nearest Neighbor
LSI	:	Latent Semantic Index
NLP	:	Natural Language Processing
PMI	:	Point-wise Mutual Information
POS	:	Parts of Speech
SA	:	Sentiment Analysis
SMS	:	Short Message Service
SO-CAL	:	Semantic Orientation Calculator
SVM	:	Support Vector Machines
WEKA	:	Waikato Environment for Knowledge Analysis

CHAPTER 1: INTRODUCTION

This chapter provides an introduction to the research work presented in this thesis. The overview of the background is briefly introduced in Section 1.1. Section 1.2 presents general thoughts about sentiment analysis and other key concepts of the research area. Section 1.3 states the problem statement. Research objectives and research questions are defined in Section 1.4. Section 1.5 summarizes the research methodology of this research and layout of the thesis‟ structure is highlighted in Section 1.6.

1.1 Background

The appearance of the second generation Word Wide Web impelled the advancement of various social networking web applications. Smart phone and micro-blogging applications like Twitter promoted the ease of communicating moment by moment opinions of users. Millions of Twitter users publically share their opinion and sentiment about daily activities, product and services, news, event, issues etc. Accordingly, an enormous opinionated data is being delivered by users; consequently the requirement for automatic techniques proficient to analyze users‟ sentiments, which is the core of sentiment analysis (Angulakshmi & ManickaChezian, 2014).

Sentiment analysis manages with the polarity of a sentence often referred as sentiment classification, in which a sentence may be utilized to represent positive, negative or neutral sentiment towards a product, service, topic, person or event (Serrano-Guerrero et al., 2015). Generally, in the real-world context sentiment analysis applications include user reviews about product and services, reputation monitoring, decision making and result prediction etc. The applications of sentiment analysis of social media are increasing day by day in every field of life like business (He et al., 2013), health care (Rodrigues et al., 2016), politics (Mohammad et al., 2015; Nasir et al., 2009), sports (Yu & Wang, 2015) and crisis incidents (Abel et al., 2012) etc. In the social domain, government, social societies and non-governmental organizations are constantly anxious with social effect, natural crisis and their viability in reacting to these challenges (Gundecha & Liu, 2012).

Twitter is one of the most famous social network application to which people turn to share their opinions about different aspects of life. Twitter messages which are called Tweets are publically available by default but users can restrict the visibility to limited audience as well. With the different powerful features like open access social network, easy and user friendly Application Programming Interface and high density sentiment availability; Twitter emerged as a gold mine for sentiment analysis researchers and industry practitioners. Twitter has been chosen in various sentiment analysis researches (Khan et al., 2014; Kolchyna et al., 2015; Mohammad et al., 2015) since the amount of relevant data is much larger with the better resemblance of public sentiment.

In the last more than one and half decades, sentiment analysis has emerged as very attractive research area due to proliferation of opinionated data and hence growing need of techniques in different fields. In the sentiment analysis research, most part of techniques has been directed on product and services (Hardeniya & Borikar, 2016; Karamibekr & Ghorbani, 2012a; Singh & Paul, 2015). However, sentiment analysis of social issues varies from sentiment analysis of social issues in certain aspects (Karamibekr & Ghorbani, 2012a). Additionally, minimal existing research is available for the sentiment analysis of social issues. Hence, investigation of new techniques and approaches in this area are imperative.

1.2 Sentiment Analysis

Sentiment analysis is the procedure of automatic identification and categorization of opinions and sentiment expressed in a bit of content, particularly in determining if the author‟s point of view towards a specific entity is positive, negative or neutral. In process of decision making or option selection, it is essential to recognize what others are considering (Pang & Lee, 2008a). Sentiment analysis deals with the polarity of textual content in which a sentence may be used to reflect positive, negative or neutral sentiment towards a particular product, topic or event etc. (Serrano-Guerrero et al., 2015). Sentiment analysis is an interdisciplinary field which is comprised of several domains such as information retrieval, natural language processing and machine learning (Cambria et al., 2013; Esuli & Sebastiani, 2007).

1.2.1 Sentiment Classification Approaches

Sentiment classification are majorly divided as machine learning, lexicon-based and hybrid approaches (Zhou et al., 2014). One can utilize the machine learning, lexicon-based or combination of both as a hybrid approach according to the specific criteria. The following sub-sections describe the approaches in brief.

1.2.1.1 **Machine Learning**

Machine learning approach is categorized into supervised and unsupervised learning techniques. In supervised learning approach, a list of features are selected and a labeled dataset called training dataset is provided to train the classifier, which can be practiced on an unlabeled dataset which is called testing dataset. The supervised learning techniques make usage of large training datasets for training and testing (Ravi & Ravi, 2015). On the other hand, the unsupervised methods are used when it is difficult to get labeled training data to classify the rest of the data (Medhat et al., 2014). In supervised learning techniques, feature selection is the vital stage which may result to a noisy classifier if not done precisely. N-grams, POS-Tagging, Syntax and Term Frequency are very compelling techniques in features selection. Support Vector Machines (SVM), Naïve Bayes and Maximum Entropy are the most important techniques utilized in the supervised learning approaches (Medhat et al., 2014).

1.2.1.2 **Lexicon Based**

Lexicon based approach is sub-divided into dictionary based and corpus based methods. The former is usually based on sentiment lexicons which contains list of words having positive or negative polarity information. The corpus based methods start with a set of opinionated seed words that widens through the search of relevant terms by utilizing statistical or semantic methods (Medhat et al., 2014). In statistical method, Latent Semantic Analysis or occurrence frequency of terms in the corpus can be used. The semantic approach use synonyms, antonyms or different semantic relationship from lexicons, for example WordNet. The corpus-based systems assist to solve the issue of providing domain specific dictionaries (Serrano-Guerrero et al., 2015).

Various automatic, semi-automatic and manual sentiment lexicons are available with variety of size and formats to be used in dictionary-based approaches. Some important lexicons used in industry and academia are SentiWordNet, WordNet-Affect, AFFIN, Opinion Lexicon, General Inquirer and SO-CAL etc. (Cho et al., 2014). General Inquirer (Stone et al., 1966), for example is a manually annotated category based lexicon which was initially intended to be aiding social science applications (Ohana et al., 2011). General Inquirer has 182 categories and each category contains some words based on four sources of categorizations i.e. Harvard IV-4 dictionary, Lasswell value dictionary, Marker categories, Semin and Fiedler categories (Semin & Fiedler, 1988), which are created manually.

1.2.2 Social Issues

Issues associated with people‟s individual lives and interactions with neighboring societies, environment and culture etc. are social issues. Social stratification, economic issues, social disorganization, public health, social inequality and work and occupation are example of some universal type of social issues (Singh & Paul, 2015). As manufacturer and companies are concerned with the public review about their product and services, in the same manner government and social organizations are keen to analyze user feedback and opinion in policy making decisions (Karamibekr & Ghorbani, 2012a).

Adjectives, adverbs and nouns are mainly used while expressing opinion about product and services and these are used as features as well in the machine learning techniques (Cambria et al., 2013). On the other hand, verbs perform essential part in the sentiment orientation of social issues‟ opinion (Karamibekr & Ghorbani, 2012b). Generally, a product has features for example battery, camera and display are features of a phone. On the other hand, a social issue is linked with other issues or sub-issues for example, unemployment, security and crimes are sub-issues related to illegal immigration.

1.3 Problem Statement

Due to overwhelming online opinionated data, there is an expanded requirement for the techniques of sentiment analysis in almost every field of life to have a more accurate insight. Sentiment analysis also has different challenges and complications like other research fields. It has been observed that same sort of sentiment classification techniques and approaches are not implementable on the data of various domains (Thelwall & Buckley, 2013).

Not all but most of the research studies have been conducted for the sentiment analysis of products and services. In the detailed review (Feldman, 2013; Medhat et al., 2014) of sentiment analysis studies, it is evident that most research is conducted for product and services. Karamibekr and Ghorbani (2012a) found that sentiment analysis of social issues is different than sentiment analysis of product and services in certain aspects. A social issue is different from product and service in a way that social issue is related to other sub-issues, on the other hand product or service is connected to features. Furthermore, for product and service, users express their sentiment or opinion by using adjectives and intestifiers while verbs and adverbs are moslt used in case of social issues (Karamibekr & Ghorbani, 2012b). Moreover, minimal literature is available for the sentiment analysis of social issues. These facts encourages to address the research area and to enhance the sentiment analysis in the domain of social issues by exploiting sentiment lexicon.

1.4 Research Objectives and Questions

The research work described in this thesis is directed at following objectives with the associated research questions:

Objective 1: To enhance lexicon-based sentiment analysis for social issue.

The main objective of the research is to enhance the lexicon-based sentiment classification of social issues. Following are the research questions associated with this research objective.

Research Question 1: What approaches can be used to improve lexicon-based sentiment analysis?

The first research question is mainly associated with the sentiment lexicon in the lexicon-based sentiment analysis. As a sentiment lexicon is required in the lexicon-based approach of sentiment analysis, improved General Inquire will be used for the purpose which was initially developed to assist social science applications. Some domain dependent words which are not included in the General Inquirer will be added in the respective positive or negative categories based on the manual labeling by domain experts.

Research Question 2: How dependency grammar with the verb improve sentiment analysis of social issues?

As verb play an import role in the sentiment analysis of social issues, so second research question is related to explore the effect of dependency grammar of verb on the sentiment analysis of social issues.

Objective 2: To assess the effectiveness of the proposed approach.

The second objective of the research presented in this thesis is to assess the effectiveness of the proposed solution. Research questions relevant to this objective are as follow:

Research Question 3: How the proposed approach can be evaluated in terms of its effectiveness?

For the evaluation of proposed approach, the benchmark evaluation metrics precision, recall, accuracy and F1 score are required which will be addressed by answering research question three.

Research Question 4: How the performance of proposed approach can be compared with the online sentiment analysis tools?

Research question four is also concerned to assess the effectiveness by comparing the performance of the proposed solution with the online sentiment analysis tools. It explores the performance based on the quantitative results from the proposed and existing online sentiment analysis tools such as Alchemy, uClassify, AiApplied etc.

1.5 Research Methodology

Research methodology adopted in this research work is graphically presented in Figure 1.1. The study is comprised of following steps:

i. Investigation of research problem by studying existing approaches for the sentiment analysis and identifying aspects which affect the performance of sentiment analysis especially in the domain of social issues.
ii. Formulation of research objectives and defining relevant research questions for each objective.
iii. Design and development of proposed system to enhance sentiment analysis of social issues. The role of verb with the grammatical dependencies is incorporated with the use of improved sentiment lexicon General Inquirer to attain the purpose.
iv. Evaluation of proposed solution with reference to benchmark metrics precision, recall, accuracy and F1 score by utilizing manually labeled data by the domain experts. Moreover, comparison with some exiting online sentiment analysis tools to further asses the effectiveness of the proposed solution.
v. Details of research findings with comparison of quantitative results.

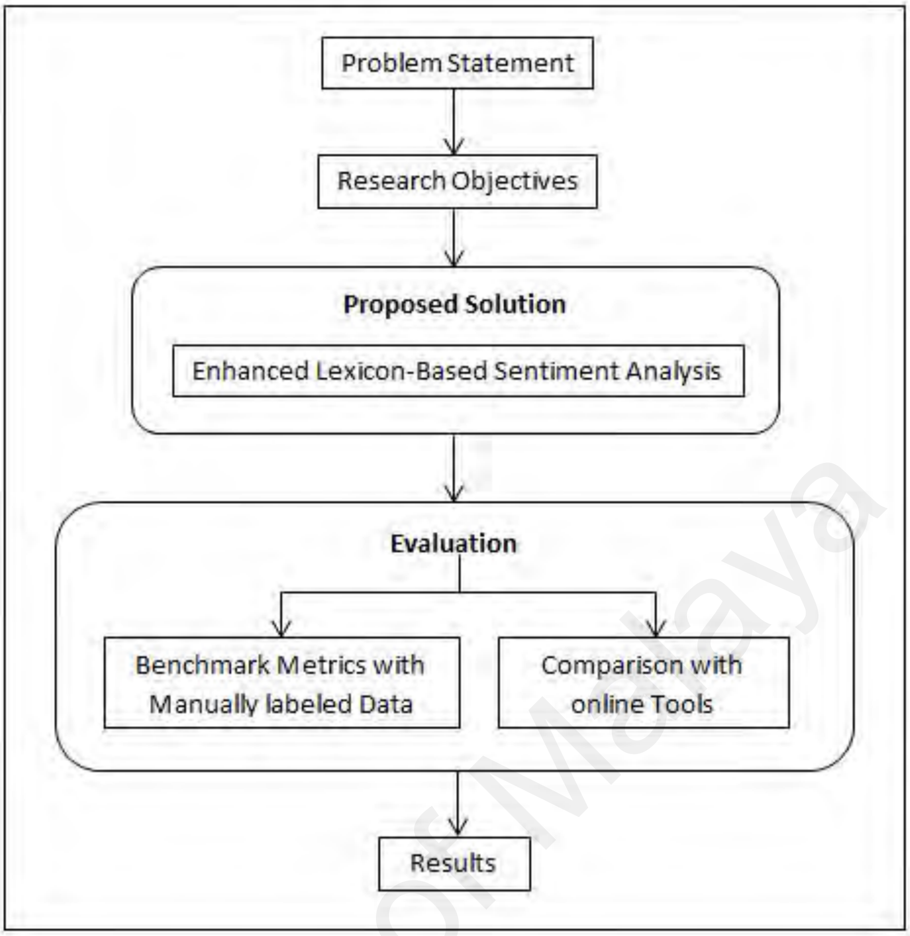

Figure 1.1: Research Methodology

1.6 Thesis Outline

This thesis is structured in five chapters that include introduction, literature review, research methodology, results and discussion, conclusion, limitation and future work. The detailed outline of the thesis is as follows:

Chapter 1 introduces the research work presented in this thesis. The overview of the background is briefly introduced with general thoughts about sentiment analysis and other key concepts of the research area. It also states the problem statement, research objectives and research questions. Furthermore, the research methodology of this research is presented and highlighted the layout of the thesis.

Chapter 2 presents a literature review of previous work related to sentiment analysis and research conducted in this thesis. A comprehensive review of sentiment analysis approaches is described. Lexicon-based methods are discussed in details as these are closely related to this research work. Moreover, the sentiment analysis in the domain of social issues and its differences from sentiment analysis of product and services are highlighted.

The deployment of proposed system of enhanced approach in lexicon-based sentiment analysis of social issues is described in the Chapter 3. It presents details of datasets and the experimental setup in the research methodology. Comparison with existing online sentiment analysis tools and benchmark metrics calculations with the help of manually labeled data by the domain specialist are also illustrated in this chapter for the evaluation of the proposed system.

Chapter 4 presents the main findings of this research. Step by step improvement of the proposed technique by incorporating different aspects is described. Experimental results achieved evaluated by benchmark metrics precision, recall, accuracy and F1 score are discussed in detail. Moreover, the effectiveness of the system is highlighted by comparing proposed technique with the online sentiment analysis tools.

Chapter 5 finally concludes the research work presented in this thesis. The chapter also states research contribution, limitations and explores avenues for the future work.

CHAPTER 2: LITERATURE REVIEW

In Chapter 1 introduction of the research work is described which is presented in this thesis. This chapter provides an overview of existing work that is related to sentiment analysis in Section 2.1. In Section 2.2, 2.3 and 2.4 existing techniques and methods are presented for machine learning, lexicon-based and hybrid approaches of sentiment analysis respectively. Section 2.3 also states about different sentiment lexicons available for the lexicon-based sentiment analysis methods. In Section 2.5, existing work that is relevant to sentiment analysis is mentioned specifically, in which Twitter is used as a dataset. Literature about applications of sentiment analysis in different domains is described in Section 2.6. Finally, the summary of this chapter is highlighted in Section 2.7.

2.1 Sentiment Analysis

With the dramatic growth of user-generated content especially on social media, a valuable source of information is available that can be used by researchers and industry practitioners. Individuals or organizations can exploit this online available information to make better decisions (Pang & Lee, 2008a; Ravi & Ravi, 2015). However, extraction of information from relevant sources, identification of sentiment and summarization into properly understandable form are difficult tasks for humans. Therefore, sentiment analysis has emerged to tackle all these challenges (Angulakshmi & ManickaChezian, 2014). Sentiment analysis, also known as opinion mining is an application of data mining, natural language processing (Jurafsky & Martin, 2014), computational linguistic (Nichols & Warnow, 2008; Rosenthal et al., 2017) and text analytics (Gandomi & Haider, 2015; Wilcock, 2009).

Sentiment analysis is mainly classified in three levels which are document level, sentence level and aspect level (Medhat et al., 2014). In document level sentiment analysis, a document is considered as an entity of opinion or sentiment. Sentence level sentiment analysis deals with the sentiment expressed in each sentence. If the different sentences in the document are related to multiple features or entities, then sentence level is more appropriate than document level (Serrano-Guerrero et al., 2015). Nevertheless, there is not much difference between document level and sentence level sentiment analysis as sentences are also short documents (Liu, 2012).

There is an remarkable feature of the document level sentiment analysis that one document may consist of sub-documents like paragraphs and sentences with different and sometimes opposite sentiments, and the overall sentiment of the document is the function of set of sentiments at sentence level (Pang & Lee, 2008a; Tang, 2015). Document level sentiment analysis assumes that document is an opinion about an entity or aspect in it (Tang, 2015; Yessenalina et al., 2010; Zhang et al., 2009). It is explored that sentence level sentiment analysis is more precise than document level sentiment analysis (Araujo et al., 2016; Conrad & Schilder, 2007). Aspect level sentiment analysis is more complex as it depends on the identification of entities and relevant aspects on the initial step (Medhat et al., 2014; Ojokoh & Kayode, 2012; Schouten & Frasincar, 2016). Based on these aspects, sentence level approach is more suitable which is not complex as compared to aspect level and more precise in terms of results, that‟s why same approach has been used in this research.

In the last more than one and half decade, there is an increased trend in the research of sentiment analysis and opinion mining. Pang and Lee (2008a) presented a comprehensive survey of more than three hundred studies to cover the techniques and approaches related to opinion-oriented information-seeking systems. They covered major tasks of opinion mining and sentiment analysis which include opinion extraction, sentiment polarity, sentiment classification, summarization, etc. In another survey conducted by Liu (2012), applications and major challenges in sentiment analysis were focused and also the techniques used to solve different problems in sentiment analysis. Medhat et al. (2014) covered most famous sentiment analysis techniques and application with the categorization of techniques as well. They also discussed other sentiment analysis related fields like emotion detection, building resources and transfer learning.

Ravi and Ravi (2015) also performed a rigorous survey on sentiment analysis and opinion mining regarding tasks, approaches and applications. They organized the earlier studies on the basis of sub tasks, machine learning and natural language processing technique and applications of sentiment analysis. Serrano-Guerrero et al. (2015) provided an insight for researchers and other users by reviewing and comparing the web services of sentiment analysis. They analyzed the online sentiment analysis tools for their capabilities of classification under different circumstances and presented the results which are useful for the users to decide about the appropriate service with the expected results. Giachanou and Crestani (2016) conducted a survey specifically for the sentiment analysis in Twitter. They briefly provided an overview for the algorithms used and categorized the studies based on the different approaches used for the sentiment analysis in Twitter. These studies had reviewed the analytical approach of different online sentiment analysis tools. It directed this research work towards the comparison of online sentiment analysis tools with the proposed technique to assess its effectiveness.

As mentioned earlier in Section 1.2.1 that sentiment analysis approaches are divided as machine learning, lexicon-based and hybrid approaches. In the following sections, existing studies are highlighted based on the main approaches of sentiment analysis.

Section 2.2 summarizes earlier studies based on supervised and unsupervised techniques of machine learning approaches.

2.2 Machine Learning Approaches

In the machine learning approaches, supervised learning methods have been mostly used in sentiment analysis. In these approaches, a pre-labeled training dataset is used for the learning of classifier based on specific mechanism (Feldman, 2013; Prabowo & Thelwall, 2009). In supervised and unsupervised methods of machine learning, extraction of proper features is very important in the success of classifier. In this process, natural languages processing techniques play an important role in some features which includes term frequency, parts of speech information, negation and syntactic dependencies (Medhat et al., 2014; Serrano-Guerrero et al., 2015).

Kennedy and Inkpen (2006) used the term frequencies for the feature selection of the classification of movie review. They also inspected the effect of valence shifter like negation, intensifiers and diminishers on the classification and showed increase in accuracy. Dadvar et al. (2011) also studied the effect of negation in the sentiment analysis in their term frequency based classifier. In the term frequency, terms may be uni-gram, bi-gram or n-gram for higher order. The researchers used movie review dataset with the term frequency in their experiment. Deng et al. (2014) proposed a supervised approach by using term weighing scheme in sentiment analysis. The authors inferred by conducting an experiment with three real time datasets that proposed technique outperformed unsupervised approaches. In a study by Pang and Lee (2008b), they claimed that uni-gram perform well as compared to bi-gram in the sentiment analysis of movie reviews. The most commonly used feature selection statistical methods include Latent Sematic Index (LSI), Point-wise Mutual Information (PMI), Chi-square, etc. Parts of speech information is also important in the sentiment analysis as it helps in finding verbs, adverbs, nouns, etc. which are useful for feature selection (Cambria et al., 2013; Pang & Lee, 2008a).

In the supervised learning methods of machine learning, the most frequently used classifiers are Naïve Bayes, Support Vector Machines (SVM), Decision Tree, Neural Network, Maximum Entropy etc. SVM works on the concept of separating hyperplanes placed between the instances of different classes. An optimal hyperplane is defined based on the labeled training dataset which is used to classify the new instances (Gautam & Yadav, 2014; Joachims, 1998). (Pang et al., 2002) applied SVM, Naïve Bayes and Maximum Entropy for the first time in machine learning for the binary sentiment classification of movie reviews and explored SVM with high accuracy using uni-gram features.

Naïve Bayes is the probabilistic classifier based on Bayes‟ theorem with assumption of strong independence between features. Naïve Bayes classifier builds the statistical model on the basis of training dataset and then utilizes this model to predict the class of new instances (Duda et al., 2001; Witten et al., 2016). Kang et al. (2012) proposed two enhanced Naïve Bayes methods of sentiment analysis for user reviews about restaurants and recommended Naïve Bayes with better accuracy than baseline methods. The authors used the proposed algorithm with unigram and bigram as features and their custom sentiment lexicon for restaurant reviews with an improvement in performance of almost 10% in recall and 29% in precision as compared to SVM. Like Naïve Bayes classifier, Decision Tree also utilizes the training data to build classifier by hierarchical decomposing training data based on the attribute value. The splitting process is done recursively until all instances in the subset belong to the same class (Duda et al., 2001; Witten et al., 2016). In a study of sentiment analysis for Roman-Urdu, (Bilal et al., 2016) presented that Naïve Bayes outperformed Decision Tree and KNN. They extracted the opinionated data from the blogs and used three above mentioned classification models in Waikato Environment for Knowledge Analysis (WEKA). According to their experiment, Decision Tree is the fastest classifier as there is not much process of calculation required and results from Decision Tree can be more accurate if larger training dataset is used.

Xia et al. (2011) used Naïve Bayes, maximum entropy and SVM classifier with wide range of comparative experiments on five different datasets of product reviews related to book, DVD, electronics and kitchen. The researchers conducted the experiments with two different set of feature of sentiment classification which are parts of speech and the word relation based feature sets. They found that by combining both feature sets with the classification algorithms together produced more accurate classification results. Dang et al. (2010) also used SVM for the sentiment analysis with different feature selection methods. They used online product reviews with combination of features like content free, content specific and sentiment features by using parts of speech and sentiment score. The accuracy of their proposed solution was almost comparable to earlier studies with the added features of cross validation and usage of stop words and filtering of features conditions.

In all supervised methods of machine learning approaches, initial labeled dataset is required to train the classifier. Whereas, unsupervised or semi-supervised machine learning techniques are used when it is not possible to get the labeled training data (Xianghua et al., 2013). This is the main drawback of supervised machine learning approaches as in large datasets it is very difficult and time consuming process to get the labeled data. Due to the same reason, this approach was not considered in this study.

2.3 Lexicon-based Approaches

In the lexicon-based approach, a sentiment lexicon is used containing words with their polarity values. Lexicon-based approaches are roughly divided into dictionary-based and corpus-based methods as mentioned earlier in Section 1.2.1.2. Three existing baseline scoring strategies for lexicon-based sentiment analysis are used in the literature i.e. term counting, maximum score and average score. In the term counting method, all the terms related to a specific category in the text are counted. As a result the text is categorized under the category with higher value of count (Ohana et al., 2011). Maximum score method classify the text with the class of term having maximum polarity value (Thelwall et al., 2012). In the average score technique, sentiment class of the text is calculated by the average of polarity values of all the terms (Bhadane et al., 2015). All these baseline scoring strategies can be used in the lexicon-based sentiment analysis with the same effect according to the required scenario. In this research work, maximum score method is adopted as scoring strategy by using positive, negative and neutral classes.

There are a lot of generic sentiment lexicons available with different format and size to help in classification of positive and negative sentiment in the text (Cho et al., 2014). SentiWordNet 3.0 is the latest version of SentiWordNet proposed by Esuli and Sebastiani (2007) which is based on the very famous English lexical database WordNet (Miller et al., 1990). OpinionFinder is a manually annotated lexical resource which is part of an online system to detect the sentiment within a document (Wilson et al., 2005). AFFIN Lexicon was developed to focus on the text from micro-blogging hence including many commonly used slang words as well (Nielsen, 2011). AFFIN Lexicon was based on ANEW Lexicon proposed by Bradley and Lang (1999) which provide emotional rating of the words.

General Inquirer is a manually annotated category based lexicon developed by Stone et al. (1966) which was initially intended to assist social science applications (Ohana et al., 2011). General Inquirer has 182 categories and each category contains some words based on four sources of categorizations i.e. Harvard IV-4 dictionary, Lasswell value dictionary, Marker categories, Semin and Fiedler categories (Semin & Fiedler, 1988), which are created manually. „Positive‟, „Pstv‟, „PosAff‟, „Pleasure‟, „Virtue‟, „Complete‟ and „Yes‟ categories cane be used as positive, while „Negative‟, „Ngtv‟, „NegAff‟, „Pain‟, „Vice‟, „Fail‟, „Negate‟ and „No‟ are some negative categories (Cho et al., 2014). The same sentiment lexicon is used in this study as well, because initially it was intended to be used in social science applications and the data used in this research is related to a social issue.

In the domain-dependent sentiment analysis, it is quite challenging to use generic sentiment lexicons (Thelwall & Buckley, 2013). However, as usually domain-dependent sentiment lexicons are not readily available, so have to be generated specifically if required. A sentiment lexicon for health is entirely different than sentiment lexicon for politics. Zhou et al. (2014) proposed topic-based lexicon expansion to overcome the polarity issues of misspelled and abbreviations in the sentiment analysis on Twitter. They expanded the general lexicon with the domain dependent words and the abbreviations with the emoticons as well. The researchers compared the proposed approach of expanded lexicon with the SentiStrength which lead to higher performance in classification. The same concept has been used in the proposed method to improve the sentiment lexicon General Inquirer with the domain related terms which are missing. Godbole et al. (2007) generated different lexicons based on WordNet in their research of sentiment analysis of news and blogs for different topics like business, crime, general, health, media, politics and sports. The researchers proposed a technique which assigns positive or negative scores to each entry in the text corpus. They evaluated the significance of scoring mechanism based on statistical evidences.

Chaumartin (2007) used WordNet, SentiWordNet and WordNetAffect as sentiment lexicons in the rule-based method to detect emotions in the news headlines. The proposed algorithm identifies opinionated words and annotates them by using pre-defined list of emotions and then categorize as positive or negative class. The researchers concluded high accuracy results on emotion and valence annotation. Qiu et al. (2010) used dictionary-based approach in the contextual advertisement for the sentiment analysis. The authors proposed a rule-based technique to retrieve opinionated topic words related to negative sentiment by using syntactic parsing and sentiment lexicon. Contextual semantic in different context perform an important role in the better performance of sentiment classification (Cho et al., 2014; Saif et al., 2016). Neviarouskaya et al. (2011) created a sentiment lexicon SentiFul by improving SentiWordNet for better performance. The researchers expanded it further by using word‟s synonyms and antonyms associations. They also utilized General Inquirer in the second phase of evaluation and achieved improved accuracy.

Taboada et al. (2011) proposed Semantic Orientation Calculator (SO-CAL), a binary-classifier to recognize the semantic orientation from text. They also incorporated negation, intensifiers and diminishers to increase the accuracy of classifier. The proposed approach performed well across different domains and for unseen data as well. Cho et al. (2014) introduced a data-driven approach by merging different sentiment lexicons and adjusting the prior polarity of the dictionaries according to domain specific data to handle the contextual polarity problem. The researchers integrated ten sentiment lexicons including AFINN, General Inquirer, Micro-WNOp, Opinion Lexicon, SenticNet, SentiSense, SentiWordNet, SO-CAL, Subjectivity Lexicon and WordNet-

Affect by using their proposed merge, remove and switch operations. They inferred that integrated dictionary approach outperformed the individual sentiment dictionaries in the sentiment classification of product reviews from different domain like books, smartphones and movies.

Gitari et al. (2015) created a lexicon-based classifier to detect hate speech from blogs and forums by using sentiment analysis techniques. The researchers crawled 100 theme based blog postings from ten different websites related to nationality, religion and ethnicity. They created a lexicon and a model classifier for hate speech detection with 70% performance. Ngoc and Yoo (2014) proposed a lexicon-based approach of sentiment analysis to incorporate sentiment within comments text along with user engagement parameters for the improved and more accurate Facebook fan page rankings. They used Social Packets Crawler to get real Facebook dataset for the proposed method. Asghar et al. (2016) created a domain dependent health-related lexicon by using their proposed hybrid approach for the more accurate classification in this domain. They used dataset of user reviews from the health related blogs and implemented hybrid approach by combining boot strap and corpus based techniques and improved the accuracy of classifier to 89% as compared to earlier studies with 76% and 78% accuracy. Beside English, voluminous research of lexicon-based sentiment analysis is being conducted for other dialect as well (Avanço & Nunes, 2014; Duwairi et al., 2015; Syed et al., 2010).

2.4 Hybrid Approaches

There are some methods which use combination of machine learning and lexicon-based techniques and are called hybrid approaches. The use of hybrid approaches is less frequent due to its computational complexities (Medhat et al., 2014). Pak and Paroubek (2010b) used two discriminatory-word lexicons with the Naïve Bayes classifier in their hybrid approach of a Twitter based system. The researchers collected the corpus by using Twitter API based on positive and negative emoticons. They used the Naïve Bayes classifier with parts of speech and n-gram features which resulted in higher accuracy. In another hybrid approach of sentiment analysis, Prabowo and Thelwall (2009) combined supervised, machine learning and rule-based methods. In their semi-automatic proposed approach, each classifier can contribute to the other classifier to attain the improved effectiveness in terms of precision and recall. The researchers used sentiment lexicon General Inquirer with rule-based and statistic-based classifiers along with SVM approach by utilizing the datasets from movie and product reviews.

Another cross-domain sentiment lexicon was proposed by Weichselbraun et al. (2013) which can be utilized in wide range applications of sentiment analysis. The authors used hybrid approach by combining lexicon analysis and machine learning to resolve the ambiguity of context in sentiment words. They used product and movie reviews for the evaluation of their proposed approach. Mukwazvure and Supreethi (2015) used sentiment lexicons for the polarity calculation of news comments and then trained the machine learning algorithm based on lexicon-based method"s results. They used the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for training the classifier and inferred that SVM performed well as compared to KNN.

Mukwazvure and Supreethi (2015) used hybrid of semantic orientation and neural network based approach to classify the sentiment from blogs on products. The proposed approach used five datasets and all resulted in good performance. Datasets from other social media like Twitter, Plurk, Facebook etc. can also be utilized in proposed strategy. Ghiassi et al. (2013) introduced a hybrid approach by utilizing a Twitter-specific lexicon and a comparable sentiment classification model using SVM and dynamic artificial neural network. The researchers demonstrated that dynamic artificial neural network yield better accuracy of sentiment classification than SVM and by using the same Twitter-based lexicon. Khan et al. (2014) presented a hybrid approach for Twitter feeds classification. Their proposed technique overcome the previous issues in classification accuracy, data sparsity and sarcasm and showed better accuracy when compared with similar techniques. Another hybrid approach using natural language processing (NLP) techniques, an enhanced sentiment lexicon based on SentiWordNet and fuzzy sets was proposed by Appel et al. (2016). The authors validated in their experiments with three datasets that proposed solution is more precise and accurate as compared to Maximum Entropy and Naïve Bayes. Although due to computational complexities, it is difficult to use hybrid approach in the sentiment analysis applications, but it provides more accurate and precise results as it uses the strength of machine learning and lexicon-based approaches collectively.

2.5 Twitter as Dataset in Sentiment Analysis

Micro-blogging platform Twitter has become the most frequently used social media for people to express their opinion and sentiment. With almost 500 million tweets per day (Haustein et al., 2016), Twitter has emerged as the gold-mine for researchers to analyze user"s opinion. This section specifically describes some sentiment analysis studies based on the data from Twitter. The current research also utilized data from Twitter for a social issue of illegal immigration.

(Pak & Paroubek, 2010a) proposed a system to automatically collect corpus from Twitter for sentiment analysis. With the linguistic analysis of collected data, they developed a sentiment classifier to determine tweets as positive, negative or neutral. In the prediction models, there are many studies based on Twitter data. (Tumasjan et al., 2011) investigated the Tweets to forecast 2009 German federal elections. They used the positive and negative tweets of party profiles to reflect their political position. Bollen et al. (2011) used the Twitter data to forecast the stock market based on the mood of public. The researchers inferred the increase in accuracy of predictions of stock market index by analyzing the Twitter‟s text. They used OpinionFinder to measure positive and negative moods, and Google-Profile of Moods State (GPMOS) to detect 6 mood states like calm, alert, sure, vital, kind and happy. (Rui et al., 2013) studied the influence of tweets on movie sales and important managerial implications. The authors found that positive word of mouth is connected with higher movie sales while lower sales are the impact of negative word of mouth. They utilized the publically available data from Twitter‟s word of mouth, details of the various followers of users and data of movie sales was analyzed via machine learning approaches to examine the pattern of movie sales.

Pete Burnap et al. (2013) investigated Twitter‟s data to identify tension by using machine learning, syntactic and lexicon-based text mining and sentiment analysis techniques. The researchers explored possibility of predicting spikes in social tension in online communities by using data from social media. Khan et al. (2014) presented an algorithm with improved accuracy for tweets classification by using a hybrid sentiment analysis approach. The researchers presented an improved algorithm to overcome the primary issues of previous techniques like low classification accuracy, data sparsity, sarcasm and high percentage of incorrect classification as neutral. They also included different pre-processing steps before classification. In another study of prediction from tweets, Mohammad et al. (2015) utilized 2012 US presidential election tweets to predict emotion and purpose from unseen tweets by using sentiment analysis techniques. The researchers automatically annotated a dataset from 2012 US elections by crowdsourcing and presented a classifier to forecast the sentiment, emotion, purpose and style in tweets.

Onifade and Malik (2015) presented a tool called Sentiment Analyzer for Social Media (SASM) that automatically fetched the opinions and analyzed their sentiments. The authors detected and removed spam, identified sarcasm and provided improved sentiment classification by using lexicon-based fuzzy linguistic hedges. They created EmotiSentiWordNet, based on the polarity scores from SentiWordNet to evaluate the sentiment classification. Yu and Wang (2015) examined the sentiment of US soccer fan"s real time tweets using Twitter search API during the games of 2014 FIFA World Cup. The authors used sentiment analysis techniques to reveal that sport fans express their sentiments via Twitter by exploring change in emotions especially after goal. Zavattaro et al. (2015) tried to explore that if positive sentiment of tweets effect the participation of citizen with government via social media. The authors used sentiment analysis to identify the tone of government agencies as well as the reaction of citizens to government policies. The current research also aims to provide the categorization of tweets based on sub-issues which may be helpful for the government and other social organizations in the policy making decisions. Wong et al. (2016) described a modified classifier for the public sentiment for breast cancer. They investigated and presented a prominent relationship between Twitter sentiment and screening uptake at state level for breast cancer. Kolchyna et al. (2015) covered lexicon-based and machine learning sentiment analysis techniques for Twitter. With the comparative analysis, they showed that enhanced sentiment lexicons with emoticons, slangs and abbreviations increase the accuracy of Twitter sentiment analysis. Same approaches of improved sentiment lexicon with slangs and abbreviations dealing in the data pre-processing are followed in the current research work for the sentiment analysis with improved accuracy.

2.6 Sentiment Analysis Applications

Sentiment analysis is a discipline that is widely used in many applications. This section highlights the application of sentiment analysis in different domains.

Sentiment analysis of product and services is the most common application. There are many online systems available which provide a summary of reviews about product and services based on the research work of sentiment analysis (Jin et al., 2009; Liu et al., 2005). By analyzing the sentiment from social media, the possible outcome of any event can be predicted (Bollen et al., 2011; Rui et al., 2013; Tumasjan et al., 2011). Sentiment analysis applications are also used in the reputation monitoring on social media (Ngoc & Yoo, 2014). Decision making system also applies the sentiment analysis as important factor which are used in policy making and other relevant activities (Hridoy et al., 2015; Khan et al., 2014; Onifade & Malik, 2015; Wong et al., 2016; Zavattaro et al., 2015). So overall sentiment analysis applications are in every field of life like business (Bollen et al., 2011; He et al., 2013; Rui et al., 2013), health care (Rodrigues et al., 2016; Wong et al., 2016), politics (Mohammad et al., 2015; Tumasjan et al., 2011; Zavattaro et al., 2015), sport (Yu & Wang, 2015), etc.

In the sentiment analysis research, most part has been directed on product and services. However, sentiment analysis of social issues varies from sentiment analysis of social issues in certain aspects (Karamibekr & Ghorbani, 2012a). Also, minimal literature is available for the sentiment analysis of social issues. Hence, investigation of new techniques and approaches in this area is vital and it directed to conduct research in this domain.

2.7 Summary

This chapter has presented an extensive literature review for the research work described in this thesis mainly focusing on sentiment analysis and existing approaches. The review considered briefly the machine learning and lexicon-based approaches with famous available sentiment lexicons. Since the current study utilized Twitter dataset, so we also discussed the studies based on Twitter dataset for the sentiment classification.

Finally, this chapter also reviewed the related work on sentiment analysis applications in different domains.

The next chapter introduces the research methodology in this research work for the deployment of proposed system of enhanced approach in lexicon-based sentiment analysis of social issues with the evaluation methods to assess its effectiveness.

CHAPTER 3: METHODOLOGY

This chapter introduces the methods and procedures to address the deployment of the proposed solution. It is comprised of five sections covering different phases of implementation from data collection to evaluation. In Section 3.1, data collection and filtering process is described in detail. Data pre-processing techniques applied on the collected data are highlighted in Section 3.2 along with manual labeling of tweets. The main experiment of enhanced sentiment analysis approach which incorporates the usage of verb, sentiment lexicon, grammatical dependencies and other important aspects is presented in Section 3.3. The categorization of data based on the sub-issues related to the main issue is also described in Section 3.3. In Section 3.4, benchmark metrics are defined for the evaluation of the proposed solution, also the performance of ten online sentiment analysis tools is compared to assess the effectiveness of the proposed technique. Finally, the chapter is summarized in Section 3.5.

The overall process flow of the proposed solution is given in Figure 3.1.

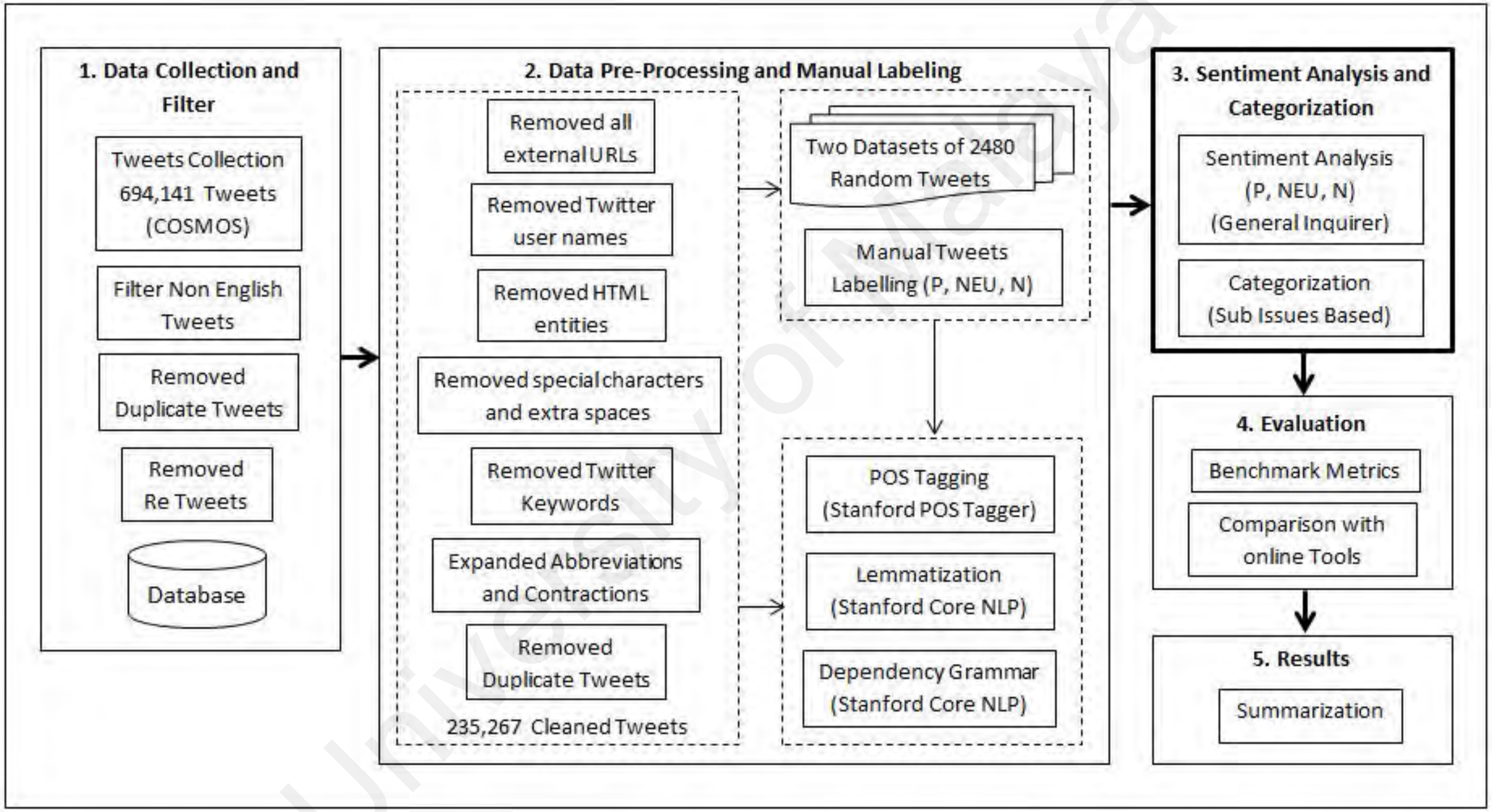

Figure 3.1: Overall Process Flow

3.1 Data Collection and Filter

The initial phase is data collection in the proposed technique. Tweets were collected for particular keywords from October 2015 till December 2015 to have ample dataset for this research. As the current study focus on social issues for the domain of illegal immigration, so „Illegal immigration", „illegal immigrant", „illegal alien", „undocumented immigrant" and „illegals" were keywords utilized as criteria for data collection.

Twitter is the well-known source of corpus for opinion mining with almost 500 million tweets per day which provide APIs for data collection (Haustein et al., 2016; Pak & Paroubek, 2010a). Twitter search API[1] enables the retrieval of past tweets as per request criteria but max to 14 days prior. Concurrently Twitter streaming API[2] provides continued live stream of new tweets according to user determined parameters like keyword, people, timeframe etc. We gathered 694,141 tweets with varieties of the above mentioned keywords in three months utilizing the Cardiff Online Social Media Observatory (COSMOS) (Peter Burnap et al., 2015) tool which works on the Twitter streaming API. After data collection, tweets were filtered based on the English dialect and all non-English tweets were expelled. After exporting the only English tweets from COSMOS tool in excel format, all duplicate tweets were expelled prior to data cleaning as highlighted in Figure 3.1. Subsequently all re-tweets starting with keyword RT (re-tweet) were expelled, as they merely refer to re-posting of the same tweets from the user"s own or other users" tweets.

[1] https://dev.twitter.com/rest/public/search (Accessed October, 2015)

[2] https://dev.twitter.com/streaming/overview (Accessed October, 2015)

3.2 Data Pre-Processing and Manual Labeling

Data from online stages are a rich wellspring of research and learning revelation however in the meantime it is very noisy and require broad cleaning for opinion mining (Dey & Haque, 2009). Pre-processing is an essential action in opinion mining to improve the exactness of results. Unwanted text filtering, stop words removal, negation handling, part-of-speech tagging, stemming and lemmatization are distinctive methodologies in pre-processing (Balakrishnan & Lloyd-Yemoh, 2014; Haddi et al., 2013; Sun et al., 2014; Uysal & Gunal, 2014).

Detailed data cleaning steps were performed in the pre-processing stage as follows:

i. Distinguished all the URLs of external links, twitter pictures and video URLs with regular expressions and expelled them from tweets.
ii. Distinguished and expelled all twitter user names specified with @username.
iii. Distinguished and expelled the HTML special entities like „&‟, „<‟ and „>‟ etc.
iv. Recognized and expelled the twitter specific keywords like via, and modified tweet (MT).
v. Recognized and expelled the special characters which are not being used in emoticons like #.
vi. Expelled the additional white spaces.

Twitter is restricted to 144 characters, so there is frequent utilization of abbreviations, slangs, acronyms and English contractions. In the lexicon-based approach, as the polarity of the terms is calculated based on the predefined polarity in the sentiment lexicon, so abbreviation, slangs or other short form of words leads to wrong classification (Zhou et al., 2014). In the pre-processing phase, abbreviations and emoticons are extended using SMS Dictionary[3] and NetLingo[4]. Some example of abbreviations and contractions are given in Table 3.1 below.

Table 3.1: Abbreviations and Contractions

Abbreviation	Expanded	Contraction	Expanded
B4	Before	That‟s, it‟s, what‟s	That is, it is, what is
b/c, bc, coz	Because	You‟re, we‟re, they‟re	You are, we are, they are
U	You	I‟m, we‟ve, we‟ll	I am, we have, we will
Frm, 4rm	From	Isn‟t, can‟t, don‟t	Is not, cannot, do not

The other challenge in the short content of social media are emoticons as "emotion signals" (Hu et al., 2013). In this dataset emoticons are exchanged with their respective emotions to be used further in the sentiment analysis. Table 3.2 lists a few emoticons in the corpus.

Table 3.2: List of Emoticons

Emoticons	Replacement
:-), :), :], :3, >:], :))	Smile
:-D, :D, 8-D	Laugh
:-(, :(, :[	Sad
:`-(, :*(, ;'(	Cry

[3] http://smsdictionary.co.uk (Accessed February, 2016)

[4] http://www.netlingo.com (Accessed February, 2016)

After applying pre-processing techniques on the filtered tweets for data cleaning, duplicate tweets from the same user account were also removed, resulting in 235,267 cleaned tweets. Table 3.3 portrays the dataset details.

Table 3.3: Data Collection Details

Month	TC	NET	DTBC	RT	DTAC	CT
October	203,186	6,433	95,622	23,457	5,747	71,927
November	285,913	12,774	135,573	32,125	10,786	94,655
December	205,042	11,105	94,756	23,470	7,026	68,685
Total	694,141	30,312	325,951	79,052	23,559	235,267

TC: tweets collected
NET: non-English tweets
DTBC: duplicate tweets before cleaning
RT: re-tweets
DTAC: duplicate tweets after cleaning
CT: cleaned tweets

3.3 Sentiment Analysis and Categorization

In the proposed technique different aspects were involved to calculate the sentiment of tweets specially the role of verb with grammatical dependencies and the improvement in the sentiment lexicon General Inquirer. Sentiment was calculated step by step by incorporating one aspect on each step to see its effect on the performance of classifier. As presented in Figure 3.2, all the subsequent aspects also include their predecessors as well to have a combined effect on the sentiment score calculation. As mentioned earlier in the objectives in Section 1.4, the enhancement in the proposed method is in step 2 (General Inquirer improvement) and step 4 (Grammatical Dependencies) which are further explained in the subsequent sections.

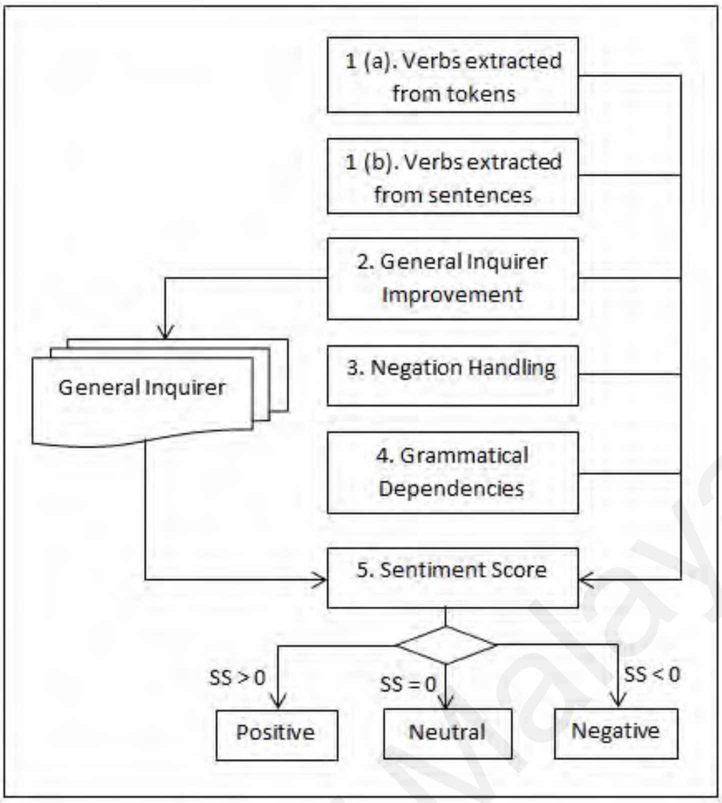

Figure 3.2: Sentiment Analysis Experiment

3.3.1 Usage of Verb

According to Karamibekr and Ghorbani (2012b), verbs play an important role in the sentiment analysis of social issues. After data cleaning, lemmatized verbs were extracted from the tokenized tweets. It was observed that verb extraction and hence sentiment analysis was quite improved when full sentences were used to extract verbs instead of only tokens. Stanford Core NLP tool (Manning et al., 2014) was used to extract lemmatized verbs.

3.3.2 Sentiment Lexicon

There are a lot of sentiment lexicons available with different format and size to help in classification of positive and negative sentiment in the text. Some of them are SentiWordNet, WordNet-Affect, AFFIN, Opinion Lexicon and General Inquirer etc. (Cho et al., 2014). In this study, we selected General Inquirer which was initially designed for social studies (Ohana et al., 2011). There are many categories available in this manually annotated lexicon (Stone et al., 1966). „Positive‟, „Pstv‟, „PosAff‟, „Pleasure‟, „Virtue‟, „Complete‟ and „Yes‟ categories were used as positive, while „Negative‟, „Ngtv‟, „NegAff‟, „Pain‟, „Vice‟, „Fail‟, „Negate‟ and „No‟ were used as negative categories (Cho et al., 2014). All positive terms in the aforementioned positive categories are assigned a sentiment score of 1, and -1 sentiment score is assigned to all words in the negative categories. Some domain dependent words which were missing in the lexicon were added in the positive and negative categories on the basis of manual labeling by domain experts (Taboada et al., 2011). The current study, for the first time used the General Inquirer for the sentiment analysis of social issues as highlighted in the phase 3 of sentiment analysis in Figure 3.1, which is also the main contribution of this study along with multi-level grammatical dependencies of the verbs mentioned in Section 3.3.4.

3.3.3 Negation Handling

Switch negation was used to handle the negation in the tweets. The obvious way to deal the negation is to reverse the polarity of the word (Taboada et al., 2011). For example support (+1) changes to (-1) if used as „do not support‟. In the third step, the effect of negation is not much prominent but as aforementioned that all the subsequent aspects also include their predecessors, so there was a positive effect of each aspect in the proposed technique. Negation terms were also identified by using Stanford Core NLP tool (Manning et al., 2014).

3.3.4 Grammatical Dependencies

Grammatical dependencies were handled in the proposed technique in which English universal grammar relations of the verbs were incorporated (De Marneffe et al., 2014; Yasavur et al., 2014). Multi-level grammatical dependencies with the verb were the main feature in the proposed technique which enhanced the classification remarkably as reflected in Figure 3.2 of sentiment analysis experiment. In this regard, not only direct dependencies of the particular verb were considered but also further n-level dependencies of those dependent objects were incorporated regardless of its type like verb, noun, adverb or adjective. For example, in Figure 3.3 verb “start” has open clausal complement (xcomp) dependency with the verb “deport” and then verb “deport” has direct object “illegals” which is noun, so these multi-level dependencies were handled with the combined effect on the polarity calculation. This technique is the main contribution of the proposed solution as highlighted in Algorithm 3.1 in Section 3.3.5, which increased the accuracy remarkably as mentioned in the step by step results of different phases in Section 4.1.

Direct object (dobj), nominal subject (nsubj), open clausal complement (xcomp), clausal complement (ccomp), adjective modifier (amod) and negation modifier (nmod) were extracted by using Stanford Core NLP tool (Manning et al., 2014) and utilized in the study. Figure 3.3 presents output from the Stanford Core NLP tool for parts of speech, lemmas and grammar dependencies for a tweet of our dataset.

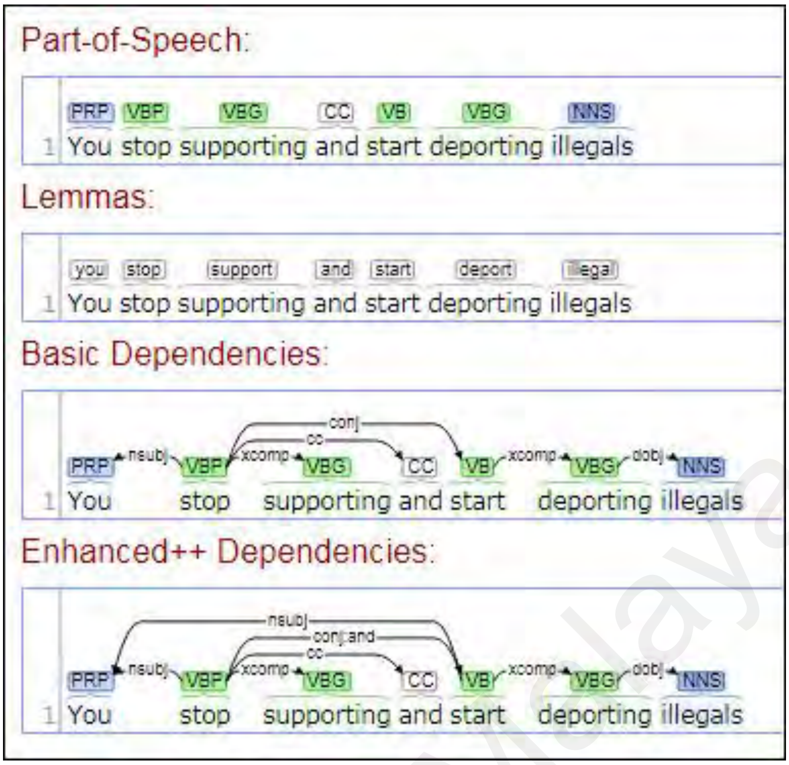

Figure 3.3: Output from Stanford Core NLP Tool

3.3.5 Sentiment Score

Sentiment score of each verb of the tweet is checked in General Inquirer and the effect of negation and multi-level grammatical dependencies were then implemented. The overall sentiment score of the tweet is calculated by adding the sentiment score of each verb as in Equation 1.

$$S_T = \sum_{i=1}^{n} S_i(V_i) \qquad (1)$$

S_T is the total sentiment calculated by sum of sentiment of all verbs ($S_i(V_i)$, i=1,…,n where n is the number of verbs) in the tweet. If the overall sentiment score is positive, then the tweet is classified as positive, and negative if the score is negative. Tweet is classified as neutral if the overall sentiment score is 0 (Bhadane et al., 2015).

The overall process for the sentiment analysis is presented in following algorithm.

<u>Algorithm 3.1: Sentiment Classification</u>

Input: Tweet

Output: Sentiment class

```
Sentiment_Score ← 0
Cleaned_Tweet ← CleanTweet (tweet)
POS ← GetPOS (Cleaned_Tweet)

Dependencies ← GetDependencies(Cleaned_Tweet)

  For every verb do
    Verb_Lemma ← GetLemma (Verb)
    Polarity ← GetPolarity (Verb_Lemma)

      If (verb has dependent) then
        For every dependent do
        Dependent_Polarity ← DependentPolarity(Verb, Dependent)
        End For
      End If

      If (verb has negation) then
      Polarity ← ReversePolarity (Polarity)
      End If

      Polarity ← CombinePolarity(Polarity, Dependent_Polarity)

      If (Polarity is positive) then
      Sentiment_Score ← Sentiment_Score + 1
      Else If (Polarity is negative) then
      Sentiment_Score ← Sentiment_Score - 1
      End If

  End For

  If (Sentiment_Score is positive) then
  Sentiment_Class ← Positive
```

Else If (Sentiment_Score is negative) **then**

Sentiment_Class ← Negative

Else

Sentiment_Class ← Neutral

End If

Return Sentiment_Class

The highlighted section in the above algorithm is the main area of contribution in this proposed approach.

3.3.6 Categorization

Identification and categorization of data in different domains depends on the perspective of that particular domain. Balakrishnan and Kaur (in press) classified the posts according to the categories used in the previous studies of the airline domain, which suits well the nature of their crawled posts. As described earlier in Section 1.2.2, a social issue is connected with other issues or sub-issues in contrast to the features for the product and services. Tweets in the dataset used in this research are generally categorized based on sub-issues using bag of words and term frequency approach (Bandhakavi et al., 2017). Terms were extracted from the complete dataset with their frequencies and then same type of terms was listed for main sub categories according to the opinion of domain experts and based on previous studies which discussed the social issues connected with illegal immigration. For example, frequently used terms including kill, crime, murder, rape, steal, fraud, etc. are grouped together for category „crimes". Similarly, other set of same type of terms are grouped for other categories as well. Crimes (Beare, 1997; Stupi et al., 2016), security (Beare, 1997; Chiricos et al., 2014; Lakoff & Ferguson, 2017; O'Doherty & Lecouteur, 2007), law and order (Lakoff & Ferguson, 2017; O'Doherty & Lecouteur, 2007) , government benefits (Chiricos et al., 2014; Stupi et al., 2016), racism (Stupi et al., 2016), terrorism (Lakoff & Ferguson, 2017), unemployment (Chiricos et al., 2014; Lakoff & Ferguson, 2017; O'Doherty & Lecouteur, 2007) and economic issues (Chiricos et al., 2014) etc. are some sub-issues based categories and tweets were classified based on the terms extracted for each category.

3.4 Evaluation

Proposed sentiment analysis solutions was evaluated by using benchmark metrics precision, recall, accuracy and F1 score with the help of manually labeled tweets by three domain specialists. Furthermore, based on these evaluation metrics the performance of the proposed solution was compared with online sentiment analysis tools to assess the effectiveness. The improvement in the accuracy by using proposed method as compared to online sentiment analysis tools showed the enhancement in the technique of sentiment analysis for social issues. Detailed results are presented in Section 4.1 and 4.2 which clearly describe the enhancements by using proposed technique.

With the dataset being significantly large and with the distinctive constraints that accompany evaluation online tools, it may take months to analyze all collected tweets. It is not important to utilize all tweets to analyze performance (Serrano-Guerrero et al., 2015). Hence we choose some tweets arbitrarily to analyze from the selected online tools. These selected tweets were additionally separated into two datasets containing 1,045 negative, 119 neutral and 76 positive tweets in each dataset to check the uniformity and reliability of results from the proposed solution and online tools.

3.4.1 Manually Labeled Tweets

To evaluate the performance of the proposed approach, manually labeled tweets by the domain specialists were used. Due to limited time and resources, randomly selected 2,500 tweets from the cleaned tweets mentioned in Table 3.3, were given to three linguistic and domain specialist to label as positive, negative or neutral. Hence 2,480 tweets were categorized as positive, negative or neutral based on the labeling and classification criteria as shown in Figure 3.4.

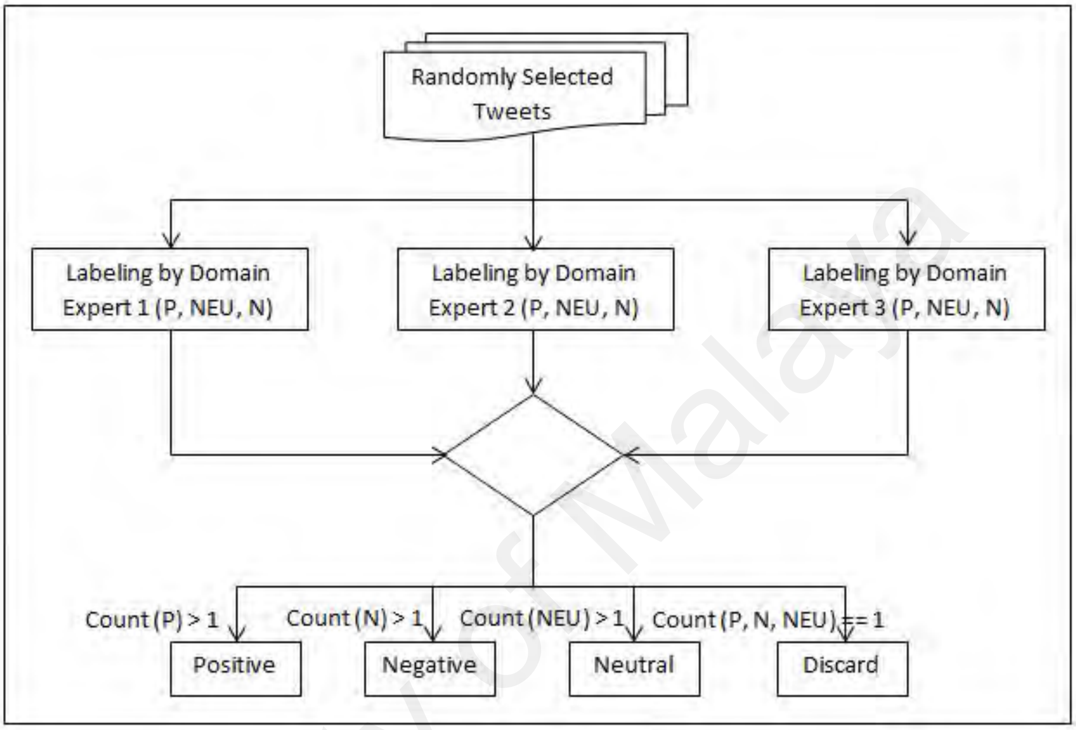

Figure 3.4: Manual Labeling of Tweets

20 labeled tweets were discarded due to lack of unanimity in categorization by the experts. The selected labeled tweets were randomly divided into two datasets to assure the consistency of the technique. These two datasets will be used as dataset 1 and dataset 2 in the rest of this thesis. The datasets contain cleaned tweets and the positive (P), negative (N) or neutral label (NEU) given by the experts which will be further used in the evaluation process along with the classification by the proposed solution. Please note that, both datasets contain same number of positive, negative and neutral tweets, the only difference is the text of tweets in the datasets. Table 3.4 shows sample data randomly from both the datasets.

Table 3.4: Sample Data

Cleaned Tweet	Label	Dataset
I can see a virtual Bankrupting of the USA just from hearing of illegals criminal alien promotion absolutely insanity	N	1
Clinton comes out in favor of healthcare for undocumented immigrants	P	1
Illegals from Mexico, Africa and China found in a truck in Texas	NEU	1
No amnesty for law breakers	N	2
Yeah right, Obama loves the illegals	P	2
What we think about the illegal alien in Mexico	NEU	2

3.4.2 Evaluation Metrics

Precision, recall, accuracy and F1 score were used as benchmark evaluation metrics.

Table 3.5: Example of Confusion Matrix

N = 100	Classified: Negative	Classified: Positive
Actual: Negative	Negative_Correct: 63	Positive_Wrong: 06
Actual: Positive	Negative_Wrong: 07	Positive_Correct: 24

Table 3.5 shows an example of confusion matrix (Maynard & Bontcheva, 2016) used for the calculation of evaluation metrics. In this context, a manually labeled dataset of 100 tweets with 69 negative and 31 positive tweets was assumed to explain the calculation of formulas given in subsequent sections.

3.4.2.1 **Precision**

Positive_Precision is the portion of positive classification which is true and computed as in Equation 2 (Serrano-Guerrero et al., 2015).

$$Positive\_Precision = \frac{Positive\_Correct\_Classified}{\text{Positive\_Correct\_Classifeid} + \text{Positive\_Wrong\_Classifed}} \quad (2)$$

Positive_Correct_Classified is the quantity of tweets which were marked as positive by the domain experts and classified positive by the proposed technique as well. Positive_Wrong_Classified is the quantity of tweets classified as positive by the technique however initially not marked as positive by the domain experts. Negative and neutral precision is figured by the similar formula by using correct and wrong classified tweets of negative and neutral class.

To calculate the Positive_Precision, as per confusion matrix shown in Table 3.5, Positive_Correct is 24 and Positive_Wrong is 6. Hence, with reference to Equation 2, Positive_Precision is 0.8 (24/(24+6) = 0.8) and by multiplying with 100, percentage becomes 80%. Similarly, to calculate Negative_Precision, Negative_Correct is 63 and Negative_Wrong is 7, so it becomes 63/(63+7) = 0.9 and multiplied by 100 turns it to 90%.

3.4.2.2 **Recall**

Positive_Recall is the portion of positive tweets caught the technique and computed as shown in Equation 3 (Serrano-Guerrero et al., 2015).

$$Positive\_Recall = \frac{Positive\_Correct\_Classified}{\text{Total\_Positive}} \quad (3)$$

Positive_Correct_Classified is the quantity of tweets which were marked as positive by the domain experts and classified positive by the proposed technique as well. Total_Positive is the total number of positive marked tweets by the domain experts in the dataset. Negative and neutral recall is also figured by the same formula by using respective values.

To calculate the Positive_Recall as per assumed data presented in Table 3.5, Positive_Correct is 24 and Total_Positive is 30. Thus, according to Equation 3, Positive_Recall is calculated as 24/31 = 0.77 which becomes 77% after multiplying with 100. Similarly Negative_Recall is calculated as 63/69 = 0.91 and 91% accordingly.

3.4.2.3 **Accuracy**

Accuracy is the portion of total correct classified tweets by the proposed technique over total tweets labeled by domain experts. Total_Correct_Classified is the quantity of tweets which were accurately classified as positive, negative or neutral tweets. Total_Tweets is the quantity of all labeled tweets in the dataset. With that, accuracy is computed as in Equation 4 (Serrano-Guerrero et al., 2015)

$$Accuracy = \frac{Total\_Correct\_Classified}{\text{Total\_Tweets}} \tag{4}$$

Based on the assumed data given in Table 3.5, accuracy is (63+24)/100 = 0.84 which means 84% of tweets are classified accurately.

3.4.2.4 **F1 Score**

The two metrics precision and recall are sometimes used together to calculate F1 score as a single measurement to evaluate the system. F1 score or F measure is the harmonic mean of precision and recall and computed as in Equation 5 (Kolchyna et al., 2015; Prabowo & Thelwall, 2009).

$$Positive\_F1\_Score \;= 2 * \frac{Positive\_Precision \; * Positive\_Recall}{Positive\_Precision + Positive\_Recall} \tag{5}$$

By using Positive_Precision and Positive_Recall calculated in the earlier sections for the same scenario, Positive_F1_Score is 2*(0.8*0.77/0.8+0.77) = 0.78 which means 78% performance of classifier based on combined effect of precision and recall.

3.4.3 Comparison with Online Sentiment Analysis Tools

To further evaluate and assess the effectiveness of the proposed solution, ten online sentiment analysis tools were used to classify the same manually labeled datasets. Some online tools were selected from previous studies (Abbasi et al., 2014; Serrano-Guerrero et al., 2015) which are currently functional and the rest are found from some other online sources. Details about these online sentiment analysis tools are shown in Table 3.6.

Table 3.6: Online Sentiment Analysis Tool's Details

Sentiment Analysis Tool	Web Address	Sentiment Polarity		
		Positive	Negative	Neutral
Alchemy	http://www.alchemyapi.com	positive	negative	neutral
Semantria	https://www.lexalytics.com	positive	negative	neutral
Sentiment140	http://www.sentiment140.com	4	0	2
TextProcessing	http://text-processing.com	Pos	Neg	Neu
TheySay	http://www.theysay.io	POSITIVE	NEGATIVE	NEUTRAL
MeaningCloud	http://www.meaningcloud.com	P, P+	N, N+	NEU
uClassify	https://uclassify.com	P > 55%	N > 55%	45 - 55%
SentimentAnalyzer	http://sentimentanalyzer.appspot.com	> 0.55	< 0.45	.45 - .55
Repustate	http://www.repustate.com	> 0	< 0	0
AiApplied	http://ai-applied.nl	positive	negative	neutral

As presented in Figure 3.5, tweets from both datasets were submitted to these online tools utilizing their Application Program Interface (API). Some tools were not able to classify some tweets from both the datasets possibly due to either the size of tweets being too short or long, or having characters like "!" or "?" etc. Table 3.7 below shows the detail for un-classified tweets for each tool.

Table 3.7: Un-classified Tweets by Online Tools

Sentiment Analysis Tool	Number of Un-Classified Tweets	
	Dataset 1	Dataset 2
Alchemy	27	25
Semantria	0	0
Sentiment140	0	0
TextProcessing	0	0
TheySay	1	0
MeaningCloud	185	183
uClassify	0	0
SentimentAnalyzer	0	0
Repustate	0	0
AiApplied	8	10
Proposed Solution	0	0

The classification result‟s format from different online sentiment analysis tools vary from each other. While a few tools compute classification result in numerical range, others categorized them as positive, negative, neutral or some different classes. Since these distinctions make the comparison troublesome, each of the tools' outcomes were standardized to regular classes as P, N, NEU and NA for positive, negative, neutral and un-classified respectively (Cho et al., 2014). The overall process of comparison with online sentiment analysis tools is given in Figure 3.5.

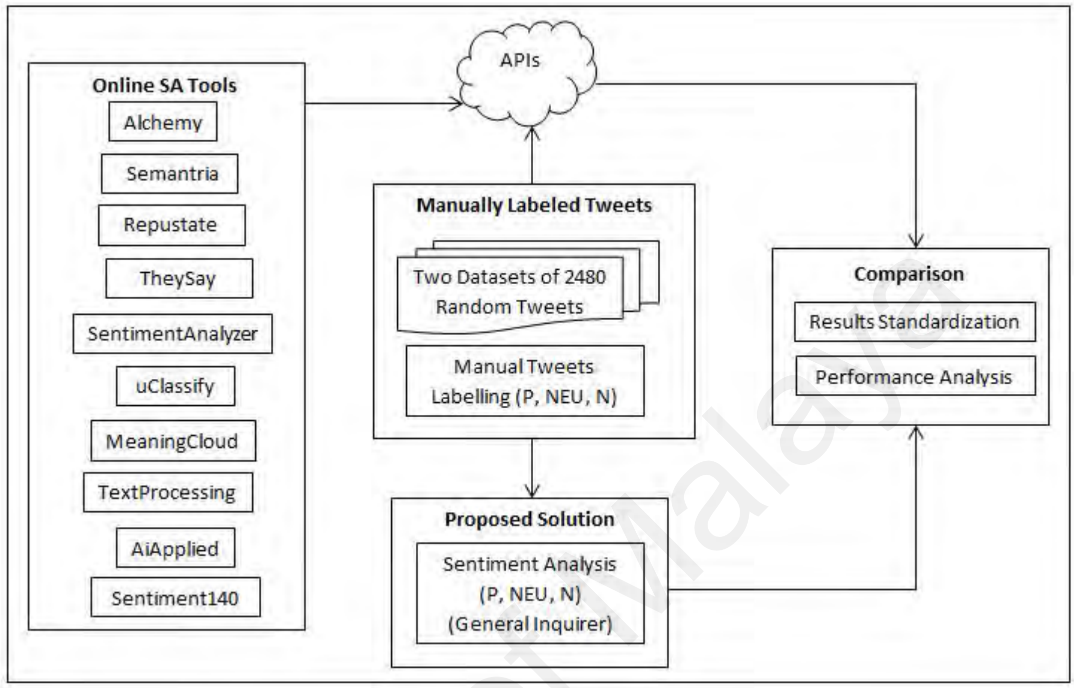

Figure 3.5: Comparison with Online Sentiment Analysis Tools

3.5 Summary

This chapter described the detailed research methodology of the proposed solution. It highlighted the details of dataset and the experimental setup for each step to achieve the goal of this research. This chapter also presented the evaluation methods by using benchmark metrics and comparison with online sentient analysis tools to assess the effectiveness of the proposed system. Examples of evaluation metrics calculation is also included according to assumed data.

The next chapter introduces the main findings of this research work with the experimental and other evaluation results in detail.

CHAPTER 4: RESULTS AND DISCUSSION

This chapter highlights the results of the enhanced lexicon-based technique along with their analysis. Step by step results are presented at every stage with the incorporation of different aspects as discussed in Chapter 3. Detailed evaluation results are described with the comparison of online sentiment analysis tools based on benchmark metrics precision, recall, accuracy and F1 score. Furthermore, an overview of overall results is also provided for labeled and complete datasets, and categorization of data based on sub-issues.

4.1 Sentiment Analysis Results and Discussion

In this research we endeavored to enhance the sentiment analysis of social issues by using lexicon-based method and incorporating grammatical dependencies with the role of verb. In the step by step implementation of our proposed technique according to Algorithm 3.1, results in Table 4.1 and Table 4.2 provide an overview of different evaluation metrics for each step for both datasets. Precision, recall and F1 score is calculated at each step for positive, negative and neutral sentiment classification and the overall accuracy as well. All the values of accuracy, precision, recall and F measure are given in % in Table 4.1 and Table 4.2. As mentioned in Section 3.3.1, lemmatized verbs are used from the tokenized tweets in step 1 (a) with accuracy of 30.1% and 32.3% for dataset 1 and dataset 2 respectively. In step 1 (b), accuracy increased to 34.5% for dataset 1 and 36.6% for dataset 2 with the usage of improved verbs detection from full tweet sentences using Stanford Core NLP tool (Manning et al., 2014) instead of tokens as used in step 1 (a).

Although General Inquirer (Stone et al., 1966) is used in all the steps in this lexicon-based approach, but in step 2 it is specifically mentioned with some improvements by

Table 4.1: Step by Step Results of Proposed Solution for Dataset 1

Step	Accuracy (%)	Positive			Negative			Neutral		
		F1	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision
1 (a)	30.1	18.1	55.3	10.8	39.9	25.6	91.4	18.9	53.8	11.4
1 (b)	34.5	18.4	53.9	11.1	46.4	31.1	91.3	19.6	52.1	12.1
2	49	25.6	59.2	16.3	63.3	47.5	94.8	23.9	56.3	15.2
3	49.5	26.7	60.5	17.1	63.8	48.1	94.5	23.3	54.6	14.8
4	**81.4**	51.4	72.4	39.9	89	83.4	95.3	53.6	68.9	43.9

Step 1 (a): Verbs extracted from tokens

Step 1 (b): Verbs extracted from sentences

Step 2: General Inquirer Improvement

Step3: Negation Handling

Step4: Grammatical Dependencies

Table 4.2: Step by Step Results of Proposed Solution for Dataset 2

Step	Accuracy (%)	Positive			Negative			Neutral		
		F1	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision
1 (a)	32.3	17.9	56.6	10.6	43.3	28	94.8	20.2	54.6	12.4
1 (b)	36.6	17.4	52.6	10.4	49.9	33.9	94.7	19.9	50.4	12.4
2	50.9	24.1	52.6	15.6	65.9	50	96.5	24.2	57.1	15.4
3	50.8	22.4	47.4	14.6	65.9	50.3	95.3	24.2	57.1	15.4
4	**82.3**	49.2	60.5	41.4	90	85.5	95	52.6	68.1	42.9

Step 1 (a): Verbs extracted from tokens

Step 1 (b): Verbs extracted from sentences

Step 2: General Inquirer Improvement

Step3: Negation Handling

Step4: Grammatical Dependencies

adding missing terms according to the corpus. These improvements in General Inquirer resulted in enhanced accuracy of 49% and 50.9 % for dataset 1 and dataset 2 respectively which is quite significant in this step. Negation handling is incorporated in step 3 which merely effected the accuracy but as mentioned earlier in Section 3.3.3, all the subsequent aspects also include their predecessors and has positive impact. Effect of grammatical dependencies with verb is evident in the last step which improved the accuracy of sentiment analysis very effectively. That"s why, this aspect of the proposed solution is the main contribution in this study as mentioned earlier in Section 3.3.4. Accuracy is enhanced to 81.4% for dataset 1 and 82.3% for dataset 2 with the integration of grammatical dependencies in step 4 as highlighted in Table 4.1 and Table 4.2.

Purpose of showing separate results on each step is to provide an insight of every aspect in the proposed technique of sentiment analysis and their significant effect on the performance of classifier. As shown in Table 4.1 and Table 4.2, results from both datasets are almost similar by using the proposed technique, thus demonstrating a consistent pattern in performance regardless of the dataset.

4.2 Results and discussion of comparison with online sentiment analysis tools

Beside the manually labeled tweets by the domain experts to evaluate the effectiveness of proposed technique, online tools of sentiment analysis were also used to compare the results with the proposed technique for both datasets as described earlier in Section 3.4.3.

Table 4.3 and Table 4.4 show an overview of results deduced from dataset 1 and dataset 2 for overall accuracy, precision, recall and F measure of positive, negative and neutral classification for each sentiment analysis tool and proposed technique as well. All the values of accuracy, precision, recall and F measure are given in % in Table 4.3

and Table 4.4. The minor differences in the values of evaluation metrics in Table 4.3 and Table 4.4 for all sentiment analysis tools are due to the different number of correct and wrong classification of positive, negative and neutral sentiment in both datasets. For each metric, values with single asterisk (*) demonstrate the best performance while the figures with double asterisk (**) show the worst results among the online sentiment analysis tools. Results from the proposed technique are shown highlighted at the end in both the tables. The following sections describe the results of each evaluation metric in detail with the graphical presentation.

Table 4.3: Results of Comparison with Online SA Tools for Dataset 1

API Name	Accuracy (%)	Positive			Negative			Neutral			Un-Classified
		F1	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	
Alchemy**	*10.5*	*4.5*	*2.6*	*15.4*	*3.6*	*1.8*	*82.6*	*16.8*	*91.6*	*9.3*	*27*
Semantria	59.5	28.4	39.5	22.2	73.6	62.4	89.7	22.5	47.1	14.8	0
Sentiment140	24	12.9	7.9	35.3	30.8	18.7	87.8	17.3	81.5	9.7	0
TextProcessing	26.2	20.1	26.3	16.3	34.5	21.7	84.1	16.1	65.5	9.2	0
TheySay**	*10.5*	*6.5*	*3.9*	*17.6*	*2.1*	*1.1*	*91.7*	*17.5*	*97.5*	*9.6*	*1*
MeaningCloud	58.5	28	60.5	18.3	76.7	64.6	94.3	4.9	4.2	5.7	185
uClassify*	*63.2*	*18.3*	*43.4*	*11.6*	*78.3*	*70.6*	*88*	*11*	*10.9*	*11.1*	*0*
Senti.Analyzer	52.8	16.4	50	9.8	69	57.4	86.5	12.3	14.3	10.8	0
Repustate	58.7	27.4	55.3	18.2	73.1	60.5	92.3	24.4	45.4	16.7	0
AiApplied*	*72.9*	*21.6*	*44.7*	*14.2*	*85.4*	*83.3*	*87.6*	*0*	*0*	*0*	*8*
Proposed Solution	**81.4**	**51.4**	**72.4**	**39.9**	**89**	**83.4**	**95.3**	**53.6**	**68.9**	**43.9**	**0**

*: Best performance among online tools **: Worst performance among online tools

Table 4.4: Results of Comparison with Online SA Tools for Dataset 2

API Name	Accuracy (%)	Positive			Negative			Neutral			Un-Classified
		F1	Recall	Precision	F1	Recall	Precision	F1	Recall	Precision	
Alchemy**	*11*	*2.2*	*1.3*	*7.7*	*3.9*	*2*	*84*	*17.6*	*95.8*	*9.7*	*25*
Semantria	58.5	27.3	40.8	20.5	73.3	61.1	91.7	22.3	47.9	14.5	0
Sentiment140	24.4	14.1	9.2	30.4	30.4	18.2	91.8	18.8	89.1	10.5	0
TextProcessing	25.1	18.5	26.3	14.3	33.2	20.7	84.7	15.6	63	8.9	0
TheySay**	*10.4*	*2.2*	*1.3*	*6.7*	*1.9*	*1*	*100*	*17.7*	*99.2*	*9.7*	*0*
MeaningCloud	57.3	27.5	61.8	17.7	74.7	62.9	92.1	6.1	5	7.7	183
uClassify*	*61.9*	*17.6*	*44.7*	*11*	*77*	*68.8*	*87.5*	*13.2*	*12.6*	*13.9*	*0*
Senti.Analyzer	52.2	13.9	42.1	8.3	69.2	57.5	86.8	9.9	11.8	8.5	0
Repustate	56	29.5	61.8	19.3	70.9	57.7	92.1	19.5	37.8	13.2	0
AiApplied*	*74.6*	*26.5*	*50*	*18*	*85.9*	*84.9*	*87*	*0*	*0*	*0*	*10*
Proposed Solution	**82.3**	**49.2**	**60.5**	**41.4**	**90**	**85.5**	**95**	**52.6**	**68.1**	**42.9**	**0**

*: Best performance among online tools **: Worst performance among online tools

4.2.1 Precision

Figure 4.1 and Figure 4.2 present the precision of positive, negative and neutral sentiment classification for online tools and the proposed solution for dataset 1 and dataset 2 respectively. Precision values in percentage are presented on Y-axis while name of the tools are mentioned on X-axis.

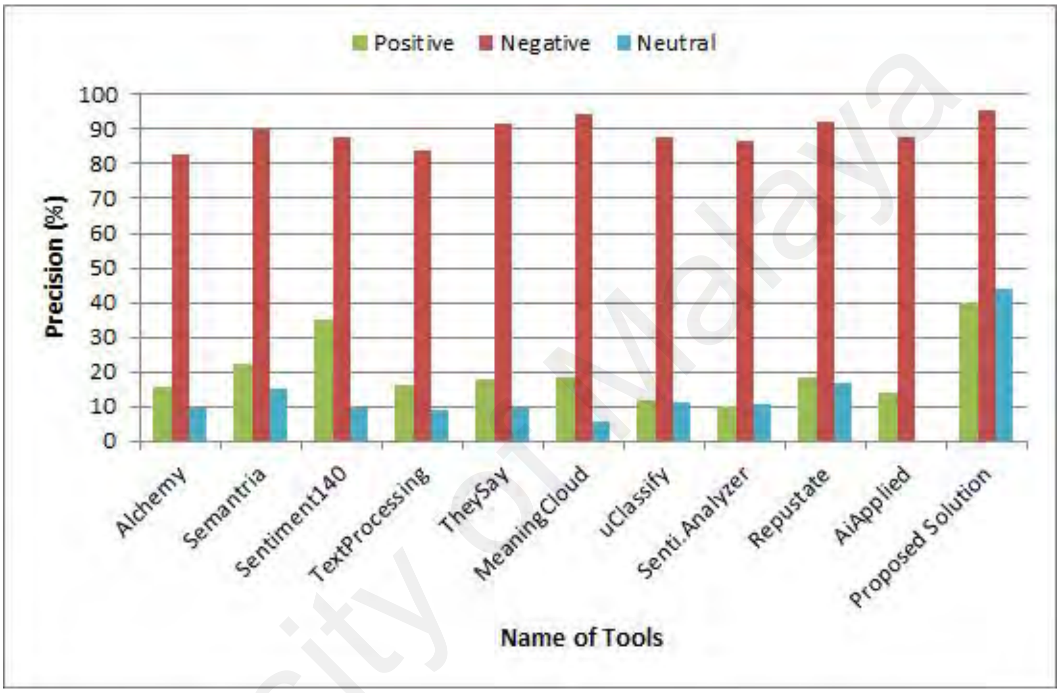

Figure 4.1: Precision for Dataset 1

For precision, all the online sentiment analysis tools in this research exhibit quite consistent behavior as depicted in Figure 4.1 and Figure 4.2. For positive precision, Sentiment140 performed better in comparison to the rest of the online tools with 30% precision, while all other tools fell to less than 20%. The proposed solution outperformed in positive precision with almost 40% in both datasets.

TheySay outperformed in negative precision with 91% and 100% in both datasets among online sentiment analysis tools, similarly MeaningCloud and Repustate also performed the best in negative precision for both datasets with more than 90% precision. Other tools‟ performance came out best as well in negative precision with more than 80%. 95% negative class was precise in both datasets with the proposed solution.

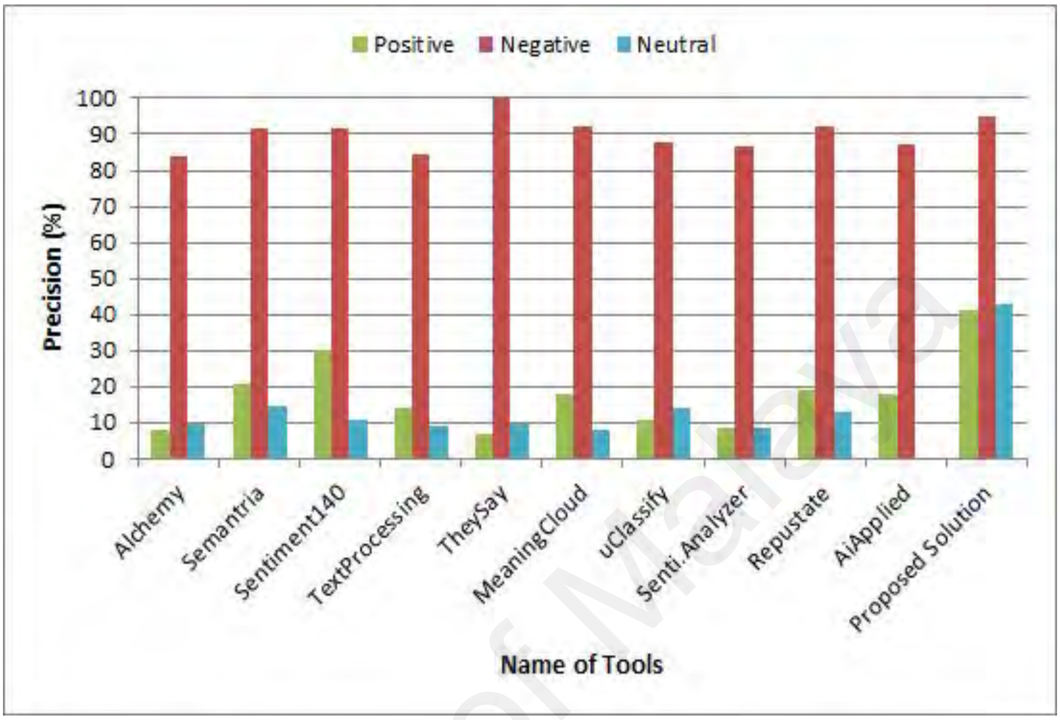

Figure 4.2: Precision for Dataset 2

For neutral precision, all online sentiment analysis tools are failed to give good results with precision between 5% and 15% for both datasets because mainly they work for positive and negative classes. The proposed solution outperformed in neutral precision with 43.9% and 42.9% for dataset 1 and dataset2 respectively. AiApplied was unable to classify neutral tweets hence precision is 0%.

From Figure 4.1 and Figure 4.2, it can be inferred that negative precision of all tools were high as compared to positive and neutral precision. This is because the dataset contains more negative tweets as compared to positive and neutral. Most of the online sentiment analysis tools are not good in classifying neutral texts regardless of the domain (Serrano-Guerrero et al., 2015). The proposed solution performed well for neutral class also as compared to other online sentiment analysis tools.

4.2.2 Recall

As illustrated in Figure 4.3 and Figure 4.4, proposed solution comparatively outperformed online sentiment analysis tools. MeaningCloud and Repustate performed well at classifying positive tweets with around 60% recall among online sentiment analysis tools. AiApplied and SentimentAnalyzer positive recall is quite good in comparison to other tools. Semantria and uClassify showed almost similar results of positive recall with around 40%. Sentiment140 performed badly with less than 10% while Alchemy and TheySay performed the worst in positive recall.

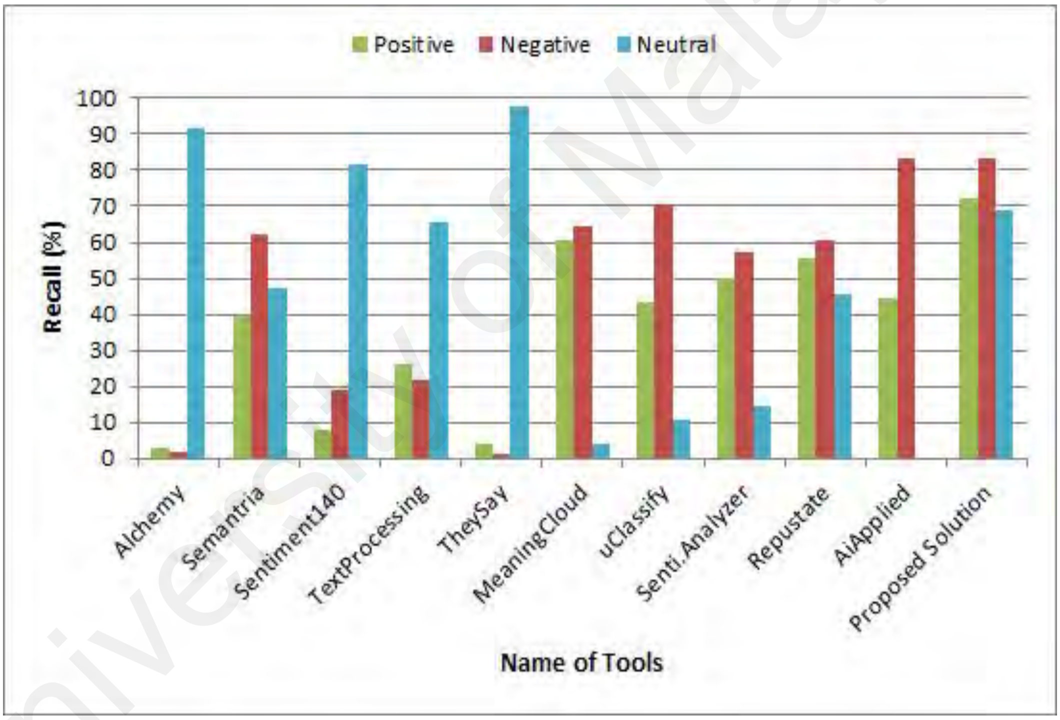

Figure 4.3: Recall for Dataset 1

The proposed solution also outperformed the online sentiment analysis tools in negative recall with 83.4% and 85.5% recall for dataset1 and dataset 2 respectively, as shown in Table 4.3 and Table 4.4. AiApplied also performed well for negative recall with more than 80% recall. uClassify came out second among online sentiment analysis tools with 70% recall of negative class. Semantria, MeaningCloud, SentimentAnalyzer and Repustate also showed good results in negative recall. Here again, Alchemy and TheySay ranked lowest for negative recall.

Neutral recall show quite different results as highlighted in Figure 4.3 and Figure 4.4. Alchemy and TheySay outperformed the rest in neutral recall with more than 90% since these two tools analyzed maximum tweets as neutral only. Sentiment140 performed well with 80% and 90% in both datasets, respectively. uClassify and SentimentAnalyzer performed badly in neutral recall while MeaningCloud and AiApplied ranked the worst with 5% and 0%. Evidently, AiApplied which outperformed all other among online sentiment analysis tools in overall performance failed to classify neutral tweets.

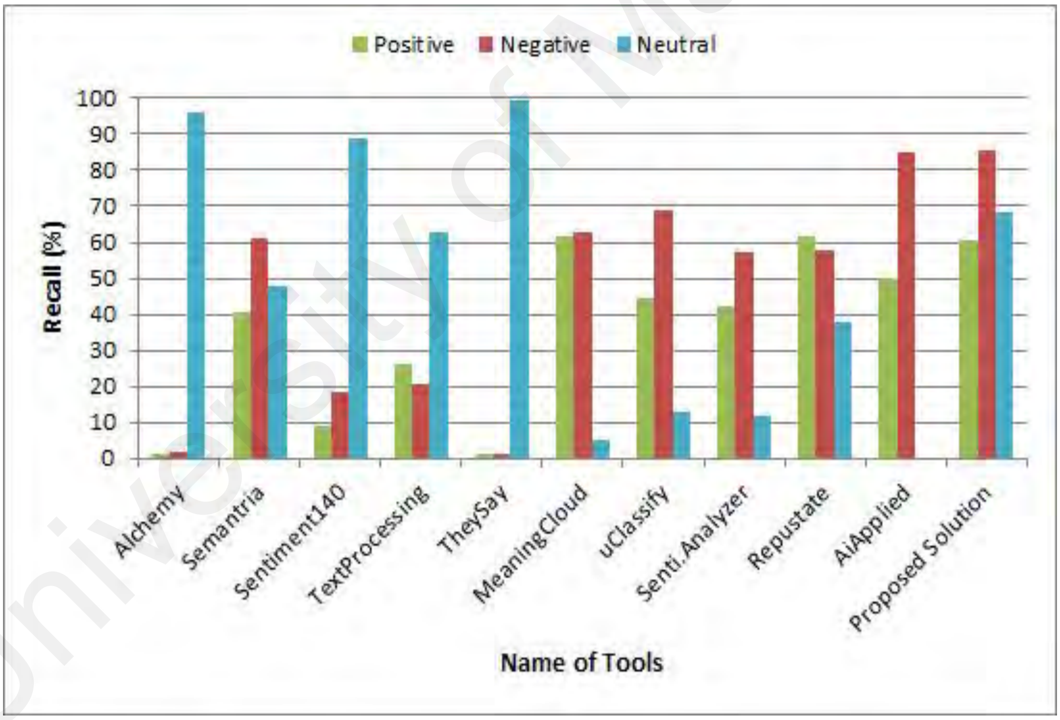

Figure 4.4: Recall for Dataset 2

The proposed solution showed comparatively low performance in neutral recall with almost 68% in both datasets but as illustrated in Figure 4.1 and Figure 4.2 in Section 4.2.1, neutral sentiment classification is more precised as compared to online sentiment analysis tools.

4.2.3 F1 Score

The combined effect of precision and recall in the form of F1 score for positive, negative and neutral sentiment classification is depicted in Figure 4.5 for dataset 1 and Figure 4.6 for dataset 2. Y-axis highlights the F1 score values in percentage whereas name of sentiment analysis tools are shown on X-axis.

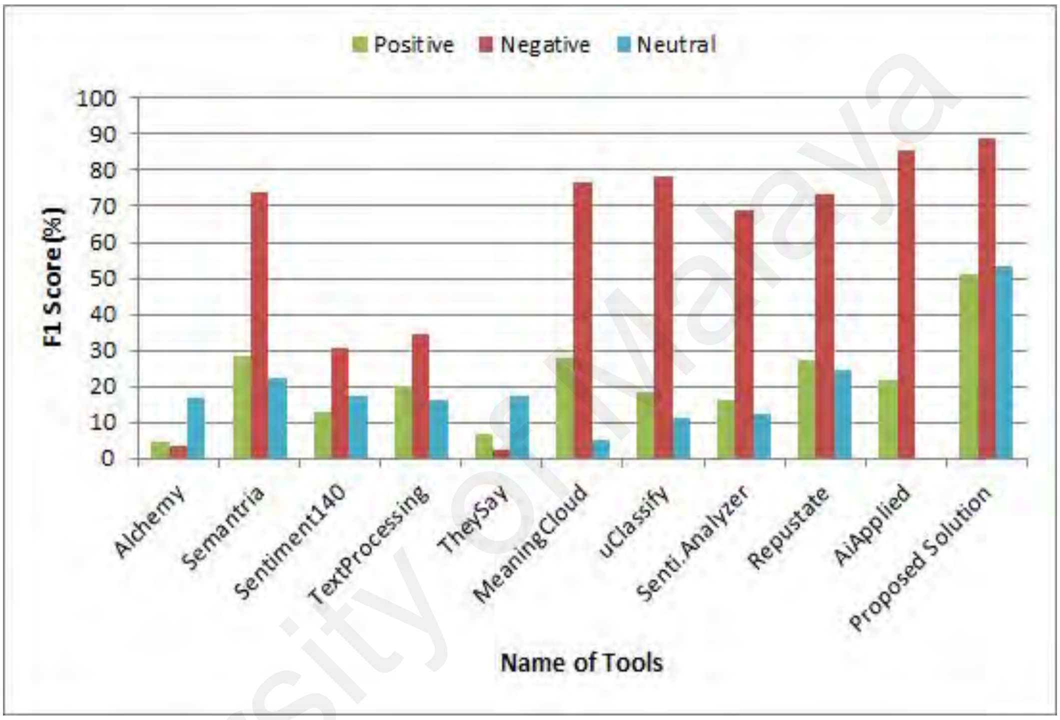

Figure 4.5: F1 Score for Dataset 1

As presented in Figure 4.5 and Figure 4.6, proposed solution did well for positive class as compared to other online sentiment analysis tools with almost 50% F1 score. Semantria, MeaningCloud and Repustate performed well among other online tools with almost 28% positive F1 score. Alchemy and TheySay graded lowest for positive F1 score with less than 5%.

Proposed solution also outperformed the online sentiment analysis tools in negative class with 89% F1 score for dataset 1 and 90% F1 score for dataset 2. AiApplied performed well in the online sentiment analysis tools with F1 score of almost 85%.

Semantria, MeaningCloud, uClassify and Repustate also worked effectively for negative tweets with more than 70% F1 score for both datasets. Alchemy and TheySay underperformed among all sentiment analysis tools under study for negative class as well.

Neutral F1 score also showed almost same type of results like positive and negative F1 score for all sentiment analysis tools. Proposed solution again outperformed with 53.6% F1 score for dataset 1 and 52.6% F1 score for dataset 2. In contrast with positive and negative F1 score, AiApplied couldn‟t classify neutral tweets. Alchemy, Semantria, Sentiment140, TextProcessing, TheySay and Repustate exhibits nearly similar handling for neutral class with almost 15% to 20% F1 score. In contrast with lowest positive and negative F1 score, Alchemy and TheySay showed better F1 score for neutral class because both the tools marked maximum tweets as neutral only.

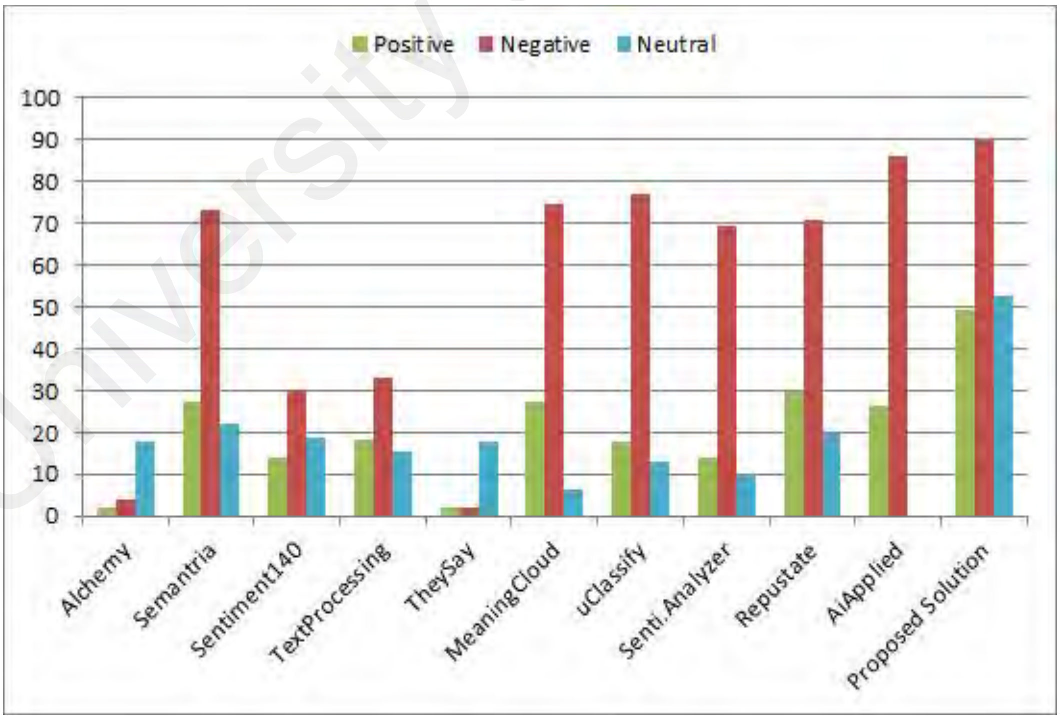

Figure 4.6: F1 Score for Dataset 2

From Figure 4.5 and Figure 4.6, it can be inferred that F1 scores of proposed solution is highest for all three classes i.e. positive, negative and neutral. This shows that the proposed technique is better for the overall sentiment classification of positive, negative and neutral tweets. AiApplied performed well among online sentiment analysis tools overall but it was not able to classify neutral tweets.

4.2.4 Accuracy

Figure 4.7 and Figure 4.8 illustrates the overall accuracy of online tools and the proposed solution for dataset 1 and dataset 2 respectively. Accuracy values in percentage are presented on Y-axis while name of the tools are mentioned on X-axis.

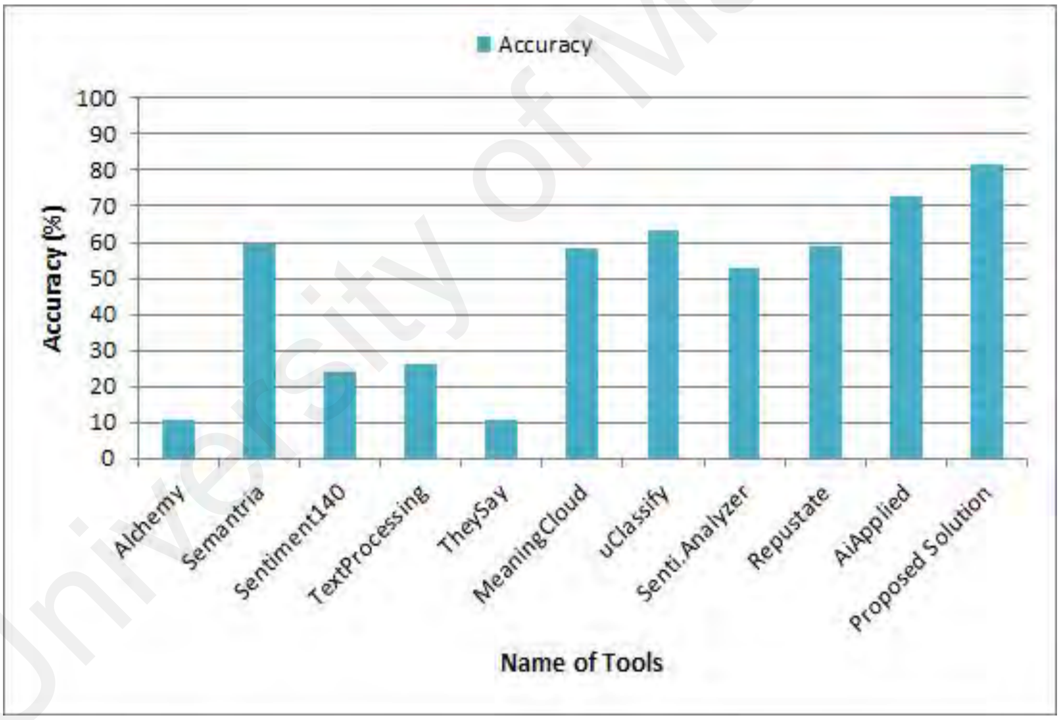

Figure 4.7: Accuracy for Dataset 1

As depicted in Figure 4.7 and Figure 4.8, proposed solution outperformed all online sentiment analysis tools with accuracy of 81.4% and 82.3% for dataset 1 and dataset 2 respectively. Among online sentiment analysis tools, AiApplied showed good overall performance with almost 74% accuracy but was unable to classify neutral tweets.

uClassify also performed quite well for overall classification with accuracy of 63.2% and 61.9%. Semantria, MeaningCloud and Repustate results are satisfactory with accuracy of 56% to 59% and no significant rise in performance rate. However, tools like Alchemy and TheySay may be discarded due to their poor performance results.

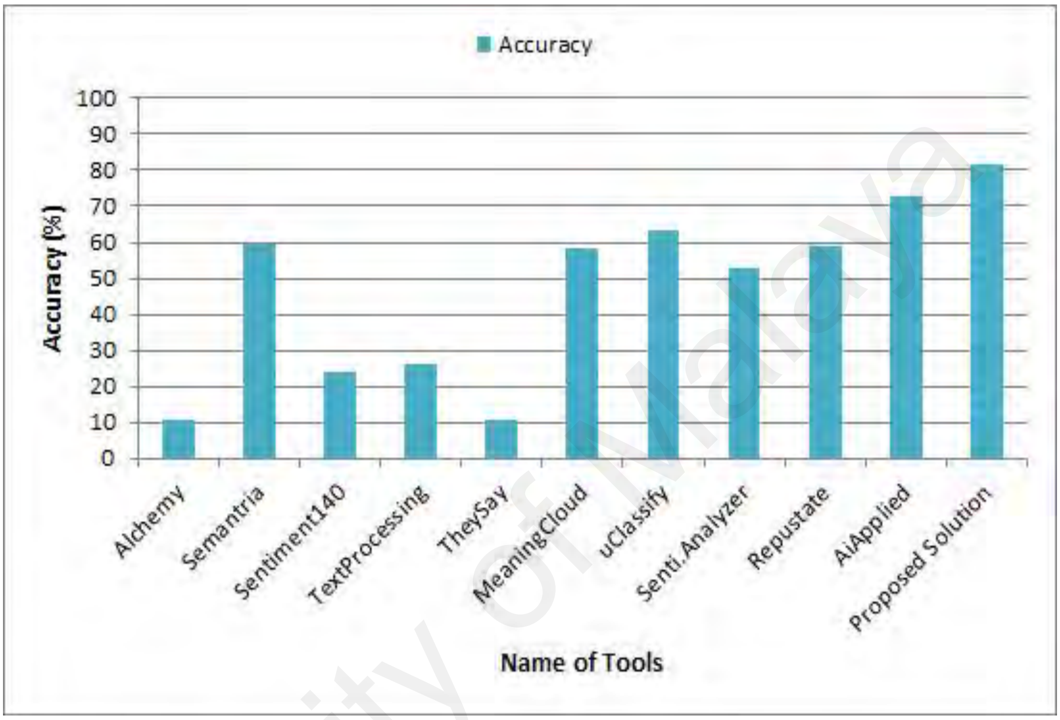

Figure 4.8: Accuracy for Dataset 2

In an earlier study by Serrano-Guerrero et al. (2015), Alchemy was considered the best online sentiment analysis tool for Twitter datasets but the current study demonstrated otherwise. It is important to note that the dataset used by Serrano-Guerrero et al. (2015) was related to products and services while this study involves the analysis of dataset from a social issue. Such differing verticals are expected to produce different results as the sentiment analysis of social issues differs from products and services (Karamibekr & Ghorbani, 2012a). They highlighted that Twitter users utilize adjectives, adverbs and nouns specifically to express their opinion of products and their features, whereas verbs are used extensively to discuss a social issue.

The proposed lexicon-based solution produced more accurate results as the sentiment lexicon General Inquirer is used with the improvements of domain dependent terms and usually domain dependent lexicons are the important factor in the improved accuracy of sentiment classification (Asghar et al., 2016). Grammatical dependencies are another reason for the better sentiment classification in the proposed technique. As discussed earlier in Section 4.1, accuracy was boosted very prominently by incorporating the grammatical dependencies of verb along with other aspects in this enhanced lexicon-based sentiment analysis approach.

4.3 Overall Results

Beside manually labeled datasets used for the evaluation of proposed solution, complete dataset of 232,785 tweets was classified as well with the same approach as mentioned in Algorithm 3.1.

Table 4.5: Overall Results

	Positive	Negative	Neutral	Total
Labeled Data	152 (6.13%)	2,090 (84.27%)	238 (9.6%)	2,480
Dataset 1	138 (11.13%)	915 (73.39%)	187 (15.08%)	1,240
Dataset 2	111 (8.95%)	940 (75.81%)	189 (15.24%)	1,240
Complete Dataset	28,931 (12.43%)	142,816 (61.35%)	61,038 (26.22%)	232,785

As shown in Table 4.5 and Figure 4.9, results from complete dataset also shows nearly same percentage values of positive, negative and neutral tweets same as manually labeled data. The dataset was collected for a social issue illegal immigration, so it was supposed to have more negative tweets than positive and neutral, and the results describe the same.

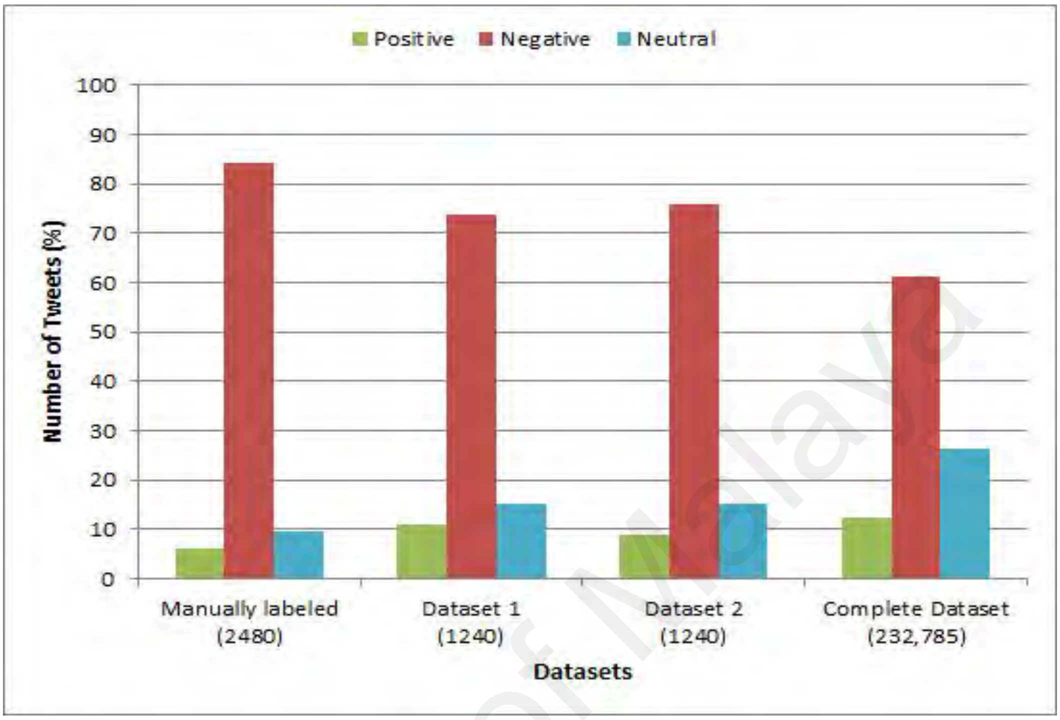

Figure 4.9: Overall Results

4.4 Categorization Based on Sub Issues

After classifying all the tweets in the dataset for positive, negative and neutral tweets by using the proposed approach, tweets were categorized based on sub-issues as described in Section 3.3.6. Table 4.6 illustrates the number of tweets in each category and also count with the percentage values of positive, negative and neutral tweets for the same.

Table 4.6: Categorization Based on Sub-Issues

Category	Positive	Negative	Neutral	Total
Unemployment	297 (14.13%)	1,140 (54.25%)	664 (31.60%)	2,101
Crimes	771 (5.70%)	10,420 (77.16%)	2,312 (17.12%)	13,503
Terrorism	71 (5.89%)	940 (78.07%)	193 (16.02%)	1,204
Security	2,270 (16.17%)	8,471 (60.36%)	3,291 (23.45%)	14,032
Law and Order	3,991 (10.09%)	27,190 (68.79%)	8,341 (21.10%)	39,522
Government Benefits	1,337 (19.91%)	3,665 (54.58%)	1,712 (25.49%)	6,714
Economic Issues	660 (11.47%)	3,815 (66.30%)	1,279 (22.22%)	5,754
Racism	198 (10.59%)	1,271 (68.04%)	399 (21.35%)	1,868
Food	123 (10.80%)	616 (54.13%)	399 (35.06%)	1,138

Categorization of data based on sub-issues in social issues or features in product and services is same and it provide a clear insight about what people are talking and their sentiment according to sub-issues or features. Such type of information may be helpful for government and other social organizations for policy making decisions. Figure 4.10 depicts the categories with sentiment classification to have a better overview.

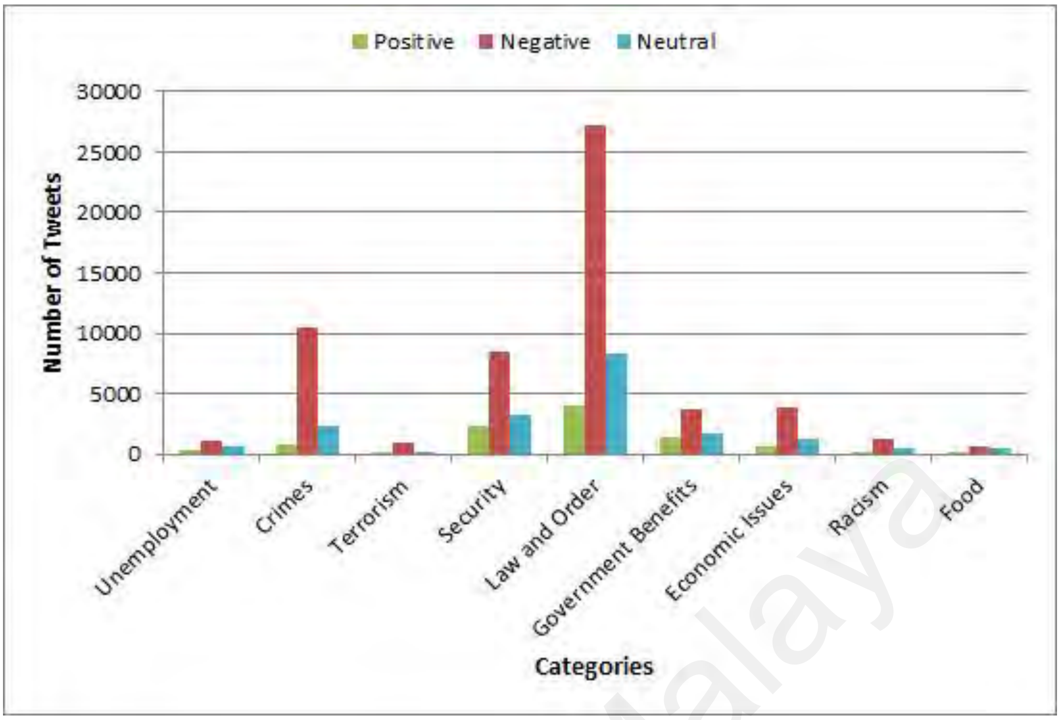

Figure 4.10: Categorization Based on Sub-Issues

4.5 Summary

This chapter has highlighted the step by step experimental results of proposed approach at each stage of implementation. It described the evaluation results also, based on manually labeled tweets and comparison with online sentiment analysis tools to assess the effectiveness of proposed solution with benchmark metrics. This chapter also presented the sentiment classification of complete dataset using the proposed approach. Moreover, categorization of tweets based on sub-issues is summarized with the sentiment classification details as well.

Finally, the next chapter concludes the study, describes the limitation and also highlights the avenues for the future work.

CHAPTER 5: CONCLUSION, LIMITATIONS AND FUTURE WORK

This chapter summarizes and concludes the overall research work in terms of its objective and major contributions. It is described briefly in Section 5.1 and Section 5.2. Further, it highlights the limitations of this research in Section 5.3. Future dimension of this research work is presented in Section 5.4.

5.1 Conclusion

With the rapid increase and online availability of opinions concerning almost every domain of life, it is pertinent to investigate the opinions to gain an insight and have more precise knowledge of certain aspects. This context has brought forward the approach of sentiment analysis. Nevertheless, like other research areas, sentiment analysis also involved various challenges as same technique of sentiment classification is not applicable on the data of different domains. It has been observed that most of the research is done on the sentiment analysis of product and services. More research is required in the sentiment analysis of social issues. Also minimal literature exists in this area. This led to the motivation of conducting research in the area of sentiment analysis of social issues.

The first objective of this research work was to enhance lexicon-based sentiment analysis for social issues. This objective led to the exploration of the approaches that can be adopted to improve lexicon based sentiment analysis. It also directs to find out the impact of dependency grammar of verb on sentiment analysis, as verb is a vital component in the sentiment analysis of social issues. This results in an enhanced approach of lexicon-based sentiment analysis with 81.4% and 82.3% accuracy as compared to the highest accuracy among online sentiment analysis tools of 72.9% and 74.6 in two datasets. The proposed approach is comprised of verbs with grammatical dependencies and an improved General Inquirer sentiment lexicon.

The second objective was to assess the effectiveness and evaluate the performance of the proposed solution. In this context the benchmark evaluation metrics such as precision, recall, accuracy and F1 score were defined and calculated to measure the system‟s effectiveness based on data manually labeled by three domain experts. Furthermore the performance of the proposed solution was comparatively analyzed with respect to the existing online sentiment analysis tools. Ten online tools were selected for the quantitative comparison. The proposed technique outperformed the online tools not only for overall accuracy but also demonstrated best results for each class of positive, negative and neutral sentiment classification.

Above all, sentiment classification of the complete dataset as positive, negative and neutral tweets was performed on the basis of the proposed approach. Additionally, the tweets in the dataset were categorized with respect to the sub-issues, under the spectrum of social issues, namely crimes, security, law and order, etc. This could prove to be helpful at the policy decision-making level.

5.2 Research Contribution

A considerable body of research is available for the sentiment analysis of product and services while less attention was paid towards the sentiment analysis of the social issues. In the context of sentiment analysis of social issues, this research has added to the body of knowledge that is stated as follows:

i. Several approaches were explored that can be adopted to improve lexicon based sentiment analysis. This results in an enhanced approach of lexicon-based sentiment analysis for social issues by incorporating grammar dependencies of verb.

ii. The current research work also results in an improved general inquirer sentiment-lexicon.

iii. It also provides sentiment classification of the complete dataset as positive, negative and neutral tweets on the basis of the proposed approach. Furthermore, the tweets in the dataset were categorized with respect to the sub-issues, under the spectrum of social issues.

5.3 Limitation

The possible limitations of this study are stated as follows:

i. In the current work, the results of the proposed solution were compared with the results of existing online tools. Nevertheless, comparison can be done by applying the proposed technique on the dataset of product and services first and then comparatively analyzed with respect to the data set of social issues.
ii. Results of the complete data sets do not have the accuracy, as the labeled data was unavailable. Due to voluminous data, manual labeling is a time consuming and cost effective process for the complete dataset.
iii. The accuracy of categorization based on sub-issues was not considered due to unavailability of manually annotated categories. Also, it was not under the scope of this research as well.

5.4 Future Work

This study can be extended further by incorporating some of the suggestion as listed below:

i. This research can be extended by focusing on some other sentiment-lexicons also under lexicon-based sentiment analysis approach such as SentiWordNet, AFFIN, SO-Cal, etc. Also the combination of various sentiment-lexicons will help to see the performance of the proposed solution (Cho et al., 2014). Some missing terms in one sentiment-lexicon might be available in other sentiment-lexicon, hence resulting in more improved accuracy of classification.

ii. Moreover, this research can be extended by the involvement of machine learning based approach in combination with lexicon-based method which is known as hybrid approach and it may get higher accuracy of the sentiment classifier (Asghar et al., 2016).

iii. Also, future studies could utilize the dataset from product and services too beside social issues‟ dataset only to have thorough and prominent comparison.

iv. Additionally, it would be more helpful to include offline sentiment analysis tools like WEKA and Spark along with online sentiment analysis tools for better comparative analysis and to provide more insight to the researchers and industry practitioners.

REFERENCES

Abbasi, A., Hassan, A., & Dhar, M. (2014). *Benchmarking Twitter Sentiment Analysis Tools.* Paper presented at the LREC.

Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., & Tao, K. (2012). *Semantics+ filtering+ search= twitcident. exploring information in social web streams.* Paper presented at the Proceedings of the 23rd ACM conference on Hypertext and social media.

Angulakshmi, G., & ManickaChezian, R. (2014). An analysis on opinion mining: techniques and tools. *International Journal of Advanced Research in Computer Communication Engineering, 3*(7), 7483-7487.

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems, 108*, 110-124.

Araujo, M., Reis, J., Pereira, A., & Benevenuto, F. (2016). *An evaluation of machine translation for multilingual sentence-level sentiment analysis.* Paper presented at the Proceedings of the 31st Annual ACM Symposium on Applied Computing.

Asghar, M. Z., Ahmad, S., Qasim, M., Zahra, S. R., & Kundi, F. M. (2016). SentiHealth: creating health-related sentiment lexicon using hybrid approach. *SpringerPlus, 5*(1), 1139.

Avanço, L. V., & Nunes, M. G. (2014). *Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese.* Paper presented at the Intelligent Systems (BRACIS), 2014 Brazilian Conference on.

Balakrishnan, V., & Kaur, W. (in press). Improving sentiment scoring mechanism: a case study on airline services. *Industrial Journal of Management System*.

Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: a comparison of retrieval performances. *Lecture Notes on Software Engineering, 2*(3), 262.

Bandhakavi, A., Wiratunga, N., Padmanabhan, D., & Massie, S. (2017). Lexicon based feature extraction for emotion text classification. *Pattern Recognition Letters, 93*, 133-142.

Beare, M. E. (1997). Illegal migration: Personal tragedies, social problems, or national security threats? *Transnational Organized Crime. Volume 3, Issue 4 (1997), p. 11-41.*

Bhadane, C., Dalal, H., & Doshi, H. (2015). Sentiment Analysis: Measuring Opinions. *Procedia Computer Science, 45*, 808-814.

Bilal, M., Israr, H., Shahid, M., & Khan, A. (2016). Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques. *Journal of King Saud University-Computer and Information Sciences, 28*(3), 330-344.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science, 2*(1), 1-8.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings: Technical report C-1, the center for research in psychophysiology, University of Florida.

Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J., . . . Conejero, J. (2015). COSMOS: Towards an integrated and scalable service for analysing social media on demand. *International Journal of Parallel, Emergent and Distributed Systems, 30*(2), 80-100.

Burnap, P., Rana, O. F., Avis, N., Williams, M., Housley, W., Edwards, A., . . . Sloan, L. (2013). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*(2), 15-21.

Chaumartin, F.-R. (2007). *UPAR7: A knowledge-based system for headline sentiment tagging.* Paper presented at the Proceedings of the 4th International Workshop on Semantic Evaluations.

Chiricos, T., Stupi, E. K., Stults, B. J., & Gertz, M. (2014). Undocumented immigrant threat and support for social controls. *Social Problems, 61*(4), 673-692.

Cho, H., Kim, S., Lee, J., & Lee, J.-S. (2014). Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. *Knowledge-Based Systems, 71*, 61-71.

Conrad, J. G., & Schilder, F. (2007). *Opinion mining in legal blogs.* Paper presented at the Proceedings of the 11th international conference on Artificial intelligence and law.

Dadvar, M., Hauff, C., & de Jong, F. M. (2011). Scope of negation detection in sentiment analysis.

Dang, Y., Zhang, Y., & Chen, H. (2010). A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems, 25*(4), 46-53.

De Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., & Manning, C. D. (2014). *Universal Stanford dependencies: A cross-linguistic typology.* Paper presented at the LREC.

Deng, Z.-H., Luo, K.-H., & Yu, H.-L. (2014). A study of supervised term weighting scheme for sentiment analysis. *Expert Systems with Applications, 41*(7), 3506-3513.

Dey, L., & Haque, S. M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJDAR), 12*(3), 205-226.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification. 2nd. *Edition. New York*, 55.

Duwairi, R., Ahmed, N. A., & Al-Rifai, S. Y. (2015). Detecting sentiment embedded in Arabic social media–A lexicon-based approach. *Journal of Intelligent & Fuzzy Systems, 29*(1), 107-117.

Esuli, A., & Sebastiani, F. (2007). SentiWordNet: a high-coverage lexical resource for opinion mining. *Evaluation*, 1-26.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM, 56*(4), 82-89.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144.

Gautam, G., & Yadav, D. (2014). *Sentiment analysis of twitter data using machine learning approaches and semantic analysis.* Paper presented at the Contemporary computing (IC3), 2014 seventh international conference on.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266-6282.

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR), 49*(2), 28.

Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering, 10*(4), 215-230.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *ICWSM, 7*(21), 219-222.

Gundecha, P., & Liu, H. (2012). Mining social media: a brief introduction. *Tutorials in Operations Research, 1*(4).

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science, 17*, 26-32.

Hardeniya, T., & Borikar, D. A. (2016). Dictionary Based Approach to Sentiment Analysis-A Review. *IJAEMS: Open Access International Journal: Infogain Publication, 2*(Issue-5).

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology, 67*(1), 232-238.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464-472.

Hridoy, S. A. A., Ekram, M. T., Islam, M. S., Ahmed, F., & Rahman, R. M. (2015). Localized twitter opinion mining using sentiment analysis. *Decision Analytics, 2*(1), 1-19.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013). *Unsupervised sentiment analysis with emotional signals.* Paper presented at the Proceedings of the 22nd international conference on World Wide Web.

Jin, W., Ho, H. H., & Srihari, R. K. (2009). *OpinionMiner: a novel machine learning system for web opinion mining and extraction.* Paper presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Machine learning: ECML-98*, 137-142.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3): Pearson London.

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications, 39*(5), 6000-6010.

Karamibekr, M., & Ghorbani, A. A. (2012a). *Sentiment analysis of social issues.* Paper presented at the Social Informatics (SocialInformatics), 2012 International Conference on.

Karamibekr, M., & Ghorbani, A. A. (2012b). *Verb oriented sentiment classification.* Paper presented at the Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01.

Kennedy, A., & Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational intelligence, 22*(2), 110-125.

Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems, 57*, 245-257.

Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. *arXiv preprint arXiv:1507.00955*.

Lakoff, G., & Ferguson, S. (2017). The framing of immigration.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.

Liu, B., Hu, M., & Cheng, J. (2005). *Opinion observer: analyzing and comparing opinions on the web.* Paper presented at the Proceedings of the 14th international conference on World Wide Web.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). *The stanford corenlp natural language processing toolkit.* Paper presented at the ACL (System Demonstrations).

Maynard, D., & Bontcheva, K. (2016). *Challenges of Evaluating Sentiment Analysis Tools on Social Media.* Paper presented at the LREC.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093-1113.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography, 3*(4), 235-244.

Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management, 51*(4), 480-499.

Mukwazvure, A., & Supreethi, K. (2015). *A hybrid approach to sentiment analysis of news comments.* Paper presented at the Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on.

Nasir, A. N. M., Selamat, A., & Selamat, M. H. (2009). Web mining for Malaysia‟s political social networks using artificial immune system *Knowledge Acquisition: Approaches, Algorithms and Applications* (pp. 137-146): Springer.

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). SentiFul: A lexicon for sentiment analysis. *IEEE Transactions on Affective Computing, 2*(1), 22-36.

Ngoc, P. T., & Yoo, M. (2014). *The lexicon-based sentiment analysis for fan page ranking in Facebook.* Paper presented at the Information Networking (ICOIN), 2014 International Conference on.

Nichols, J., & Warnow, T. (2008). Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass, 2*(5), 760-820.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

O'Doherty, K., & Lecouteur, A. (2007). "Asylum seekers","boat people" and "illegal immigrants": Social categorisation in the media. *Australian journal of psychology, 59*(1), 1-12.

Ohana, B., Tierney, B., & Delany, S.-J. (2011). *Domain independent sentiment classification with many lexicons.* Paper presented at the Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on.

Ojokoh, B. A., & Kayode, O. (2012). A feature-opinion extraction approach to opinion mining. *Journal of web engineering, 11*(1), 51-63.

Onifade, O., & Malik, M. (2015). *SASM: A tool for sentiment analysis on Twitter.* Paper presented at the Web Applications and Networking (WSWAN), 2015 2nd World Symposium on.

Pak, A., & Paroubek, P. (2010a). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining.* Paper presented at the LREc.

Pak, A., & Paroubek, P. (2010b). *Twitter based system: Using Twitter for disambiguating sentiment ambiguous adjectives.* Paper presented at the Proceedings of the 5th International Workshop on Semantic Evaluation.

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

Pang, B., & Lee, L. (2008b). *Using Very Simple Statistics for Review Search: An Exploration.* Paper presented at the COLING (Posters).

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up?: sentiment classification using machine learning techniques.* Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.

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

Qiu, G., He, X., Zhang, F., Shi, Y., Bu, J., & Chen, C. (2010). DASA: dissatisfaction-oriented advertising based on sentiment analysis. *Expert Systems with Applications, 37*(9), 6182-6191.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems, 89*, 14-46.

Rodrigues, R. G., das Dores, R. M., Camilo-Junior, C. G., & Rosa, T. C. (2016). SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks. *International journal of medical informatics, 85*(1), 80-95.

Rosenthal, S., Farra, N., & Nakov, P. (2017). *SemEval-2017 task 4: Sentiment analysis in Twitter.* Paper presented at the Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).

Rui, H., Liu, Y., & Whinston, A. (2013). Whose and what chatter matters? The effect of tweets on movie sales. *Decision Support Systems, 55*(4), 863-870.

Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management, 52*(1), 5-19.

Schouten, K., & Frasincar, F. (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering, 28*(3), 813-830.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of personality and Social Psychology, 54*(4), 558.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences, 311*, 18-38.

Singh, S. K., & Paul, S. (2015). Sentiment Analysis of Social Issues and Sentiment Score Calculation of Negative Prefixes. *International Journal of Applied Engineering Research, 10*(55), 1694-1699.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Stupi, E. K., Chiricos, T., & Gertz, M. (2016). Perceived criminal threat from undocumented immigrants: antecedents and consequences for policy preferences. *Justice Quarterly, 33*(2), 239-266.

Sun, F., Belatreche, A., Coleman, S., McGinnity, T. M., & Li, Y. (2014). *Pre-processing online financial text for sentiment classification: A natural language processing approach.* Paper presented at the Computational Intelligence for Financial Engineering & Economics (CIFEr), 2104 IEEE Conference on.

Syed, A. Z., Aslam, M., & Martinez-Enriquez, A. M. (2010). Lexicon based sentiment analysis of Urdu text using SentiUnits *Advances in Artificial Intelligence* (pp. 32-43): Springer.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics, 37*(2), 267-307.

Tang, D. (2015). *Sentiment-specific representation learning for document-level sentiment analysis.* Paper presented at the Proceedings of the Eighth ACM International Conference on Web Search and Data Mining.

Thelwall, M., & Buckley, K. (2013). Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the Association for Information Science and Technology, 64*(8), 1608-1617.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the Association for Information Science and Technology, 63*(1), 163-173.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social science computer review, 29*(4), 402-418.

Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management, 50*(1), 104-112.

Weichselbraun, A., Gindl, S., & Scharl, A. (2013). Extracting and grounding contextualized sentiment lexicons. *IEEE Intelligent Systems, 28*(2), 39-46.

Wilcock, G. (2009). Introduction to linguistic annotation and text analytics. *Synthesis lectures on human language technologies, 2*(1), 1-159.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., . . . Patwardhan, S. (2005). *OpinionFinder: A system for subjectivity analysis.* Paper presented at the Proceedings of hlt/emnlp on interactive demonstrations.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Wong, K. O., Davis, F. G., Zaïane, O. R., & Yasui, Y. (2016). *Sentiment Analysis of Breast Cancer Screening in the United States using Twitter.* Paper presented at the KDIR.

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences, 181*(6), 1138-1152.

Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems, 37*, 186-195.

Yasavur, U., Travieso, J., Lisetti, C., & Rishe, N. D. (2014). *Sentiment analysis using dependency trees and named-entities.* Paper presented at the The Twenty-Seventh International Flairs Conference.

Yessenalina, A., Yue, Y., & Cardie, C. (2010). *Multi-level structured models for document-level sentiment classification.* Paper presented at the Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing.

Yu, Y., & Wang, X. (2015). World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans‟ tweets. *Computers in Human Behavior, 48*, 392-400.

Zavattaro, S. M., French, P. E., & Mohanty, S. D. (2015). A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Government Information Quarterly, 32*(3), 333-341.

Zhang, C., Zeng, D., Li, J., Wang, F. Y., & Zuo, W. (2009). Sentiment analysis of Chinese documents: From sentence to document level. *Journal of the Association for Information Science and Technology, 60*(12), 2474-2487.

Zhou, Z., Zhang, X., & Sanderson, M. (2014). Sentiment analysis on twitter through topic-based lexicon expansion *Databases Theory and Applications* (pp. 98-109): Springer.