SENTIMENT ANALYSIS FOR AIRLINE SERVICES ON TWITTER USING DEEP LEARNING WITH WORD EMBEDDING

MAWADA MOHAMED NOUR EL DAIM EL KHALIFA

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

2020

SENTIMENT ANALYSIS FOR AIRLINE SERVICES ON TWITTER USING DEEP LEARNING WITH WORD EMBEDDING

MAWADA MOHAMED NOUR EL DAIM EL KHALIFA

DISSERTATION SUBMITTED IN PARTIAL

FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SOFTWARE ENGINEERING

(SOFTWARE TECHNOLOGY)

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

2019

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **Mawada Mohamed Nour El Daim El Khalifa**

Matric No: **WOC160008**

Name of Degree: **Master of Software Engineering (Software Technology)**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Sentiment Analysis for Airline Services on Twitter Using Deep Learning with Word Embedding

Field of Study: **Data Mining**

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:

Designation:

SENTIMENT ANALYSIS FOR AIRLINE SERVICES ON TWITTER USING DEEP LEARNING WITH WORD EMBEDDING

ABSTRACT

The use of social media platform in the airline industries have increased rapidly to allow analysis introduce the quality and performance of the services. The role of Sentiment Analysis (SA) is to classify people's opinions into different categories, such as positive and negative from text, using existing algorithms. However, existing approaches such as the Bag of Words (BOW) model is frequently used for text classification, where a document is mapped to a feature vector before the construction of the actual model, using machine learning techniques, like Logistical Regression and Support Vector algorithms. This problem has led to low accuracy in predicting Airline Services using Twitter data. Meanwhile, in recent years, Deep Learning algorithms for Sentiment Analysis has emerged as one of the most popular algorithms, which provides automatic feature extraction, rich representation capabilities, and better performance than most of the traditional learning algorithms. This research proposes deep learning with word embedding prediction model in order to improve the accuracy of the sentiment analysis of the airline services. The objectives of this research are as follows: to achieve the research aim, wherein firstly a taxonomy to classify different airline services models in the current literature based on sentiment classifiers is designed. Secondly, this research achieved the development of deep learning with a word embedding (DLWE) prediction model that extract features automatically through the layers in the model to classify and improve the prediction accuracy of sentiment analysis of the airline services. Thirdly, experiment to evaluate and compare the performance of the deep learning based sentiment classifier with some of the machine learning baselines has been conducted. For the experiment, the research used a public dataset that were extracted from the microblogging and airline services domain, while the training and testing process was performed using TensorFlow's library, for the evaluation process, classification accuracy and confusion matrix were employed. Regarding the results, the highest prediction accuracy achieved is 82.61% by using the word embedding model as a features extractor, and deep neural network learning as a classifier, as to when compared to the existing benchmark which used SVM and LG machine learning classifiers and word to vector model as a features extractor with 72% prediction accuracy. Thus, the result obtained from the proposed deep learning with word embedding model, can be used to help improve the sentiment analysis accuracy of airline services and other related fields of future research.

Keywords: Sentiment analysis, deep learning, word embedding, Twitter, airline services

ANALISIS SENTIMEN UNTUK PERKHIDMATAN PENERBANGAN DI TWITTER MENGGUNAKAN MENDALAM PEMBELAJARAN DENGAN PERKATAAN BENAMAN

ABSTRAK

Penggunaan platform media sosial dalam industri penerbangan telah meningkat pesat untuk membolehkan analisis memperkenalkan kualiti dan prestasi perkhidmatan. Peranan Analisis Sentimen (SA) adalah mengklasifikasikan pendapat orang dalam pelbagai kategori, seperti positif dan negatif dari teks, menggunakan algoritma yang sedia ada. Walau bagaimanapun, pendekatan yang sedia ada seperti model Bag of Words (BOW) sering digunakan untuk teks klasifikasi, di mana dokumen dipetakan ke vektor ciri sebelum pembinaan model sebenar, menggunakan teknik pembelajaran mesin, seperti algoritma Logistik Regresi dan Sokongan Vektor. Masalah ini menyebabkan ketepatan yang rendah dalam meramalkan Perkhidmatan Penerbangan menggunakan data Twitter. tahun-tahun kebelakangan ini, algoritma Deep Learning untuk Analisis Sentimen telah muncul sebagai salah satu algoritma yang paling popular, yang menyediakan pengekstrakan ciri automatik, keupayaan perwakilan yang kaya, dan prestasi yang lebih baik daripada kebanyakan algoritma pembelajaran tradisional. Kajian ini mencadangkan pembelajaran mendalam dengan model ramalan menaip perkataan untuk meningkatkan ketepatan analisis sentimen perkhidmatan penerbangan. Objektif o Kajian ini adalah seperti berikut: untuk mencapai matlamat penyelidikan, di mana pertama taksonomi untuk mengklasifikasikan model perkhidmatan penerbangan yang berbeza dalam kesusasteraan semasa berdasarkan kelas sentimen direka. Kedua, kajian ini mencapai perkembangan pembelajaran mendalam dengan model ramalan perkataan (DLWE) yang mengekstrak ciri secara automatik melalui lapisan dalam model untuk mengklasifikasikan dan meningkatkan ketepatan ramalan analisis sentimen perkhidmatan penerbangan. Ketiganya, eksperimen untuk menilai dan membandingkan

prestasi kelas sentimen berasaskan pembelajaran mendalam dengan beberapa pembelajaran baseline pembelajaran telah dijalankan. Untuk eksperimen, penyelidikan menggunakan dataset awam yang diekstrak dari domain perkhidmatan microblogging dan perkhidmatan penerbangan, manakala proses latihan dan ujian dilakukan menggunakan perpustakaan TensorFlow, untuk proses penilaian, ketepatan klasifikasi dan matriks kekeliruan yang digunakan. hasilnya, ketepatan ramalan tertinggi yang dicapai adalah 82.61% dengan menggunakan model embedding perkataan sebagai pengekstrak ciri, dan pembelajaran rangkaian neural yang mendalam sebagai pengelas, apabila dibandingkan dengan penanda aras sedia ada yang menggunakan pengkaji dan mesin pembelajaran SVM dan LG ke vektor model sebagai pengekstrak ciri dengan ketepatan ramalan 72%. Oleh itu, hasil yang diperoleh daripada pembelajaran yang dicadangkan dengan model embedding kata, boleh digunakan untuk membantu meningkatkan ketepatan analisis sentimen perkhidmatan penerbangan dan bidang penyelidikan masa depan lain yang berkaitan.

Kata kunci: Analisis sentimen, pembelajaran mendalam, penyemakan perkataan, Twitter, perkhidmatan penerbangan

ACKNOWLEDGEMENTS

I would like to firstly return all the glory and honour to God Almighty for granting me the privilege and opportunity to go through my master's program, most especially the period of my conducting and writing my dissertation, in good health and great success. He is the most beneficent and greatest help.

I would say a very big thank you to my supervisors, **Dr. Mumtaz Begum Binti Peer Mustafa**, and **Associate Professor Dr. Zarinah Binti Mohd Kasirun**, for their continuous guidance and assistance in my research work. Indeed, you have moulded me and helped me to appreciate my research work better.

If I don't appreciate my lovely mum, then I would be an ingrate. She has been so caring and always check on me to know how I fare and how the research is going, thank you for your support and unending love.

Special appreciation also goes to my dear husband for his all-round support, both financially, emotionally, among others. Thank you for making me believe in myself better, and being there when I was discouraged and confused.

Finally, my appreciations won't be complete if I forget to acknowledge my friends and colleagues, especially **Fatokun Faith Boluwatife**, and **Ali Baba** for your selfless assistance to me, and all those who have in one way or the other show me some love, and assisted me during my course of study, you are all indeed great, and I say thank you.

I pray that all those mentioned above, as well as those who haven't been mentioned, but have shown some love and care to me during my master's program, will all find everlasting joy and have good health. Thank you again!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

With the increasing number of airlines and the rising number of travelers around the world, some airlines are beginning to use different ways to promote themselves and attract more people, and they do this through the upgrade of the services they provide, with the aim of satisfying the needs of their customers (Bott, 2014). The level of this service is considered as one of the most important factors which represent the final impression of the passengers on a whole travel experience.

Customer service starts from the moment the passenger enters the company's website until he or she receives the last bag in preparation to depart from the airport (Misopoulos, Mitic, Kapoulas, & Karapiperis, 2014). Therefore, airlines engage in verifying the opinions and observations of customers about the services they provide, whereas these reviews are very important, as it is a means through which the quality of airline services will be developed and improved (Horiguchi, Baba, & Kashima, 2017).

In the past, customer opinions are collected and analyzed using traditional methods such as questionnaires, which are considered inaccurate, stressful, and time consuming (Liou & Tzeng, 2010). In contrast to questionnaires, social media is an important source of data to determine the customers sentiment and reviews regarding airline services (Almas, 2017). The rapid increase in the usage of the internet and social media has boosted the economy in airline product sales and profitability (Wang, 2017). The internet has become the central tool for the business strategy development of the airline company.

Implementation of social media in marketing has been defined as a ‘connection between brands and consumers (Paquette, 2013). The adoption of social media technologies has already become the driving force for establishing and reinforcing the competitiveness among large number of companies across numerous industries and domains (Plant, 2016). Thus, the potential of using social media as a marketing tool has been recognized by a vast majority of organizations, who have applied such tool into their business strategies, to maintain and improve competitiveness in the market (Plant, 2016). For instance, valuable information can be obtained using a marketing survey conducted by online users, customer feedbacks, and reviews written by the consumers. The survey can then be analyzed to obtain the knowledge about the products that supports the business decision makings process (Khan, Borah, & Pradhan, 2016). From the companies' perspective, it can be extremely valuable for their future business strategies, such as for determining the customer's post purchase responses and reviews. Interacting face-to-face with individual customers could be very time consuming and challenging, especially when there is an attempt to determine the customers' opinions on the products sold (Duan, Chen, Liu, & Ding, 2015).

Thus, the use of social media tools can be leveraged for essential business strategies, such as interacting with customers and monitoring targeted online customer groups (Bonchi et al., 2011). Among the existing social media platform, Twitter is a good platform as a huge repository of many important data, as it is a major destination for data analysts and researchers, thus suitable for performing the sentiment analysis. Twitter is more accurate than questionnaires (Okazaki et al., 2015). The emergence of modern techniques specialized on data collection has led to the widespread of data accumulation in various fields (Miedema, 2018).

1.1 Background

Sentiment Analysis, which is also known as opinion mining is a branch of Natural Language Processing (NLP), designed to build systems capable of identifying and extracting opinions from different text sources (Leung & Chan, 2008). The most commonly used method in sentiment analysis is the text classification, which classifies text into classes or polarity classes, such as positive, negative or neutral (Leung & Chan, 2008). The applications of text classification were used across multiple domains such as business, communications, and politics. For example, sentiment analysis can be applied on social media data to find out people's opinion of a certain candidate in the elections, or identifying areas that supports a candidate more than others, and up to the extent of even predicting the election results (Bahrami, Findik, Bozkaya, & Balcisoy, 2018).

Sentiment analysis application is extensively used to determine the predictive power towards sales in a company, as well as in the products domain, where companies wish to release their product to know the customer's feedback and reviews on a particular product, through their interaction on social media, by applying sentiment classification (Gautam & Yadav, 2014).

Twitter is one of the most widely used social media data in sentiment analysis, where Twitter data is used to predict various domains such as baby clothes or accessories, movies (Gautam & Yadav, 2014; Dijkman & Ipeirotis, 2015), and public protests which often lead to violence and destruction, making it a necessity for these protests to be predicted in advance, so as to avoid damages (Bahrami et al., 2018).

Furthermore, travelers also make use of the Twitter platform to share their opinion and experiences with airline companies; in reverse, the airline companies use this twitter data to analyze customers feedback in order to improve their quality and services to achieve customer satisfaction (Paper & Inspired, 2017).

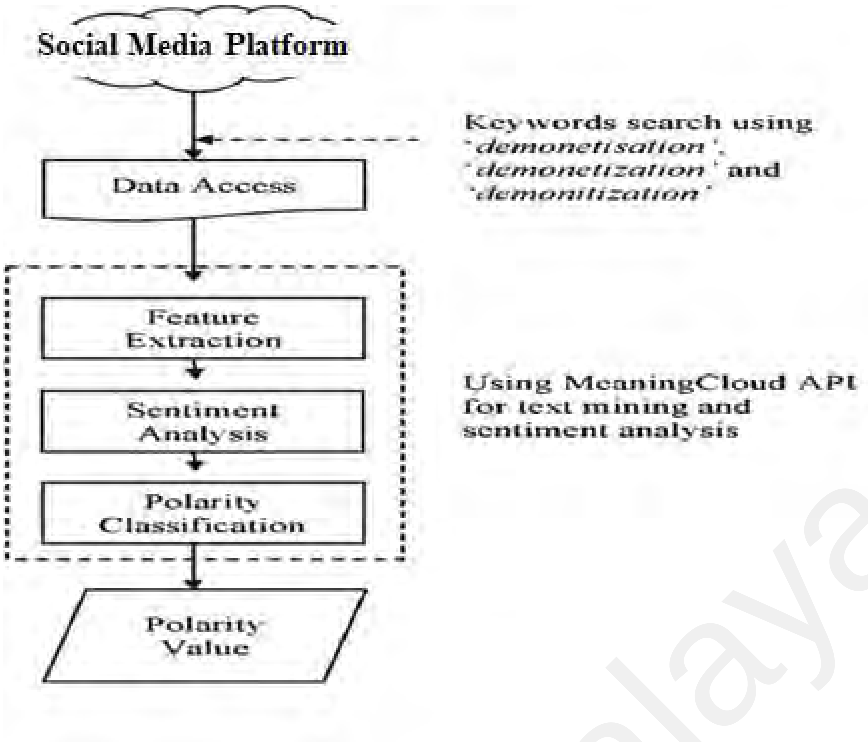


Figure 1.1: Framework of sentiment analysis (Darliansyah, Wandabwa and B, 2018)

Figure 1.1 depicts the common framework of sentiment analysis based on Twitter. From Figure 1.1, the first step in the sentiment analysis is to fetch data from the social media platform. The second step is the feature extraction, which refers to the transformation of input data into a set of features, after which the sentiment analysis takes place, which is a special task of the text classification, as it's a process of determining whether a written tweet is positive, negative or natural. The last step is the polarity classification, which refers to the identification of the sentiment's orientation (positive, neutral, and negative) (Darliansyah, Wandabwa and B, 2018). It is important also to note that a lot of research has been carried out in the field of sentiment analysis based on Twitter data, via the use of the same framework (Omar, Njeru, Paracha, Wannous, & Yi, 2017; Data, Using, Framework, Science, & Gurgaon, 2016; DEITY, 2014).

Thus, the use of social media tools can be leveraged for essential business strategies, such as interacting with customers and monitoring targeted online customer groups

(Bonchi et al., 2011). Twitter is considered as a suitable social media platform as it has a huge repository of many important data, as it is a major destination for data analysts and researchers, especially those interested in sentiment analysis, reason being that it is considered more accurate than questionnaires (Okazaki et al., 2015).

The emergence of modern techniques such as specialized on data collection has led to widespread data accumulation in various fields. Traditional data query methods are inadequate to extract useful information from large databases of various types. The Internet has resulted in an exponential increase in data shared by users, making it increasingly difficult for humans to manage and analyze the data (Okazaki et al., 2015). As a result, data mining science is becoming more important to deal with large volumes of data through the use of various exploration techniques. More particularly, sentiment analysis techniques have the ability to assist researchers and decision makers in airway companies make sense of customer reviews (Paper and Inspired, 2017). With the inclusion of the classification technology used in this research, it is classified in the field of predictive learning based on previous information, in order to predict future information (Acosta et al., 2017). Also, it produces predictive models that define categories or classes for diverse purposes based on the features and attributes that have been identified in the learning or training stage, which is very important in accuracy improvement (Gautam and Yadav, 2014).

In the past, most of the sentiment analysis researchers employed the machine learning approach (ML) to build the predictive models (Gautam and Yadav, 2014). Currently, with advances in data analysis, many have begun to use the deep learning approach (DL) to improve the performance and accuracy, due to classification problems (Araque et al., 2017). However, there are many sentiment classification researches, which focuses on various domains using Twitter data such as products, movies, among others. Nonetheless, very few works that have been carried out on twitter as the source of sentiment analysis of airline services, which requires a new approach to improve the predictive model due to lack of classification accuracy (Miedema, 2018; Chatterjee et al., 2018).

1.2 Research Motivation

Airline services companies must decipher a substantial amount of client input about their items and services. In any case, normal strategies to gather clients' feedback for aircraft service organizations is to examine through appropriating and gathering questionnaires, which can be tedious and inaccurate.

It consumes time and energy to disperse and collect surveys from clients and furthermore, it will require an excess of effort to record and document those questionnaires considering the number of travelers who take flights each day. Apart from that, not all clients will pay attention to questionnaires and simply fill questionnaires arbitrarily, resulting in inaccurate data for sentiment analysis. In contrast to an investigative questionnaire, twitter is a vastly improved information source for classification in the airline services. As a result, advancements in data analysis, Big Data, are turned out to be simple in gathering a huge number of tweets and performing information analysis. This has spared tons of work costs which questionnaire investigations may require. Above all, individuals post their certifiable sentiments on Twitter, which makes the data more exact than investigative questionnaires. In questionnaire investigations, there are different impediments that the inquiries are good, but it is difficult to uncover the data. Accordingly, text sentiment analysis has recently turned out to be exceptionally famous for automatic consumer loyalty analysis of online services. Sentiment analysis is a branch of information mining that is used for the analysis of large-volume data to uncover hidden vital information. Clearly, the upsides of automatic analysis for large datasets make sentiment investigation ideal for aircraft services.

Sentiment classification techniques are capable of supporting analysts and leaders in airline companies to comprehend and improve client feedback and satisfaction. Scientists and administrators can use these procedures to consequently sample the clients' input on micro-blogging sites like Twitter. These techniques can be used to develop business analysis software. Lots of researches on text classification and sentiment classification have been conducted to analyze airline sentiment analysis on customers. It is additionally important to build up another approach to improve the classification accuracy.

1.3 Problem Statement

Text sentiment analysis has become very popular in recent years for automatic customer satisfaction analysis of online services. Obviously, the advantages of automatic analysis of massive datasets makes sentiment analysis to be preferred by airline companies (Wan & Gao, 2016). Sentiment classification techniques help researchers and decision makers in airline companies to have a better understanding of customer feedback and satisfaction; also researchers and decision makers can utilize these techniques to automatically classify customers' (Tripathy & Rath, 2016).

Few work has been done on twitter sentiment classifications of airline services (Wan & Gao, 2016). Conventional sentiment classification approaches, such as Naive Bayes, have been applied to some tweet data and the performance was acceptable (Gautam & Yadav, 2014). The overall performance of their work was over 80%, however, some limitations existed in the work, as there was reported to have been a lack of direct comparison with some other classifiers which have the incremental abilities, some of which are the Incremental Tree Induction, which is based on fuzzy logic. Also, the other advanced techniques with incremental classifiers could be able to predict and mine the satisfaction of users via online reviews, instead of manual methods as used in this research (Gautam & Yadav, 2014).

There are many related research efforts that have been conducted, and that employed Twitter data of an airline company, as well as performed a series of sentiment classification techniques using binary classifiers such as Naive Bayes, SMO (SVM), Random Forest to evaluate the public opinion on airline services (Lacic, Kowald, & Lex, 2016).

However, there is a lack in evaluating and improving the accuracy of the methods in sentiment analysis. (Acosta, Lamaute, Luo, Finkelstein, & Cotoranu, 2017) used the existing machine learning approaches, and more specifically, the algorithms used included Naïve Bayes, Logistical Regression, and Support Vector Machine algorithms. However, the implementation achieved only 72% accuracy, using Logistical Regression and Support Vector algorithm. The reason for low accuracy could be due to the fact that machine learning approaches make frequent use of the Bag of Words (BOW) model, even though it is an efficient and easy approach, yet there is a loss of a reasonable amount of the original natural language obtained information (Xia & Zong, 2010).

1.4 Research Questions

RQ1: What are the limitations of existing data classification techniques for prediction of airline services sentiment?

RQ2: What is the suitable sentiment classification technique for the extraction of customer sentiment in airline services?

RQ3: What is the performance improvement of the developed/implemented sentiment classification technique?

1.5 Research Objectives

The aim of this research is the proposal of a predictive model using suitable technique to identify the customer's sentiment polarity in airline services. To achieve this aim, the following objectives are highlighted:

1. To review existing data classification using deep learning for sentiment analysis.
2. To propose deep learning sentiment analysis technique with automatic feature extraction to classify customer sentiment of airline services via Twitter.
3. To evaluate the performance of the proposed deep learning sentiment technique for airline services.

1.6 Research Scope

This study focuses on predicting the sentiment of airline services from twitter data and designing a suitable classification technique that can achieve the best classification accuracy rate for the sentiment of airline services. The dataset was collected from Kaggle, the number of airline service providers obtained are Virgin America airline, United airline, Southwest airline, Delta airline, US Airways and American airline (Acosta et al., 2017). In this study, deep learning neural network with word embedding model was employed for the model design. Also, classification accuracy and confusion matrix were used to measure the performance of the model.

1.7 Research Significance

This study can help researchers and decision makers in airline companies to have an understanding of customer feedback and satisfaction. Researchers and decision makers can utilize these techniques to classify sentiment from customers automatically.

It also helps airline companies to identify the customers' needs and desires as well as make effort to acquire and retain the customer's quality. Furthermore, this study helps the airline companies by providing useful information to set appropriate policies in making sure that the customers' experience a maximum level of satisfaction.

1.8 The Structure of the Study

This section discusses the purposes of each chapter of this study. The dissertation contains six chapters. Chapter 1 introduces the research, with the inclusion of the introduction, background study, research motivation, and problem statement, coupled with the objectives, questions and scope of the study. Chapter 2 discusses the related works on sentiment analysis. Precisely, the chapter reviews the related work on machine learning classifiers, airline sentiment as well as deep learning. Chapter 3 discusses the proposed research methodology in details. The chapter describes the four vital steps used in the study. Chapter 4 discusses the experimental design of the study by distinctively explaining the implementation process and procedures used in obtaining the desired research objectives based on the methodology in chapter 3. The results and findings from the implementation of the experimental design is discussed in Chapter 5. This involves the comparison of the predictive performances of the stated classification models proposed and used in this research. As a wrap up, chapter 6 gives a brief conclusion, together with some future work suggestions.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter presents and discusses the review of related literature in the area of sentiment classification processes and analysis, in advanced sentiment classification techniques, and the application of sentiment classification in the airline services sphere. The discussion emphasizes on the present realm of specific sentiment analysis techniques using information extracted from Twitter.

2.2 Sentiment Classification

2.2.1 Definition of Sentiment Classification

Sentiment classification is a procedure of extracting the primary opinion of the provided text or document. It is described in the area of natural language processing and/or opinion mining. (Pang & Lee, 2008) informed that the phrase opinion mining or sentiment analysis was first introduced in the field of marketing research by Das and Chen, who were interested in finding the market sentiment through web technologies (Gupta & Lehal, 2014).

In the early stages of sentiment classification research, the phrase text categorization is commonly used in the field of study. Text categorization is the classification of documents into predefined categories (Sebastiani, 2001); for instance, book genres can be categorized into action, novel, etc.

The understanding of the text often needs assistance from a variety of external knowledge, as the text contains limited information and occasionally, comes with uncertainties. Therefore, it is necessary to learn how to add more information to a deeper learning process. Fortunately, this requirement may be fulfilled because of a variety of textual knowledge. First, because the text is based on morphology and linguistic rules, including a popular syntactic and morphology knowledge (Cao & Lu, 2017).

Spelling information can be part of a text (POS) tag and can contain different speech rules, such as the relativeness and superiority of the synthesis, past and participation of the verb, as well as the plural form of a name (Cao & Lu, 2017).. Secondly, there are rich research trends on semantic knowledge learning from a wide range of text information, for instance, WordNet (Gautam & Yadav, 2014)Freebase ((G. Hu, Bhargava, Fuhrmann, Ellinger, & Spasojevic, 2017)), and Probase (G. Hu et al., 2017). Such semantic information has the tendency to reflect the category of subjects and words / subjects, like the synonyms and the antonyms, fitting in to the end. For instance, Portland - Oregon; Portland is the city. Assuming the existence of morphological, syntactic and semantic knowledge, a critical problem will continue to be developed to utilize a different approach of deep learning algorithms and systems to produce a high quality vocabulary.

Similar discussion and papers were published in the Natural Language Processing (NLP) field by (Pang & Lee, 2008). The definition of opinion mining is also described as a new sub-field at the crossroad of data fetching and computational linguistics related to the content of a document (Vinodhini, 2012).

2.2.2 Natural Language

The sentiment classification applications and techniques are developed as a result of certain needs. Primary motives are a fast surge of sentiment text information produced in social media with the potentiality for the text classification suggestively important to different aspects of managerial decision.

In this research, the need arises to discuss the causes of the formulation of these classification techniques, that is, opinion rich text data itself. In the area of mining data, the analysis of the set of data is always crucial in mining information from the unstructured form. But, textual data is always unstructured due to the complex nature of the data in the actual situations and the vagueness of the human language. Hence, this requires in-depth analysis and pre-processing prior to classification (S. Gharehchopogh & Khalifehlou, 2011).

The textual data do not follow statistical rules, but grammar and syntax rules are defined in different languages accordingly. There are many challenges and issues while using textual data that is unstructured (S. Gharehchopogh & Khalifehlou, 2011). The main issues in text mining are identified as follows;

1) Stop words considerations
2) Stemming words considerations
3) Existence of noisy data in text.
4) Word sense disambiguation
5) Part of speech tagging.
6) Technical or compound terms
7) Tokenization considerations
8) Word context, order, and background knowledge.

The above list stated the considerations that needed to be addressed during the processing of text-based information for sentiment classifications. Whenever there is a consideration of the natural language matter adeptly, there is an increase in the performance of the results of the sentiment classification. Natural language is proven to be an effective approach in many research works (VijayGaikwad, Chaugule and Patil, 2014; (Savoy, 2012). Also, in the area of sentiment classification, many researchers adopted the Natural Language Processing techniques to address the issues in the textual data. The application of NLP in sentiment classification is reviewed in the next section.

2.3 Sentiment Classification Process

2.3.1 Sentiment Subjectivity Classification

The process of sentiment classification assumed that the opinions already exist in the text, when classifying the underlying sentiment for the specific piece of text. In sentiment classification, it is necessary to identify the individuality of the text and unbiased evidence (Liu, 2010). There are cases where provided document does not contain any opinions or polarities. The given documents are objective and only explain the fact of some matters. For instance, most articles from newspapers are stating the facts that already happened in the past, meaning that there are no indicators of bias behind the facts described. If the documents are not bias, then the opinion of the document cannot be correctly extracted. Identifying the subjectivity within the given text is essential in the sentiment analysis process (Liu, 2010). Many published research papers have suggested in relation to separate the subjective information and opinions of the documents. According to some scholars, determining the strength of subjectivities within deep clause-level of each text can improve the accuracy significantly over baseline (In & Other, 2016).

2.3.2 Sentiment Polarity Classification

The polarity classification is one of the categories in sentiment classification. Primarily, the purpose of the polarity classification is to assure the document under consideration belongs to the class of either positivity or negativity in the opinion expressed. For example in an email spam filter, whether it is spam or not and in product reviews, whether the review leads to a positive sentiment or negative sentiment. The sentiment classification type is referred to as polarity classification (Pang & Lee, 2008). Nonetheless, this classification is similarly ruled as binary classification, because there are only two types of classification results. As for classifications like categorizing search engine query results into multiple predefined categories, it is often referred to as multi-class categorizations.

2.4 NLP and Sentiment Analysis

One of the concerns discussed in the previous section is the consideration of stop words in the textual dataset. For sentiment classifications, stop words such as ‘a’, ‘the’, ‘be’ etc. in the content of a sentiment classification, have less impact on the final result of the sentiment. However, sometimes stop words can be relevant to the content of the document that can influence the sentiment classification result, depending on the objective. For example, when the sentiment’s strength is considered as one of the classification results, the word ‘very’ should not be removed in the text, as it can be treated as a solid indicator of the sentiment polarity strength.

The stemming technique in NLP has provided a way of reducing the variation of the terms used in the text. For instance, the word ‘cancelled’, ‘canceled’, are all replaced with ‘cancel’. The purpose of implementing the procedure in the NLP is tantamount to decrease the number of attributes used during the process. However, it still depends on the objective of the text mining task. This process also reduces the information provided for the classifier when training and classifying text. The noisy data in the text refers to the incorrectly spelled words, shortened terms such as abbreviations and acronyms, and mark-up language tags. These noisy data normally exists in unprocessed data, which requires correction of misspelled words and expansion of abbreviations and acronyms (Stavrianou, Andritsos, & Nicoloyannis, 2007).

The word sense disambiguation (WSD) refers to the same term used in a different context, which is, expressing completely different meanings. One typical example can be the term ‘touch down’, it refers to scoring a goal in sports and it could also mean the airplane landing on destinations. The use of this term in a different context could be resulted in distinct meanings. The word sense disambiguation has been an active research topic in recent decades and challenging task to complete with ideal results.

Many approaches have been developed over the years and they all follow the rules of using external lexical resources to determine the actual meaning of the term. However, applying different methods from supervised machine learning and semi-supervised machine learning approach to unsupervised machine learning approach (Pal & Saha, 2015). The WSD in text mining tasks requires special considerations, because of the domain dependencies of the different terms used. The part of speech tagging is to determine the grammatical class of the term in the text. In the area of natural language processing, tagging part of a speech is typically used in text mining, because the same term can make completely different senses in different grammatical classes.

For instance, word ‘bass’ is referring to the lowest singing voice in musical domains from the context of an adjective, but also refers to a kind of freshwater perch in the noun context. In sentiment classification, tagging the terms can influence the overall sentiment, since it has different sentiment scores in the corpus, such as SentiWordNet (Haque and Hoque, 2015). The technical terms used in sentiment classifications are considered accordingly, the reason being is that the impact of using technical terms is small when only classifying the overall polarity of the text, because they are not always indicators of the emotions in the text.

The considerations of tokenization in text refers to the tasks of converting the textual data into different collections of words or terms as tokens, which are then being used for text mining, thus increasing the overall performance of the sentiment classification process and providing an easier manipulation process for text mining (Savoy, 2012). Sometimes in long documentation classification processes, the tokenization process can separate the whole document into different segments that can be considered and analyzed individually depending on the objective of the data mining task. The tokenization of the text can be represented in different formats.

2.5 Text representation in sentiment analysis

Sentiments analysis from twitter data using deep learning classification technique of auto-encoder was used by (Almas, 2017) to classify sentiment overall opinion and attitude of people which involve emotional dispositions formed towards an object over time such as sentiments that included hate, friendship, loyalty, patriotism, etc., the accuracy for sentiment classification performance on Twitter data was 80%.

The most common text representation used for sentiment classification is the word vector model proposed in (Turney & Pantel, 2010). The implementation of converting textual data into word vector has created the opportunity for applying statistical analysis in textual data. Table 2.1 shows the example of word vector text representation.

Table 2.1: Word matrix example

ID	Thanks	Cancel	Flight	Delay	Book	…..
1	1	1	0	0	0	…..
2	0	0	0	1	0	…..
3	0	1	0	0	0	…..
4	0	0	1	0	0	…..
5	0	0	1	0	1	…..

2.6 Sentiment Classification Approaches and algorithm

There are many approaches and algorithms to perform sentiment analysis systems, which can be categorized as:

1. Rule-based approaches which implements sentiment analysis based on a group of manually set rules (Liu, 2010).

2. Analysis method based on automated learning approaches (machine learning / deep learning); this method needs a data to train the classifier, the primary goal of this approach is the process of data collection from various sources available on the internet; after which the data can be processed and prepared in order to train and test the classifier (Almas, 2017).

3. Hybrid system; this combines both automatic and rule based approaches.

The current study focuses on an analysis method based on automated learning approaches (machine learning / deep learning) (Araque, Corcuera-Platas, Sánchez-Rada, & Iglesias, 2017).

2.6.1 Sentiment Machine Learning Classification Techniques

There are many variants of ML techniques to be applied as solution to classification problems. However, it is not easy - if not impossible - the current technique is to decide which one is better than others (Tripathy & Rath, 2016). The difficulty of choosing the appropriate classifier is usually dependent on the nature of the problem, including data, volume and quality (Tripathy & Rath, 2016).It also depends on the amount of time that is required to prepare the model. However, there are other considerations that need to be taken into account when selecting techniques like accuracy, number of parameters, number of properties and linearity (Wang, 2017). Recreation(N. Hu, Koh, & Reddy, 2014), Logistic regression, Naïve Bayes Classifier, Decision Trees and Support Vector Machines, are some techniques that can be used in classification. Table 2.2 summarizes the existing classifications and their work on various textual classification issues.

Table 2.2: Sentiment Classification Techniques

Authors	**Domain & purpose**	**Data/source**	**Technique**		**Performance of classification Accuracy**
			Feature Extraction	Classifier	
Wang, J., Cong, G., Zhao, X. W., & Li, X. (2015, February).	Not specified	Twitter data 3000 tweets	Parts of Speech Tagging	Semi supervised Graph	Food 54.63% Travel 58.64% Self 45.73% Goods 43.25% Event 27.13% Trifle 20.04% Non intent 35.56%
Dijkman, R., Ipeirotis, P., Aertsen, F., & van Helden, R. (2015).	books and movies	Twitter data 12,780	-	decision tree	Type of Tweet 86% Type of User 86% Sentiment 76%
Tripathy, A., Agrawal, A., & Rath, S. K. (2016).	movie reviews	IMDb dataset 50,000 reviews	n-gram techniques	Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM)	83.6% 88.5% 85.1% 83.7%
Hamroun, M., Gouider, M. S., & Said, L. B. (2015)..	electronics and retail banking	Twitter data 16 million tweets	lexico semantic patterns	OWL ontologies	precision: 55.59%, recall: 55.28%
Mahmud, J., Fei, G., Xu, A., Pal, A., & Zhou, M. (2016, March).	Airline Brand	Twitter data 7534	Unigram Context-based Sentiment/Opinion Domain-specific sentiment Length of use Frequency feature	binary classifiers	(60-65%).
Acosta, J., Lamaute, N., Luo, M., Finkelstein, E., & Andreea, C. (2017).	Airline	Twitter 14,640	bag-of-words (CBOW) skip-gram (SG)	Naïve Bayes, Support Vector Machines and Logistic Regression.	72%

2.6.2 Sentiment Deep Learning Classification Techniques

2.6.2.1 The Advent of Deep Learning

The advent of deep learning has drawn attention to train complex and deep models on huge amount of data (Socher, 2016). The underlying reason for deep learning is to compute dispersed unstructured information in the form of continuous vectors (Araque et al., 2017). The conventional NLP approaches typically describe words that are expressed in terms of words, but the words learned from deeper learning with words are a multitude of semantic relationships, as well as a new placement of legitimacy and patterns of language target specific coding (Bian, Gao, & Liu, 2014).

Most of the available work uses extensive in-depth learning algorithms to demonstrate success in speech and image fields to study text and text-based tasks (Wehrmann, Becker, Cagnini, & Barros, 2017). (Zhang, Zhang, & Vo, 2016) proposed a broadly used model architecture to evaluate the model for neural network language. Subsequently, in specific studies (Zhang et al., 2016) used related nerve networks architecture to study the vocabulary for improving and simplifying NLP applications. It has recently been proposed to study similar but effective ways to achieve syntactic and semantic word similarities in the two models (Zhang et al., 2016).

These efforts will become a common ground for enhanced learning, and the text creates a unique characteristic of other fields, such as text and speech (Araque et al., 2017). In the field of speech and visual arts, however, the ability to find significant indicators from successful noisy entries, the main problem for the comprehending of the text is defective data and semantic uncertainty (Sundermeyer, Schlueter, & Ney, 2012).

2.6.2.2 Methods in Sentiment Deep Learning Classification

Many of the deeper studies in the area of NLP are focused on methods involving the study of vocabulary vectors using the Neural Languages model (Kim, 2014). As a vector, the continuous expression of words has been established as an effective tool for various NLP processes, comprising sensory analysis (Tang, Wei, Qin, Liu, & Zhou, 2015) . This can be regarded as word2vec is one of the most famous approaches that can process words as vector (Mikolov, Yih, & Zweig, 2013). Word2vec is centered on Skip Grams and CBOW models to complete the distributed authorization. Although CBOW aims to interpret a word with its content, SkipGram defines a context in a single word. Word2vec constitutes a vast array of words in a continuous vector expression of words. Calculated word vectors preserve a large quantity of syntactic and semantic concepts that are expressed in the vector field of communication (Mikolov et al., 2013).

At the level of these word-level combinations, the embedded $W \in IRd \times | V |$ is encoded with column vectors, where $| V |$ denotes the word of wealth. Each pointer to the word $WI \in IRd$ corresponds to the i-th angle vector of words. Converting the word "V" to the conversion of the word vector to rw using a matrix vector product: $rw = Wvw$ where vw is a $| V | |$ and w is the value index and the remainder is zero. The matrix component W is a hyper-parameter for the measurement of the studied parameters and the word vector d. The vector representations calculated using these methods are very effective when using conventional classical (eg, logistic regression) emotional classification, as indicated by (Karmaker Santu, Sondhi, & Zhai, 2016).

The Word2vec-based method is doc2vec (Carli, 2014), which models all the essences or documents as vectors. An added method of vocational training is auto-methodology, i.e. the type of widespread neural network used to untrained education. Automatic coding devices have been used to explore new powers for varieties of automated learning tasks, such as (Chen, Weinberger, Sha, & Bengio, 2014). An exciting approach to deep learning for SA is to extend the knowledge available in vectors to other sources of information. This additional information can be found in (Tang *et al.*, 2014), or similar handwriting features and combinations of these emotions (Tang *et al.*, 2014).

In the works presented by (Karmaker Santu et al., 2016), the features derived from the vocabulary are enriched with animated themes and incorporated into the ensemble chart. They also show that these enriched authorities are effective in increasing the effectiveness of the classification of polarization. Another approach that adds new content to the film is described in (Severyn & Moschitti, 2015), where an in-depth study is used to capture the sensory properties, depending on the semantic properties. (Karmaker Santu et al., 2016) describe a method that uses remote control information to retrieve neural network parameters from untested neural language models. You can also use the shared filtering algorithm as described in more detail with (Kim, 2014), where the authors added emotions from a small portion of the information. Adding Sample Data (Zhang et al., 2016) describes how to use the Rheumatic Neural Network (RNN) emotion in parallel to another neural network architecture.

Generally speaking, there is a growing tendency to add more information to the layout of words created by the in-depth studied network. Interesting business is an interesting job described in (Vo & Zhang, 2015), where both emotional and standard additions are used in conjunction with a variety of functional features to capture the targeted intentions of Twitter's comments. Enrichment of information in word placement is not the only trend for deep studying of SA. As Vo & Zhang (2015) have shown, the study of the composition of emotion classification has been proven to be relevant.

2.6.2.3 Deep Learning with Word Embedding

Most of the existing works employed the technique of deep learning with word embedding, and these have been proven to be successful in various domains in sentiment analysis. Tang et al. (2014) developed a deep learning system using Deep Neural Network to learn word embedding in order to improve and simplify Twitter message sentiment classification. Thereafter, (Araque et al., 2017) developed a deep learning based sentiment classifier using a word embedding's model, and applied it on seven public datasets, which were extracted from a movie reviews domain and other microblogging, all with the purpose to improve the performance of deep learning techniques. Most recently, (Deng, Lei, Li, & Lin, 2018) introduced an improved deep neural network approach to learn about word embedding in order to predict an employee's future career details, based on the online resume data. However, there is a lack in evaluating and improving the accuracy of this method in sentiment analysis. Table 2.3 summarizes the existing deep learning with word embedding and their work on various sentiment analysis domains.

Table 2.3: Sentiment Classification Using Deep Learning with Word Embedding

Authors	**Research Focus**	**Features/Domain**	**Classifier**	**Classifier Performance**
Araque, O., Corcuera-Platas, I., Sanchez-Rada, J. F., & Iglesias, C. A. (2017).	Enhancing deep learning sentiment analysis with ensemble techniques in social applications Oscar	Word embedding/ movie reviews domains	Deep classifiers (CEM ^{Mel} $_{SGA}$, M_{GA} , M_G)	CEM MeL SGA deep learning classifier was best performance with 94% accuracy.
Tang, D., Wei, F., Qin, B., Liu, T., & Zhou, M. (2014, August)	develop a deep learning system for message-level Twitter sentiment classification	Sentiment-Specific Word Embedding (SSWE) / message-level Twitter sentiment classification	Deep Neural Network(DNN)	The DNN with word embedding performed better for positive ,negative, neutral with average of 86% accuracy
Deng, Y., Lei, H., Li, X., & Lin, Y. (2018, May).	Predict an employee's future career details, based on the online	Word Embedding/Job matching which	Improved Deep Neural Net-work(IDNN)	Improved DNN model is more effective and accurate compared to

	resume data using deep neural network Model			different machine learning methods
Yu, L. C., Wang, J., Lai, K. R., & Zhang, X. (2017)	Refining word embedding using intensity scores for sentiment analysis	Word Embedding/ Scores for Sentiment Analysis	CNN FOR BINARY AND TERNARY CLASSIFICATION	The Refined Embedding yielded better performance of 90.5% by using CNN Binary classifier.

2.6.2.4 Deep learning framework

Deep learning techniques have two phases, which are the training phase and the inferring/prediction phase. During the training phase, the network tries to learn from the data. And each layer of data is assigned some random weights and the classifier runs a forward pass through the data, predicting the class labels and scores using those weights.

Deep learning has the ability to automatically discover the representations needed for feature detection or classification from raw data. This replaces the manual feature engineering and allows a machine to both learn the features and use them to perform a specific task. DNN and CNN are some examples of popular Deep learning techniques. (Deng et al., 2018) have successfully applied the DNN technique with an effective and accurate performance.

After the training is completed, the Inference / testing phase will start, it is the stage in which a trained model is used to infer/predict the testing samples and comprises of a similar forward pass as training to predict the results. Unlike training, it doesn't include a backward pass to compute the error and update weights. Figure 2.1 shows a typical Deep Learning framework to classify texts.

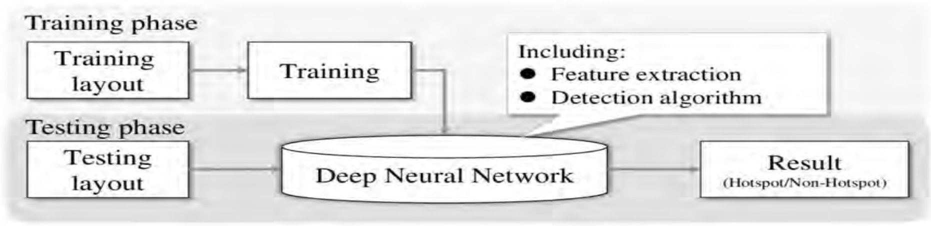


Figure 2.1: Deep Learning Framework to Classify Text (Matsunawa, Nojima, & Kotani, 2016)

2.6.2.5 Structure of Deep Neural Network

(Deng et al., 2018) proposed a deep multi-input neural network model that is able to predict an employee's future career details, which includes position name, salary, and company scale based on the online resume data. The model has five layers which are the embedding layer as input is at the bottom of the basic structure to turn positive integers into dense vectors of fixed sizes. The flattened layer will be taken as an input tensor by the next layer, the hidden layer, which is fully connected. It enhances the ability of mining hidden nonlinear characteristics for the model. Lastly, a merge layer under the SoftMax layer concatenates the different tensors output from two branches. Experiments demonstrates that improved DNN model is more effective and accurate. Figure 2.2 shows a Deep Neural Network model structure.

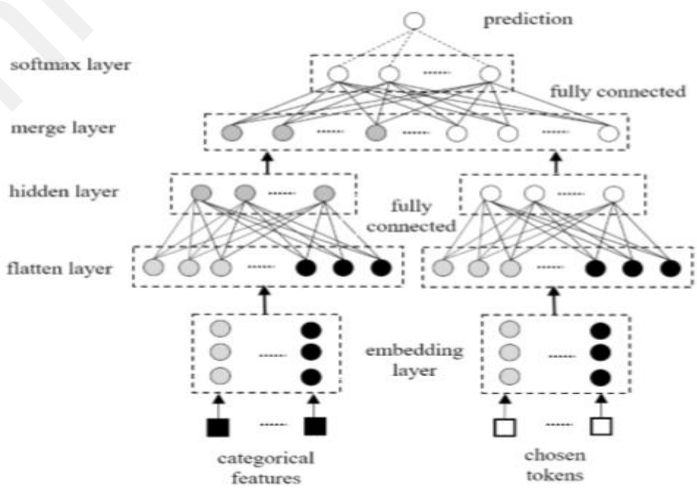


Figure 2.2: Improved Deep Neural Network Model (Deng et al., 2018)

2.7 Twitter Sentiment Classifications

Since Twitter has enabled public admittance to the stream and historical data, it has become a useful source for analyzing emotions, and has covered a lot of work in this area. Using the phrase ":-)" and ":-(", J.Read uses the Naive Bayes Approach to tweet and tweet them to positive tweets and negative tweets, and tweet to the Vector Machine. Wilson hashtags were used, and they attempted to resolve the issue of different data files, which led to the proposal of a general approach (Hoffmann, 2005). Wilson and other tweets containing positive emotions, negative emotions, and neutral feelings, have brought about the consideration of three polarities in the classmate classification, which are also classified as classifying features, and include emotions and text, other non-specific characteristics have as well been regarded, coupled with individual experiences.

It has shown that hashtag training data can regularly train better classifiers cure. But in the studies, the data were collected from libraries and the deprived of that hashtags and tweets were just a minor part of the tweets data from the real world. Pak and Paroubek suggested that the Twitter API should be able to classify emotion-based tweets and classify their sentiments (Pak and Paroubek, 2010). Based on the quiz outcome, the classifier that uses the large-scale functions has determined the maximum classification precision, since it takes a good balance between coverage and accuracy. The BBC's putting their business on the mine concept does not have a domain; that is, using their methods in domain-oriented mining helps produce dissimilar results. In addition, the origin of the data is similar because they have only received tweets containing phrases and neglected all the logos. In this case, they do not take into account the presence of neutral feelings, and the classification of these tweets is very useful for examining the tweet sense.

2.8 Sentiment Classification on Twitter Data about Airline Services

Sensitivity analysis in the airline companies is similar to the classification of product returns that are similar to the product review. However, there are many differences in domain concentration and result concentration. For example, classification of books from Amazon, "This book is worth reading." This is a positive view of the book. The word "book" is the name of this sentence, but the word "book" in the airline has more opportunities to play as "I cannot write this flight."; thus many studies have been conducted with different sources, and most of them are dependent on the domain.

There were few published studies on amateur emotion analysis using data from Twitter. Although there were published experiences, they were available online. These experiments used various technologies, such as Java, R, and Python. For example, the approach that is used as part of the classification process for Rni classification is recommended in the post (Oliver Breen, 2012). The results of the emotions are also well developed in his blog, which ultimately provides an intuitive insight into the public. Based on the evidence analyzed in these experiments, approaches to such classifications, including computer teaching techniques and lexicon-based methods are used. "One of Twitter's tweets is impressive when it comes to Twitter," says Elrhul, 2014. Twitter blogger Meghann Elrhoul has taken a photo with Millward Brown, showing how "travel" signs use Twitter to enhance their image. This blog pointed out that a software customer service strategy on Twitter may reflect the image of a direct travel brand Elrhoul, M. (2015, July 30).

The fact that official Twitter companies have an increasing number of active twitter accounts denied that this blog is trustworthy has proved the importance of adopting Twitter platform for travel companies. For example, many airline companies have been on Twitter for 24 hours, and they have responded to complaints and requests. Twitter also allows the company to deliver the latest promotional products to customers and directly connect with customer feedback or interpretations, which are other advantages of Twitter to the customers. In this context, applying automatic emotional analysis methods can be very effective when compared to using human resources when marketing analysis tasks are required for a specific advertisement. Additionally, many customers take into consideration the opinions of other Twitter users who have experience in purchasing products and services. This can have a significant impact on the consumer's decision, they plan to choose which airline.

Lee et al. (Pang, Lee, & Vaithyanathan, 2002) have used consumers as a source of information for analyzing their aviation services data. They explored twinets from two airlines of Malaysia, JetBlue Airlines and SouthWest Airlines. They used traditional text analysis techniques in the study of relationships between twitter users, and offered campaigns for airline companies. In their research, they did not classify emotions in the tweets that were easier to understand for the dealers. There was little work on the classification of airline companies in their twitter. Approaches to classifying traditional feelings, like the Naive Bayes method, was applied to some tweets, and the result was good (Pak and Paroubek, 2010).

Likewise, some studies presented an overview about the use of machine learning techniques for sentiment classification, and identified three classifiers; (Logistic regression and Support Vector Machine (SVM)) (Acosta et al., 2017). The researchers (Acosta et al., 2017) used the aforementioned classifiers (Logistic regression ,and SVM) on the US Airline dataset consisting of Tweeter posts in relation to users' experiences with U.S airlines. The dataset contained 15 attributes with 14,640 tweets in the original tweet text, the class label of the sentiment and the Twitter user data. This approach was assessed using accuracy measure and the outcome was 72%.

Table 2.4: Acosta Accuracy Results (Acosta et al., 2017)

Classifier	Training Model	Accuracy
Support Vector Classifier	Skip-Gram	72
Logistic Regression	Skip-Gram	72

TABLE II. Classification Report of Logistic Regression Classifier

	Precision	Recall	F1-score	Support
Negative	0.87	0.75	0.81	2750
Neutral	0.51	0.62	0.56	936
Positive	0.57	0.70	0.63	706
Average / Total	0.75	0.72	0.73	4392

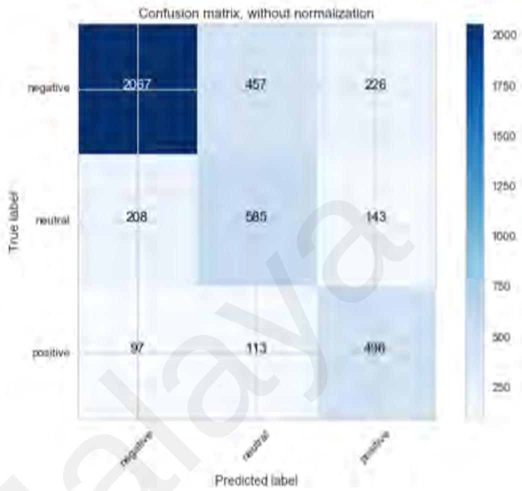


Figure 2.3: Classification Report and Confusion matrix of LG (Acosta et al., 2017)

As shown in Figure 2.3, the classification report details how the logistic regression classifier performed for each sentiment class. For the negative class, the classifier accurately predicted 75% of 2750 negative test instances correctly with a precision rate of 87% and an F1-score of 81%. For the neutral class, the classifier accurately predicted 62% of 936 neutral test instances correctly with a precision rate of 51% and an F1-score of 56%. For the positive class, the classifier accurately predicted 70% of 706 positive test instances correctly with a precision rate of 57% and an F1-score of 63%. Clearly, the logistic regression classifier along with other training algorithms struggled to precisely classify neutral and positive tweets accurately. Many instances that were incorrectly predicted were labeled as neutral. For example, in Figure 2.3, 457 out of 2750 negative tweets (17%) were predicted to be neutral while 113 out of 706 positive tweets (16%) were predicted to be neutral.

TABLE III. Classification Report of Support Vector Classifier

	Precision	Recall	F1-score	Support
Negative	0.87	0.76	0.81	2750
Neutral	0.52	0.63	0.57	936
Positive	0.59	0.70	0.64	706
Average / Total	0.75	0.72	0.73	4392

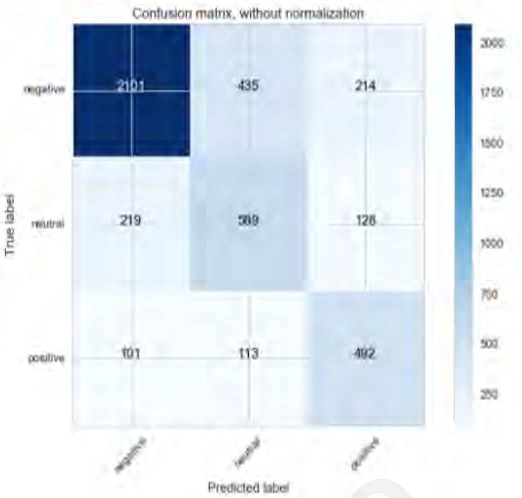


Figure 2.4: Classification Report and Confusion matrix of SVM (Acosta et al., 2017)

Clearly Support Vector Classifier, similar to the logistic regression classifier, struggled to precisely classify neutral and positive tweets accurately. As shown in Figure 2.4, many instances incorrectly predicted were labeled as neutral. Similar to the other classifier, shown in Figure 2.4, 457 out of 2750 negative tweets (17%) were predicted to be neutral while 113 out of 706 positive tweets (16%) were also predicted to be neutral.

2.9 Summary

This chapter reviews the related studies to identify the limitations of the existing sentiment classification techniques, particularly in the airline services domain. Many researchers have applied machine learning classification techniques in the airline services domain, however, there is a lack in evaluating and improving the accuracy of the methods in sentiment analysis. The literature review have focused on the existing solutions to find suitable technique(s) that can overcome the research problem. Based on the literature, the framework of Deep Learning for sentiment classification and prediction of text can be a suitable solution to overcome this problem.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Overview

This chapter explains how the research is conducted by introducing the methodology used to achieve the research objectives. The research proposes a predictive model based on deep learning technique to identify customer's sentiment polarity in airline services. The following research methodology is adopted to achieve the research aim. This includes several key steps such as the research problem, dataset collection, proposed solution and development, evaluation, and discussions of the major findings. Figure 3.1 provides the flow of the process of this study used in the objectives.

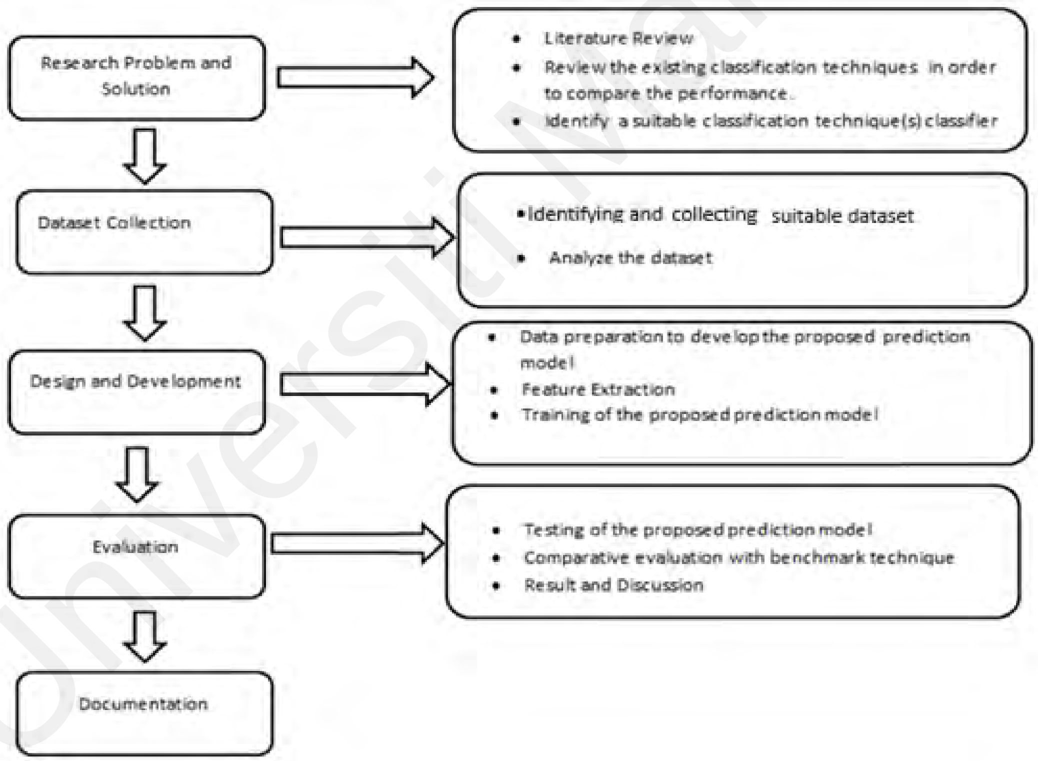


Figure 3.1: The Research Methodology Flowchart

3.2 Research Problems and Solutions

Acosta et al. (2017) used traditional machine learning approaches, included Naïve Bayes, Logistical Regression and Support Vector Machine algorithms (Acosta et al., 2017), In the model performance accuracy (72%), the training datasets that was used to train the word2vec skip-gram model was lower, and this had an impact on the overall accuracy of the testing data. Improving the predictive model is a challenging task due to the lack of accuracy of the performance (Acosta et al., 2017).

Many researches have been conducted in the area of sentiment analysis using various classifiers. Recently, deep learning is being increasingly considered as it is more robust as compared to other classical machine learning algorithms. Moreover most of the results show that the deep learning techniques significantly outperform many classical techniques (Lu et al., 2017; Araque et al., 2017; Chatterjee et al., 2018).

With reference to the current issues in airline services sentiment analysis, it is necessary to develop a prediction model that can predict the customer's sentiment of airline services, via the use of Deep Learning technique with word embedding, instead of the currently utilized Machine Learning techniques. The prediction model should have the ability to improve the prediction accuracy based on the previous result; therefore it must be able to learn to make the prediction of airline sentiment polarity, via training and testing processes.

The proposed prediction model would be able to classify and predict the airline services sentiment polarity. The proposed system's development includes data collection, data preprocessing, automatic feature extraction, prediction, and evaluation.

3.3 Dataset Collection and Analysis

By looking at advance techniques in the area of sentiment classification, the sentiment classification selection dataset relies on numerous features, the objective of the classification, the domain focus, as well as the data structure. In view of the objective and the primary focus discussed in Chapter 1, the dataset is required to be associated to sentiments on airline services strictly and consists of polarity sentiment regarding the service. As such, the initial step to the development of the proposed prediction system aimed to build a suitable text dataset. The dataset should contain the selected airline services data. In addition to that, the dataset must be cleaned and pre-processed to assure that all data extracted is appropriate and suitable. The airline services dataset must be labeled properly based on the meaning of the sentences. Finally, data must be converted to a compatible format that can be transferred and used for training and testing. As the popularity of employing Twitter data is increasing for sentiment classification purpose (Wang, 2017); (Wan & Gao, 2016), employing Twitter data is also considered in this study. The detail description about the dataset collection and preparation for this study will be discussed in detail in Chapter 4, Section 4.1.

3.4 Development of Sentiment Classification Prediction Model

The development of the deep learning with word embedding (DLWE) prediction model comprised of four main stages, initiating with the pre-processing of converting the raw data into a clean data set, followed by defining the Model architecture (deep learning model will be used to extract the feature vectors automatically with the embedding layer). The next process is the training of the model, and finally the estimation of the model's performance (by loading the test data and going through the pre-processing step and then predicting the classes using the trained model, after which the evaluation will take place).

This section describes the exemplary framework of deep learning model that can be used during the development of the proposed system. The proposed work classified several techniques to be applied in predicting the Sentiment Analysis for Airline Services. The selection is based on their overall performance whilst it is used in the problems regarding to the text classification. The consequence of this process is the selection of an appropriate technique to achieve the research's second objective. Implementation of the proposed prediction model consists of two main phases, which are the training phase and the inferring phase.

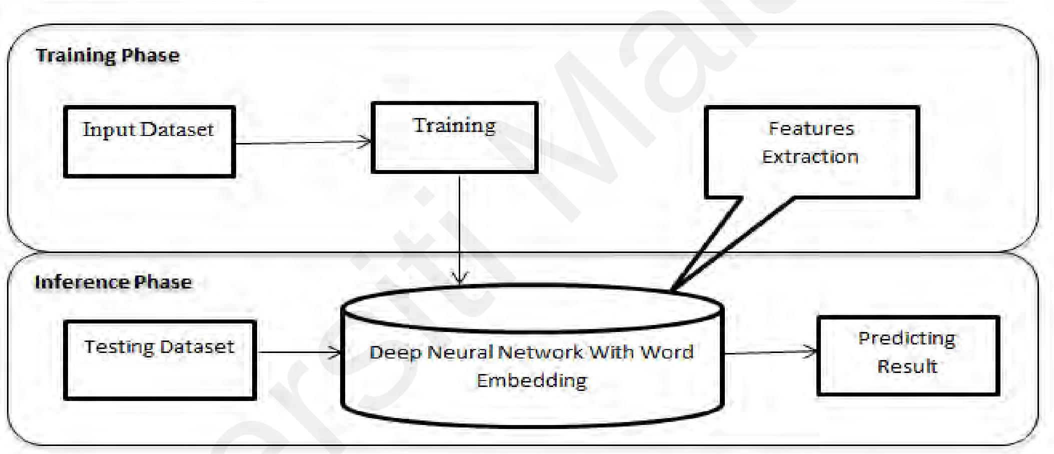


Figure 3.2: The Framework for Airline Services Sentiment Classification using the proposed technique

Through the training process, a large labeled dataset will be gathered and network architecture will be designed to learn the features and model. The deep Neural Network model works by extracting features directly from text dataset using embedding hidden layers, each layer of the data is assigned some random weights and the classifier runs a forward pass through the data, thus predicting the class labels and scores using those weights. The relevant features are not pre-trained, they are learned while the network trains on a collection of the text dataset. The class scores are then compared against the actual labels and an error is computed via a loss function. One complete pass through all of the training samples is called an epoch. After the training is completed, an inference phase commences, this is the stage in which a trained (DLWE) model is used to infer/predict the testing samples and it comprises of a similar forward pass as training to predict the results. Figure 3.2 depicts the framework for airline services sentiment classification using the proposed technique.

3.4.1 Structure of Deep Neural Network with Word Embedding

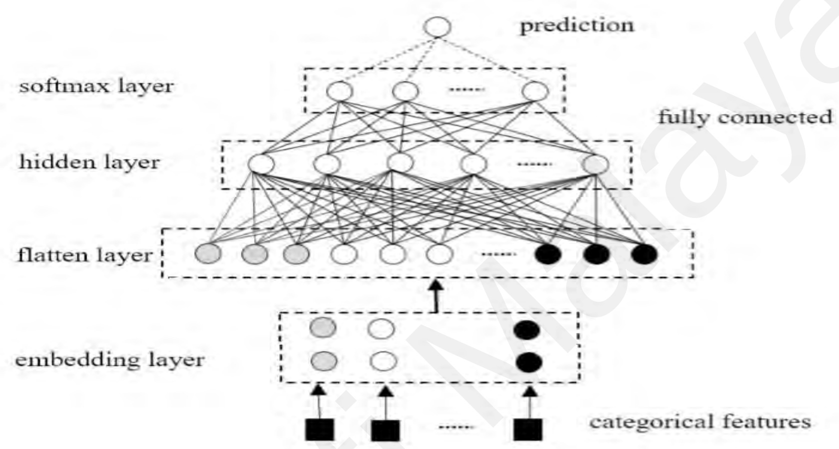


Figure 3.3: Structure of Deep Learning with Word Embedding (Deng et al., 2018)

There are four layers in Deep Learning with Word Embedding model structure, the first layer is the embedding layer, which is an input that is at the bottom of the basic structure, which is fed with sequences of categorical features. Also, there is a set of an embedding layer to turn indexes of categorical features into dense vectors of fixed sizes. The dense vectors will be continuously trained in end-to-end supervised learning, so that the vectors that share common contexts in the corpus will be finally located near one another in the vector space.

The second layer is the flattened layer, which will combine the components of the two-dimensional vector matrix produced by the embedding layer into one dimensional vector in order that will be taken as an input tensor by the next layer. The third layer is fully connected, as it enhances the ability of mining hidden nonlinear characteristics for the model. The last layer is the softmax layer, which plays the role of a multi-category output layer to predict the training labels. The number of nodes in this layer will be set the same as the predication list. Figure 3.3 depicts the structure of deep Learning with word embedding.

3.4.2 Deep Learning Classifier

In this research, the Deep Learning classifier is selected based on the recommendation from the existing literature for classifying text. The deep learning technique is DNN with word Embedding. This classifier was trained and tested to evaluate the performance of the classifier.

Deep Neural Network with Word Embedding is the Deep Learning Technique selected in this research due to several reasons. Firstly, it is a suitable technique for text classification and is one of the most popular deep learning techniques for Sentiment Analysis, which is gradually becoming widespread (Araque et al., 2017). According to Lu et al., (2017), the technique offers an extraction with well-known representation capabilities and a well-designed presentation. This study is a literature classifier to classify the selected text based on the recommendations contained in the classifier of the in-depth study.

These classifiers are taught and tested to evaluate the performance of the classifier. In this study, a deep learning was selected for several reasons. First of all, deep learning algorithms learn the high level features of data (Guo, 2017). This is a better way to learn more about the key stages of the study and the conventionally known machine learning. Thus, deep learning decreases the function of developing a fresh feature to solve the problem (Guo, 2017). Generic word vectors can be captured by word embedding methods such as word2vec (Mikolov et al., 2013) and GloVe (Pennington, Socher, & Man- ning, 2014).

General vectors are obtained in an uncontrolled manner, that is, are not trained for a specific task. The vectors of this word contain semantic and syntactic information, though without specific emotions. However, except the information is available in these general word vectors, this work developed a "sentiment analyzer" model based on the technique of deep text insertion for extraction of properties, called "Sentiment Analyze." (Araque et al., 2017).

3.5 Evaluation Methods

The general aim of the evaluation process is to measure the predictive performance of the proposed model. There are two main objectives for performing the evaluation, first is to measure the performance of the proposed model in predicting the sentiment of airline services and compare the performance of the selected classifier. The second objective is to depict the actual performance of the developed model as well as to understand the overall accuracy of the performance.

There are two evaluation methods that have been used in this research: the classification accuracy method along with the Precision and Recall measures; as they are commonly used in many machine learning researches (Fawcett, 2006; Ross et al., 2009). Apart from that, the confusion matrix is being used to show the performance of the proposed model.

3.5.1 Classification Accuracy

There are numerous means of assessing categorization and prediction of model performance. A common method for assessing the ML / DL performance for a multi-class classification problem is the Precision and Recall performance measures.

Precision and recall are mainly applied in assessing the classification model (Ikonomakis, Kotsiantis, & Tampakas, 2005). The Precision is the fraction of the properly predicted event from the overall prediction. In brief, precision implies the fittingly made prediction. Conversely, recall is the outcome as the number of events predicted by the ML techniques.

1) **Precision** = True Positive / True Positive + False Positive
2) **Recall** = True Positive / True Positive + False Negative

The classification accuracy calculation is denoted as follow:

$\boldsymbol{Classification\ Accuracy}$ = The $number\ of\ correct\ predictions\ made$ / Total number of predictions made $*$ 100

3.5.2 Confusion Matrix

Confusion matrix, which is also referred to as a contingency table, depicts the actual performance of a classification model for each class (Fawcett, 2006). For example, assuming that there are two classes X and Y, and the general accuracy was 96%. It does not mean that the prediction rate is 96% for all the classes, as the classifier may have a good prediction rate on class X and may have 0% prediction rate for class Y.

CHAPTER 4: DEVELOPMENT OF SENTIMENT CLASSIFICATION PREDICTION MODEL

This chapter provides the details of the experiments conducted in this study as well as the parameter specifications of the methods considered in this study and other details related to their implementation. Since this chapter is a direct reflection of the methodology chapter, it has been structured in a similar way to Chapter 3 but with more focus on the implementation side. In addition, all the experimental works are explained in a way that can be replicated by any researcher, in particular, to make the task easier for new researchers or those who aim at repeating our experiment.

4.1 Datasets Collection and Analysis

The dataset here is obtained from Kaggle (https://www.kaggle.com/) and it is a slightly formatted version of the original data which was obtained from Crowdflower's (Wan & Gao, 2016), as it suits the prerequisite of this research and it's efficient to use. Because this dataset does not require further labelling of the sentiment, it has already been labelled manually to state the actual sentiment of the tweet, along with the reason of the sentiment. Additionally, the dataset selected contains full content of the tweet where the Twitter Search API can only obtain a limited content of the tweet. It is also the most up-to-date dataset that contains only airline service relevant tweets. The dataset obtained contains 14,640 labelled tweets with the tweet created date, username, content, sentiment confidence and labelled class, etc. The provided dataset from Kaggle specified that the sentiment confidence is the probability of the class being labelled. This feature contains numeric values in range zero to one with three decimal numbers, of which one indicates that the categorized sentiment is fixed and zero, indicated that it is not fixed. In this research, the content of each tweet is considered as the determining factor for the result of the tweet sentiment. The tweets collected in each dataset are labelled in three classes, positive (+ve), negative (-ve) and neutral (n). The number of tweets distribution for individual class is shown in Figure 4.1 below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	tweet_id	airline_se	airline_se	negativer	negativer	airline	airline_se	name	negativer	retweet_c	text	tweet_coo	tweet_cre	tweet_loc	user_timezone				
2	5.7E+17	neutral	1			Virgin America		cairdin		0	@VirginAmerica Wh		#########		Eastern Time (US & Canada)				
3	5.7E+17	positive	0.3486		0	Virgin America		jnardino		0	@VirginAmerica plu		#########		Pacific Time (US & Canada)				
4	5.7E+17	neutral	0.6837			Virgin America		yvonnalynn		0	@VirginAmerica I di		#########	Lets Play	Central Time (US & Canada)				
5	5.7E+17	negative	1	Bad Flight	0.7033	Virgin America		jnardino		0	@VirginAmerica it's		#########		Pacific Time (US & Canada)				
6	5.7E+17	negative	1	Can't Tell	1	Virgin America		jnardino		0	@VirginAmerica and		#########		Pacific Time (US & Canada)				
7	5.7E+17	negative	1	Can't Tell	0.6842	Virgin America		jnardino		0	@VirginA		#########		Pacific Time (US & Canada)				
8	5.7E+17	positive	0.6745		0	Virgin America		cjmcginnis		0	@VirginAmerica yes		#########	San Franci	Pacific Time (US & Canada)				
9	5.7E+17	neutral	0.634			Virgin America		pilot		0	@VirginAmerica Rea		#########	Los Angel	Pacific Time (US & Canada)				
10	5.7E+17	positive	0.6559			Virgin America		dhepburn		0	@virginamerica Wel		#########	San Diego	Pacific Time (US & Canada)				
11	5.7E+17	positive	1			Virgin America		YupitsTate		0	@VirginAmerica it w		#########	Los Angel	Eastern Time (US & Canada)				
12	5.7E+17	neutral	0.6769		0	Virgin America		idk_but_youtube		0	@VirginAmerica did		#########	1/1 loner :	Eastern Time (US & Canada)				
13	5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica I &l		#########	NYC	America/New_York				
14	5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica Thi:		#########	NYC	America/New_York				
15	5.7E+17	positive	0.6451			Virgin America		mollanderson		0	@VirginAmerica @v		#########		Eastern Time (US & Canada)				
16	5.7E+17	positive	1			Virgin America		sjespers		0	@VirginAmerica Tha		#########	San Franci	Pacific Time (US & Canada)				
17	5.7E+17	negative	0.6842	Late Flight	0.3684	Virgin America		smartwatermelon		0	@VirginAmerica SFC		#########	palo alto,	Pacific Time (US & Canada)				
18	5.7E+17	positive	1			Virgin America		ItzBrianHunty		0	@VirginAmerica So		#########	west covi	Pacific Time (US & Canada)				
19	5.7E+17	negative	1	Bad Flight	1	Virgin America		heatherovieda		0	@VirginAmerica I fl		#########	this place	Eastern Time (US & Canada)				
20	5.7E+17	positive	1			Virgin America		thebrandiray		0	I â¤ï¸ flying @VirginA		#########	Somewhe	Atlantic Time (Canada)				
21	5.7E+17	positive	1			Virgin America		JNLpierce		0	@VirginAmerica you		#########	Boston | V	Quito				
22	5.7E+17	negative	0.6705	Can't Tell	0.3614	Virgin America		MISSGJ		0	@VirginAmerica wh		#########						
23	5.7E+17	positive	1			Virgin America		DT_Les		0	@VirginA	[40.74804	#########						
24	5.7E+17	positive	1			Virgin America		ElvinaBeck		0	@VirginAmerica I lo		#########	Los Angel	Pacific Time (US & Canada)				
25	5.7E+17	neutral	1			Virgin America		rjlynch21086		0	@VirginAmerica will		#########	Boston, M	Eastern Time (US & Canada)				

Tweets

Figure 4.1: Part of data sample containing 14,640 number of labelled tweets

- **Positive sentiment**

Positive tweets contain positive words such as "pretty", "great" , "fabulous" and so on. Figure 4.2 shows the Positive labeled tweets.

@VirginAmerica I <3 pretty graphics. so much better than minimal iconography. :D
@VirginAmerica This is such a great deal! Already thinking about my 2nd trip to @Australia & I haven't even
@VirginAmerica @virginmedia I'm flying your #fabulous #Seductive skies again! U take all the #stress away from
@VirginAmerica Thanks!

Figure 4. 2: Positive labeled tweets

- **Negative sentiment**

The negative tweets contain negative words such as "late flight", "no interesting", "cancelled", "disappointing", and so on. Figure 4.3 show the Negative labeled tweets.

@VirginAmerica you're the best!! Whenever I (begrudgingly) use any other airline I'm delayed and Late Flight :(
@VirginAmerica I have no interesting flying with you after this. I will Cancelled Flight my next four flights I planned.#neverflyvirgir
@VirginAmerica it was a disappointing experience which will be shared with every business traveler I meet. #neverflyvirgin

Figure 4.3: Negative labeled tweets

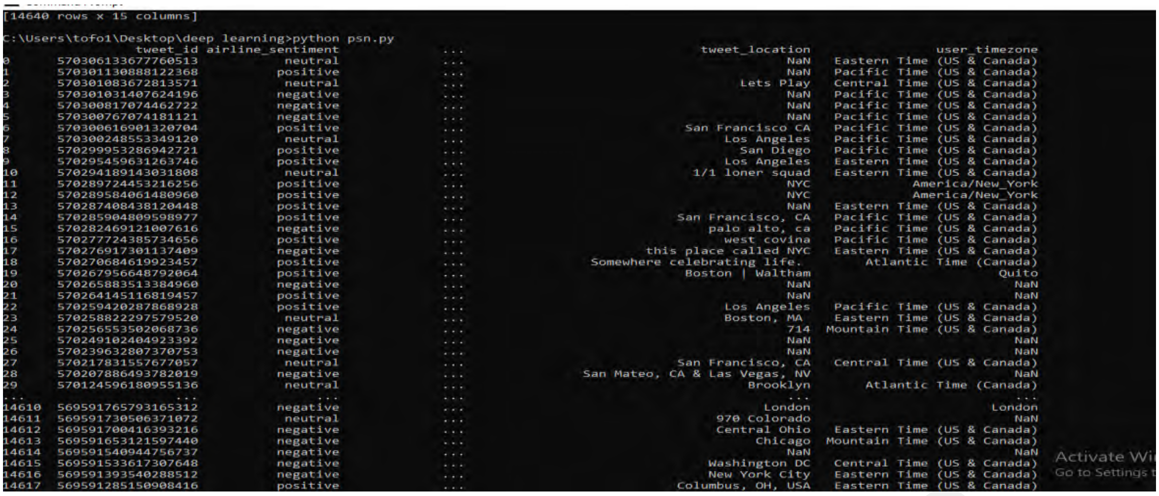
- **Neutral sentiment**

The neutral tweets seem to focus on asking airlines for information Figure 4.4 shows Neutral labeled tweets.

@SouthwestAir I've been a fan of imagine dragons since 2012 and they're my fave band and #DestinationDragons is during my bday	#####
@SouthwestAir Whats up with flight 4464 Ft myers to milwaukee???	#####
@SouthwestAir My birthday is during #DestinationDragons could I get tix?? http://t.co/07XhCacjax	#####
@SouthwestAir Please bring back RDU to FLL direct route! I noticed it is missing starting in Late Flight August! The flights are alway	#####

Figure 4.4: Neutral labeled tweets

In the collected dataset from Kaggle, the numbers of airline service providers obtained are Virgin America Airline, United Airline, Southwest Airline, Delta Airline, US Airways and American Airline. In the process of simplifying the computation, the dataset is downloaded and read into Pandas DataFrame. Some exploratory data analysis was conducted on the dataset prior to the training of the data in the deep learning models. This is to obtain an insightful knowledge of the process. The following Figure 4.5 and Figure 4.6 show the visual representation of the data.



```
[14640 rows x 15 columns]

C:\Users\tofo1\Desktop\deep learning>python psn.py
                tweet_id airline_sentiment  ...               tweet_location               user_timezone
0      570306133677760513           neutral  ...                          NaN  Eastern Time (US & Canada)
1      570301130888122368          positive  ...                          NaN  Pacific Time (US & Canada)
2      570301083672813571           neutral  ...                    Lets Play  Central Time (US & Canada)
3      570301031407624196          negative  ...                          NaN  Pacific Time (US & Canada)
4      570300817074462722          negative  ...                          NaN  Pacific Time (US & Canada)
5      570300767074181121          negative  ...                          NaN  Pacific Time (US & Canada)
6      570300616901320704          positive  ...             San Francisco CA  Pacific Time (US & Canada)
7      570300248553349120           neutral  ...                  Los Angeles  Pacific Time (US & Canada)
8      570299953286942721          positive  ...                    San Diego  Pacific Time (US & Canada)
9      570295459631263746          positive  ...                  Los Angeles  Eastern Time (US & Canada)
10     570294189143031808           neutral  ...             1/1 loner squad  Eastern Time (US & Canada)
11     570289724453216256          positive  ...                          NYC            America/New_York
12     570289584061480960          positive  ...                          NYC            America/New_York
13     570287408438120448          positive  ...                          NaN  Eastern Time (US & Canada)
14     570285904809598977          positive  ...            San Francisco, CA  Pacific Time (US & Canada)
15     570282469121007616          negative  ...                palo alto, ca  Pacific Time (US & Canada)
16     570277724385734656          positive  ...                  west covina  Pacific Time (US & Canada)
17     570276917301137409          negative  ...       this place called NYC  Eastern Time (US & Canada)
18     570270684619923457          positive  ...  Somewhere celebrating life.      Atlantic Time (Canada)
19     570267956648792064          positive  ...            Boston | Waltham                       Quito
20     570265883513384960          negative  ...                          NaN                         NaN
21     570264145116819457          positive  ...                          NaN                         NaN
22     570259420287868928          positive  ...                  Los Angeles  Pacific Time (US & Canada)
23     570258822297579520           neutral  ...                 Boston, MA   Eastern Time (US & Canada)
24     570256553502068736          negative  ...                          714 Mountain Time (US & Canada)
25     570249102404923392          negative  ...                          NaN                         NaN
26     570239632807370753          negative  ...                          NaN                         NaN
27     570217831557677057           neutral  ...            San Francisco, CA  Central Time (US & Canada)
28     570207886493782019          negative  ... San Mateo, CA & Las Vegas, NV                         NaN
29     570124596180955136           neutral  ...                     Brooklyn      Atlantic Time (Canada)
...                   ...               ...  ...                          ...                         ...
14610  569591765793165312          negative  ...                       London                      London
14611  569591730506371072           neutral  ...                 970 Colorado                         NaN
14612  569591700416393216          negative  ...                 Central Ohio  Eastern Time (US & Canada)
14613  569591653121597440          negative  ...                      Chicago Mountain Time (US & Canada)
14614  569591540944756737          negative  ...                          NaN                         NaN
14615  569591533617307648          negative  ...                Washington DC  Central Time (US & Canada)
14616  569591393540288512          negative  ...                New York City  Eastern Time (US & Canada)
14617  569591285150908416          positive  ...           Columbus, OH, USA  Eastern Time (US & Canada)
```

Figure 4.5: Screenshot of Visual representation of the data

```
[14640 rows x 15 columns]
negative    9178
neutral     3099
positive    2363
Name: airline_sentiment, dtype: int64
```

Figure 4.6: Screenshot of Label distribution of Tweets

Table 4.1: Label distribution of Tweets

	Negative	Positive	Neutral
Number of Tweets	9178	2363	3099

The analysis also indicated a greater quantity of negative occasions for each of the 6 airlines (Virgin America, American, Delta, US Airways, Southwest, and United) compared to positive and neutral that can also cause for situations to lean towards being grouped as negative in cases that contained the name or Twitter handle of the airline come-up in the text. Figure 4.7 shows the total number of tweets for each airline. For US Airways, United Airlines and American Airlines, there are 4 - 5 times additional Twitter posts grouped as negative more than positive and this can affect the training with regards to how the classifiers regulate the sentiments of the test data or future posts. Figure 4.8 depict the airline sentiment distribution.

```
Total number of tweets for each airline
  airline
United            3822
US Airways        2913
American          2759
Southwest         2420
Delta             2222
Virgin America     504
Name: airline_sentiment, dtype: int64
```

Figure 4.7: Total Number of Tweets for each Airline

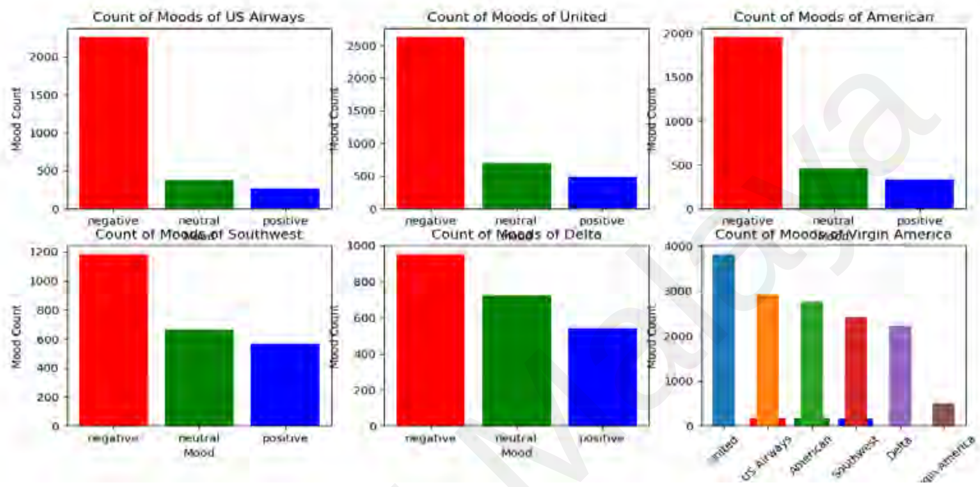


Figure 4.8: Airline Sentiment Distribution

4.2 System Requirements and Tools

The airline sentiment prediction model was implemented and evaluated using Python programming language code with Sublime Text operated and Command Prompt on Windows 10 operating system, and Intel® Core™ i5-5200U CPU and 2.20GHz with Random-access memory of 6.00 GB System Type: 64-bit Operating System, x64-based processor. Implementing the development of the model using python sublime Text consists of three main steps, which are data pre-processing, classification, and evaluation. The implementation of the system begins with the installation of Keras packages and Tensorflow library, which was used for data processing, extraction of the features, training, and evaluation of the classifier. Keras packages and TensorFlow library were also used for conducting the evaluation of the proposed model.

4.3 Data Preparation

4.3.1 Reading and cleaning data

The csv file was loaded on to the tweets and the entries were shuffled at random. This practice is to have an even distribution of sentiment classes over train and test sets to be free of bias. Here, the airline sentiment column is represented as the target and the text column represents the input.

Next, we remove stopwords, as they do not have any contribution in the sentiment prediction. Therefore, we removed the stopwords since we will build a general model to be used for other airlines, as shown in Figure 4.9 below.

```
    tweets_dir = "twitter-airline-sentiment/"
df = pd.read_csv('Tweets.csv')
df = df.reindex(np.random.permutation(df.index))
df = df[['text', 'airline_sentiment']]
df.text = df.text.apply(remove_stopwords).apply(remove_mentions)
```

Figure 4.9: Screenshot of the Reading and Cleaning Data Code

4.3.2 Train-Test split

Exemplary performance evaluation should be performed using a separate test. How good is the overallization of the model? This is done by scikit-learning's train_test_split function in Python to divide the dataset automatically as shown in Figure 4.10.

```
X_train, X_test, y_train, y_test = train_test_split(df.text, df.airline_sentiment, test_size=0.1, ra
print('# Train data samples:', X_train.shape[0])
print('# Test data samples:', X_test.shape[0])
assert X_train.shape[0] == y_train.shape[0]
assert X_test.shape[0] == y_test.shape[0]
```

```
# Train data samples: 13176
# Test data samples: 1464
```

Figure 4.10: Screenshot of Train-Test split Code and Output

4.3.3 Converting words to numbers

Translating tweets to targets refers to translating these words into integers that point to an index and store the most commonly used words in the dictionary. Then place the text by applying filters and putting words in lowercase letters and split them into spaces as shown in Figure 4.11.

```
tk = Tokenizer(num_words=NB_WORDS,
               filters='!"#$%&()*+,-./:;<=>?@[\\]^_`{|}~\t\n',
               lower=True,
               split=" ")
tk.fit_on_texts(X_train)
print('Fitted tokenizer on {} documents'.format(tk.document_count))
print('{} words in dictionary'.format(tk.num_words))
print('Top 5 most common words are:', collections.Counter(tk.word_counts).most_common(5))
X_train_seq = tk.texts_to_sequences(X_train)
X_test_seq = tk.texts_to_sequences(X_test)
```

```
Fitted tokenizer on 13176 documents
10000 words in dictionary
Top 5 most common words are: [('flight', 3540), ('not', 1406), ('no', 1329), ('get', 1215), ('t', 1095)]
```

Figure 4.11: Screenshot of Converting Words to Numbers Code and Output

4.3.4 Creating word sequences of equal length

The sequences have different lengths and Keras prefers inputs to be vectorized and before the word embedding is computed, it needs to make sure the sequences are of equal length. We will pad all input sequences to have the length. Again, we can do this with a built in Keras function, in this case the pad_sequences() function, By stating maxlen, the truncated sequences or lengthened with zeros as shown in Figure 4.12.

```
seq_lengths = X_train.apply(lambda x: len(x.split(' ')))
seq_lengths.describe()
```

```
X_train_seq_trunc = pad_sequences(X_train_seq, maxlen=MAX_LEN)
X_test_seq_trunc = pad_sequences(X_test_seq, maxlen=MAX_LEN)
X_train_seq_trunc[10]  # Example of padded sequence
```

Figure 4.12: Screenshot of creating word sequences of equal length Code

4.3.5 Encoding the target variable

The objective classes are strings that need to be transformed into number vectors. This is obtained with the LabelEncoder from Sklearn and the to_categorical method using Keras. Figure 4.13 shows screenshot of encoding the target variable code.

```
le = LabelEncoder()
y_train_le = le.fit_transform(y_train)
y_test_le = le.transform(y_test)
y_train_oh = to_categorical(y_train_le)
y_test_oh = to_categorical(y_test_le)
```

Figure 4.13: Screenshot of Encoding the Target Variable Code

4.3.6 Splitting off validation data

Here the splitting is completed for the validation set and the data is ready. The model performance will be validated based on this validation set with the tuned parameters of the model. Figure 4.14 shows screenshot of splitting off validation data code and output.

```
X_train_emb, X_valid_emb, y_train_emb, y_valid_emb = train_test_split(X_train_seq_trunc, y_train_oh, test_s

assert X_valid_emb.shape[0] == y_valid_emb.shape[0]
assert X_train_emb.shape[0] == y_train_emb.shape[0]

print('Shape of validation set:',X_valid_emb.shape)
```

```
Shape of validation set: (1318, 24)
```

Figure 4.14: Screenshot of Splitting off Validation Data Code and Output

4.4 Classification Model based on Deep Learning

One of the research objectives of this study is to identify suitable classification technique(s) that have the ability to predict the sentiment of airline services. The classification technique used in this research is (DLWE) Deep learning with word embedding. The classifier was trained and tested, where the word embedding and embedding layer is applied as an automatic feature extractor for the deep learning classifier. These embedding features are shown to be comparable and sometimes better than ngram features in several tasks, especially if classes are semantically separated.

4.4.1 Building Deep learning Classification Model with Word Embedding

In the Deep learning model, there is contained an embedded layer of 15 tightly linked layers. The input_shape for the initial layer is equal to the number of words kept in the dictionary which was created from one-hot-encoded features. As required to predict 3 different sentiment classes, the last layer has three elements. The softmax activation feature assures that the three probabilities are totaled to 1. The Embedding needs the vocabulary size (VS) specification, the size of the real-valued vector space, and the highest length of input files. The VS is the sum of words in the vocabulary, together with the number of unknown words. This could be the vocabulary set length, referred to as vocab set within the tokenizer for encoding the documents files.

Embedding layer which is provided by Keras is defined as part of the proposed neural network model as it helps to train the specific word embedding based training data. The Embedding contained a vocabulary of the size 10,000. An input length of 24 and 8 dimensions will then be used as a small embedding space.Figure 4.15 shows deep learning classification code.

```python
emb_model = models.Sequential()
emb_model.add(layers.Embedding(NB_WORDS, 8, input_length=MAX_LEN))
emb_model.add(layers.Flatten())
emb_model.add(layers.Dense(3, activation='softmax'))
emb_model.summary()
```

Figure 4.15: Screenshot of Deep learning Classification Code

After defining the Keras, a summation of the layers is published to check for validation against the expected output of the Embedding layer. This proved to be 24×8 matrix and is compressed to a 192-element vector by the Flatten layer. The quantity of parameters to train is computed as (nb inputs x nb elements in hidden layer) + nb bias terms. The number of inputs for the first layer equals the number of words in the corpus. The next layers contained the number of outputs of the earlier layer as inputs. By implication this is presented as:

Initial layer : (10000 x 8) + 0 = 80000

Middle layer : (80000 x 0) + 0 = 0

Final layer : (192 x 3) + 3 = 579

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 24, 8)             80000
_________________________________________________________________
flatten_1 (Flatten)          (None, 192)               0
_________________________________________________________________
dense_1 (Dense)              (None, 3)                 579
=================================================================
Total params: 80,579
Trainable params: 80,579
Non-trainable params: 0
_________________________________________________________________
```

Figure 4.16: Screenshot of Deep learning with Word Embedding Classification Output

Categorical_crossentropy as the loss function and softmax as the final activation function is utilized to fit the model on the train data and then validated on the validation set.Figure 4.16 shows deep learning with word embedding model output.

CHAPTER 5: EVALUATION, RESULTS AND DISCUSSION

5.1 Chapter overview

This chapter discusses the development processes, which includes the implementation of the proposed prediction model as well as the evaluation of the proposed model.

5.2 Classification Accuracy

The development process was conducted using Keras packages and Keras provides an embedding layer which helps us to train specific word embedding based on our training data. The deep learning with word embedding (DLWE) classifier was trained and tested separately. Scikit-learn was used to separate our dataset to a training set and test set. The classifier was trained on 90% (13176 tweets) of the data and tested on 10% (1464 tweets) A large split was choosing is because of it is a noisy problem and a well-performing model requires as much data as possible to learn the complex classification function. The accuracy of the trained model and its performance on classifying and predicting the airline sentiment classification are discussed in the next sub-section. Figure 5.1 show training and testing of the deep Learning classifier code and output.

```
emb_history = deep_model(emb_model, X_train_emb, y_train_emb, X_valid_emb, y_valid_emb)
```

Figure 5.1: Screenshot of Training and Testing of the Deep Learning Classifier Code

The validation accuracy was around 81%. This indicates the number of words in the tweets is slightly low, proving the result to be better. The model starts over fit from epoch 6 based on the comparison of the training and validation accuracy with the loss as shown in Figure 5.2 and Figure 5.3.

```
eval_metric(emb_history, 'acc')
```

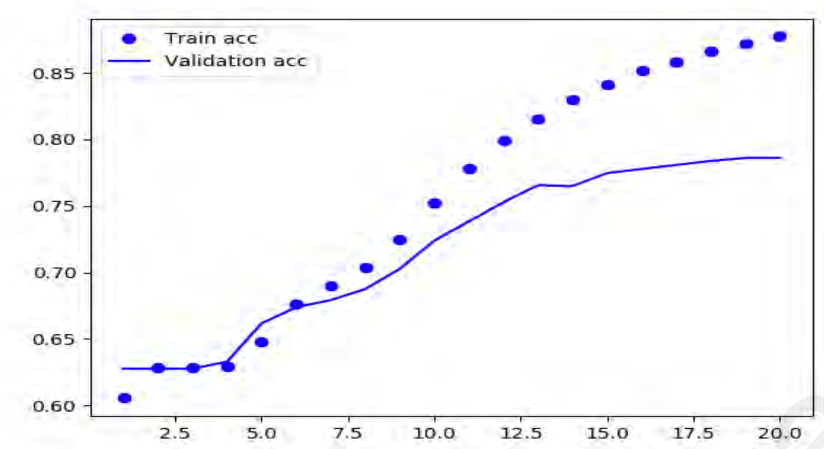


Figure 5.2: Screenshot of Validation Accuracy Output


```
eval_metric(emb_history, 'loss')
```

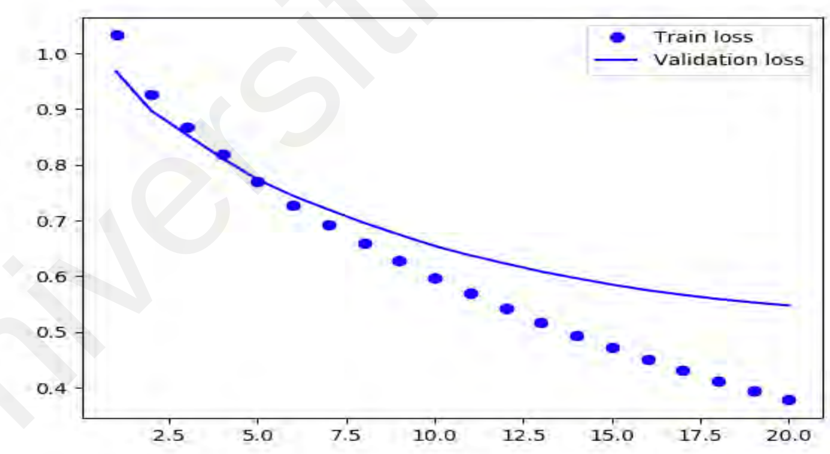
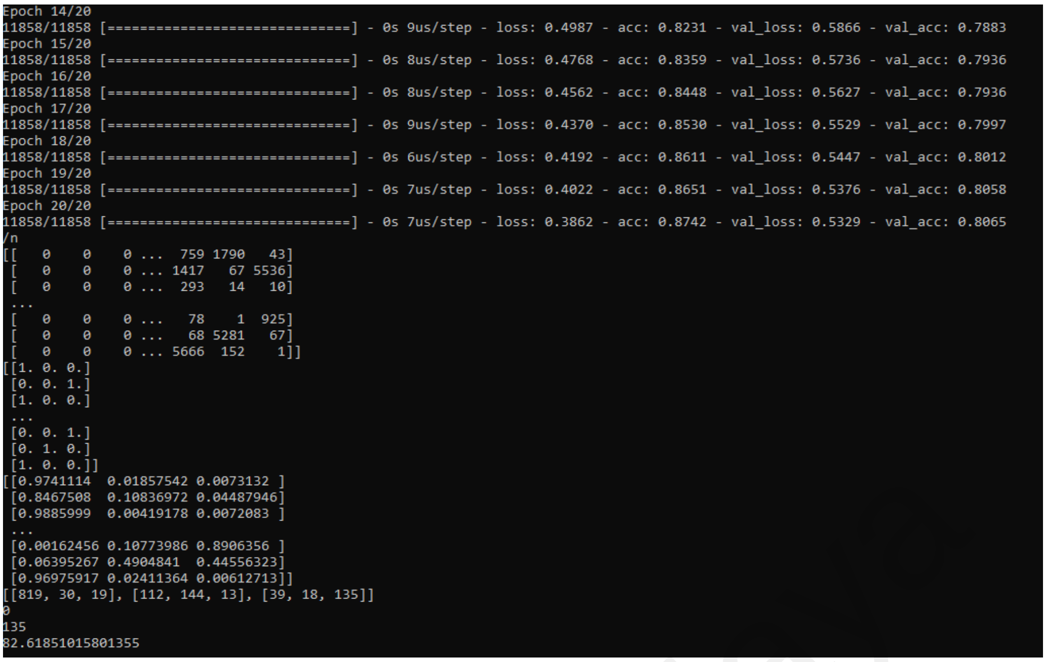


Figure 5.3: Screenshot of Loss Validation Output

5.2.1 Test Result

Figure 5.4 below shows the prediction accuracy of the selected classifier. Deep Learning With Word Embedding (DLWE) achieved the high prediction accuracy with the average accuracy of 82.6%.



```
Epoch 14/20
11858/11858 [==============================] - 0s 9us/step - loss: 0.4987 - acc: 0.8231 - val_loss: 0.5866 - val_acc: 0.7883
Epoch 15/20
11858/11858 [==============================] - 0s 8us/step - loss: 0.4768 - acc: 0.8359 - val_loss: 0.5736 - val_acc: 0.7936
Epoch 16/20
11858/11858 [==============================] - 0s 8us/step - loss: 0.4562 - acc: 0.8448 - val_loss: 0.5627 - val_acc: 0.7936
Epoch 17/20
11858/11858 [==============================] - 0s 9us/step - loss: 0.4370 - acc: 0.8530 - val_loss: 0.5529 - val_acc: 0.7997
Epoch 18/20
11858/11858 [==============================] - 0s 6us/step - loss: 0.4192 - acc: 0.8611 - val_loss: 0.5447 - val_acc: 0.8012
Epoch 19/20
11858/11858 [==============================] - 0s 7us/step - loss: 0.4022 - acc: 0.8651 - val_loss: 0.5376 - val_acc: 0.8058
Epoch 20/20
11858/11858 [==============================] - 0s 7us/step - loss: 0.3862 - acc: 0.8742 - val_loss: 0.5329 - val_acc: 0.8065
/n
[[   0    0    0 ...  759 1790   43]
 [   0    0    0 ... 1417   67 5536]
 [   0    0    0 ...  293   14   10]
 ...
 [   0    0    0 ...   78    1  925]
 [   0    0    0 ...   68 5281   67]
 [   0    0    0 ... 5666  152    1]]
[[1. 0. 0.]
 [0. 0. 1.]
 [1. 0. 0.]
 ...
 [0. 0. 1.]
 [0. 1. 0.]
 [1. 0. 0.]]
[[0.9741114  0.01857542 0.0073132 ]
 [0.8467508  0.10836972 0.04487946]
 [0.9885999  0.00419178 0.0072083 ]
 ...
 [0.00162456 0.10773986 0.8906356 ]
 [0.06395267 0.4904841  0.44556323]
 [0.96975917 0.02411364 0.00612713]]
[[819, 30, 19], [112, 144, 13], [39, 18, 135]]
0
135
82.61851015801355
```

Figure 5.4: Screenshot of Accuracy Test Resul

Table 5.1: Accuracy Result

Deep learning classification		
Word Embedding Model	Embedding layer	82.6%
	Dense layer	
	Flatten layer	

5.3 Confusion Matrix

There are many standard performance measures that can be used to evaluate the prediction techniques, such as classification accuracy for evaluating the machine learning classifier. The confusion matrix was used in this research to present the prediction accuracy of the classifiers. The confusion matrix is useful to show both the actual class distribution and the classifiers predicted class distribution.

For evaluating the performance of the classifiers, a set of experiments were carried out using the collected dataset as described in this chapter 4. In these experiments, 90% of the dataset was set for the training and 10% for the testing because deep learning required large amount of data to train the model. The evaluation was conducted for the deep learning classifier. Table 5.2 shows the confusion Matrix for deep learning. Based on Table 4.3, 819 negative data out of 868 negative samples were correctly predicted as True Negative, with a precision rate of 84%, recall of 94% and F-score of 89%. For the Neutral data, 144 were predicted correctly as True Neutral from 269 neutral tested data sample with a precision rate of 75%, recall of 54% and F-score of 62%. Last and not the least, the positive sample correctly predicted 135 out of 192 as True Positive with a precision rate of 80%, recall of 70% and F-score of 75%. Deep learning with word embedding performs better than other classifiers in previous study with 82.61% accuracy, with the highest precision rate 86 % and recall percentage of 82%.

Table 5.2: Confusion Matrix of Deep Learning with word embedding

Classifier	Deep Learning with word embedding								
Actual	Negative	Neutral	Positive	Invalid prediction	Precision	Recall	F1-Score	Support	
Negative	819	30	19	-	0.84	0.94	0.89	868	
Neutral	112	144	13	-	0.75	0.54	0.62	269	
Positive	39	18	135	-	0.80	0.70	0.75	192	
Invalid prediction	-	-	-	135	-	-	-	-	
Predict	970	192	167	Average / Total	0.86	0.82	0.84	1.329	

5.4 Discussion

Based on the evaluation results, the proposed prediction system that uses the selected classifier, which is Deep learning with word embedding DLWE, performed much better than the existing machine learning classifiers (refer to chapter 2) in sentiment prediction classification from the text input. It should be noted that the dataset used for testing of the existing systems - results were reported in Chapter 2 (pg. 29-31) - exist in the table of accuracy and confusion matrix, which was used for testing the proposed prediction model. Table 5.3 shows the benchmark comparison between the proposed DLWE technique VS the existing machine learning classification techniques. The average performance refers to accuracy of sentiment classification of machine learning benchmark work in comparison to the deep learning with word embedding of the current work.

Table 5.3: Benchmark Comparison between the Proposed DLWE Technique VS the Existing ML classification Techniques

Model	**Deep Learning with Word Embedding (DLWE)**	**SVM Classifier with Word2Vec Model**	**Logistic Regression with Word2Vec Model**
Average Performance	82.61%	72%	72%

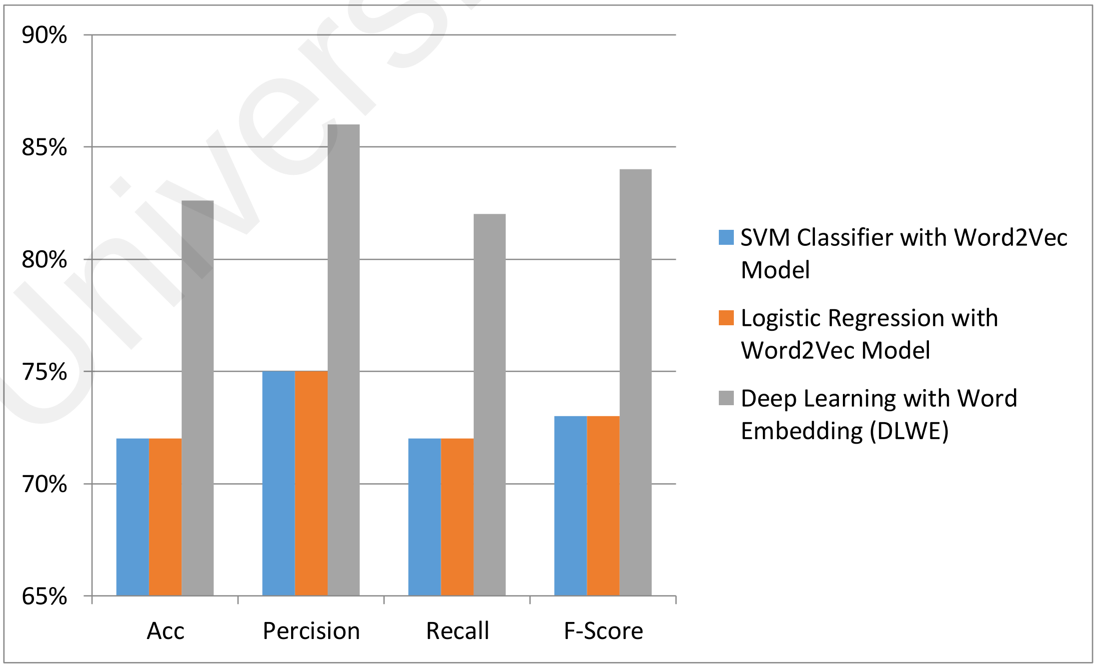


Figure 5.5: Diagram of Classifications Comparative

From Table 5.3 and Figure 5.5, it is clear that the proposed prediction system using Deep Learning with Word Embedding can predict the sentiment of airline services with an average accuracy of 82.61%. On the other hand, the performance of the existing airline services sentiment system, the SVM and Logistic Regression was 72%. It can also be observed that the proposed model performed much better than the existing systems with an improved accuracy of more than 10%. This is because it used deep learning with word embedding that have layers. The current research employed the technique of deep learning with the word embedding feature. Word embedding provides an automatic extraction of features and has richer capabilities as well as outperforms very well, with regards to representations, it has proven to be better than the existing feature based techniques, such as machine learning (Lu et al., 2017; Araque et al., 2017; Chatterjee et al., 2018). This is because those machine learning approaches in existence perform on the basis of features that undergo manual extraction, which firstly makes them complex to use, and secondly questions the process of extraction, especially in methods that are feature-driven.

CHAPTER 6: CONCLUSION AND FUTURE RESEARCH

6.1 Overview

This chapter is the final part of the dissertation. The major findings of the study and how the objectives were met is presented in this Chapter. The research contributions, limitations, and the future work are also presented here.

6.2 Meeting of Research Aims and Objectives

The primary goal of this research is to propose a predictive model based on deep learning technique to identify the customers' sentiment polarity in airline services and improve the accuracy of the existing predicting model. To achieve the aim of this study, the following objectives are highlighted:

6.2.1 Fulfilling Objective 1

The first objective is to review and analyze the limitations of the existing techniques in predicting sentiment classification of Airline Services. Some of such limitations include a lack of direct comparison with some other classifiers which have the incremental abilities, some of which are the Incremental Tree Induction, and is based on fuzzy logic. Also, the other advanced techniques with incremental classifiers could be able to predict and mine the satisfaction of users via online reviews, instead of manual methods as used in this research. Another limitation is that machine learning approaches make frequent use of the Bag of Words (BOW) model, even though on one hand, it might be an efficient and easy approach, yet there is a loss of a reasonable amount of the original natural language obtained information. To meet the objective, review of literature was carried out to acquire information about the current sentiment classification techniques, especially the airline services domain.

6.2.2 Fulfilling Objective 2

The second objective is to suggest a suitable sentiment classification method to classify the general public opinion on airline services from tweets. To achieve this objective, deep learning with word embedding classifier was identified and comparison was made to determine its capability in text categorization in line with how they performed in term of classifying objects. The researcher had to perform several tasks to achieve these objectives, which are as follows:

The initial task is looking for an appropriate deep learning classification framework to be applied for the prediction, thus meeting our objectives.

• The second task is to choose the appropriate twitter dataset matching airline services. To ease the second task, the dataset was previously labeled, dataset does not require further labelling of the sentiment, since it has already been labelled manually to state the actual sentiment of the tweet, which helped in saving valuable time.

• The third task in fulfilling objective 2 is application of the prediction model on the twitter dataset. The procedures applied the Sublime and Python Programming language to develop the deep learning model with embedding layer. Based on this combination, the features automatically from the data were extracted, trained and tested on the proposed prediction model.

In addition, one of the existing sentiment classification models was selected and evaluated to determine its ability in classifying sentiment polarity. This successfully achieved high classification accuracy rates in various classification problems.

6.2.3 Fulfilling Objective 3

The third and final objective of this research was to evaluate the performance of the proposed technique in classifying the opinion mining of airline services.

• Accuracy and Loss validation were applied for validating the dataset and estimating the exactness of the predictive model. Confusion matrix is utilized in the presentation of the predicted accurateness of the selected classifier.

In summary, the research with regards to developing deep learning prediction model for classifying and predicting the sentiment analysis polarity of airline services was successfully conducted. The proposed solution for the prediction has confirmed its efficiency with a better level of accuracy.

6.3 Conclusion

Sentiment classification has been quite studied by the academia and firms. Due to its importance and acceptability in business domains, different procedures have been developed to get as much accuracy in the classifications. Many industries are moving into the big data arena and leveraging technology to get deeper into business potentials. Among the leveraging technology is the sentiment classification. This technology classifies the client's opinion and offers a broad aspect of the client's feedback. Among the platforms leveraging technology to mine clients' opinion, the most popular and accurate is Twitter. The platform has been utilized in recent studies to classify client sentiments.

This dissertation contributed to the classification of sentiments using the deep learning approach with word embedding to improve the sentiment classification performance. Specifically, the embedding layer technique for sentiment classification was used as the features extractor. The deep learning technique was selected due to its effectiveness in classifying and predicting sentiment analysis. The research described in detail the development of the proposed prediction model, including the tools and resources required for accomplishing the development process to achieve desired performance. The process encompasses of a text corpus, data pre-processing, model training and testing. Furthermore, the outcome of the experiment and the analysis for the tweets gathered is vital information pertaining to the improvements of the services provided by airlines. As a conclusion, it is important to state that the correctness of the classifier in various sentiment classes moreover, reveals the clients' behaviors on Twitter.

6.4 Research Outcomes and Contributions

It can be observed that previous studies focused on general classification of Twitter sentiment rather than specific and deeper sentiments. Based on the outcomes, the classification proved that the accurateness of the sentiment classification in this study is higher than the twitter sentiment classification in general.

Few studies were conducted in the domain of twitter sentiment analysis on airline services. Previous studies equate different conventional classification techniques and pick out the most precise technique and implement the sentiment classification (Joshua, 2017). But, the deep learning approach in this work, improves the accuracy by applying deep learning with word embedding sentiment classifier. Again, this research reveals that the class categorization of the sentiments is not balanced and the negative sentiment and neutral sentiment tweets are greater than the positive sentiment tweets. This is an indication that Twitter users find it more favorable to tweet their bad emotions rather than the good ones.

6.5 Limitations and Future Research

The services in the airline firms need further improvement, and using a sample tweet dataset larger than the one used in this research will greatly help in identifying clients need and improve the airline services performance. Combination of different techniques and further developing of new algorithms as well as applications on the sentiment in airlines will reveal more clients' opinion.

REFERENCES

Acosta, J., Lamaute, N., Luo, M., Finkelstein, E., & Cotoranu, A. (2017). Sentiment Analysis of Twitter Messages Using Word2Vec. *Proceedings of Student-Faculty Research Day, CSIS, Pace University*, C8-1-C8-7. Retrieved from https://pdfs.semanticscholar.org/784d/1b2aebda3b80567bf8244e89499c31cf42a9.pdf

Almas, M. (2017). *Sentiment Analysis from Social Data using Deep Learning Autoencoder Technique*. *6*(11), 620–625.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, *77*, 236–246. https://doi.org/10.1016/j.eswa.2017.02.002

Bahrami, M., Findik, Y., Bozkaya, B., & Balcisoy, S. (2018). *Twitter Reveals: Using Twitter Analytics to Predict Public Protests*. 1–24. Retrieved from http://arxiv.org/abs/1805.00358

Bian, J., Gao, B., & Liu, T. Y. (2014). Knowledge-powered deep learning for word embedding. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8724 LNAI*(PART 1), 132–148. https://doi.org/10.1007/978-3-662-44848-9_9

Bott, G. A. (2014). *Implicit Customer Loyalty Factors in the Leisure Airline Business*. *36*(1), 60.

Cao, S., & Lu, W. (2017). Improving Word Embeddings with Convolutional Feature Learning and Subword Information. *Aaai*, *1*, 3144–3151.

Carli, D. (2003). Honk If You Support JDF. *Print on Demand*, *9*(2), 42. https://doi.org/10.1145/2740908.2742760

Chen, M., Weinberger, K., Sha, F., & Bengio, Y. (2014). Marginalized Denoising Auto-encoders for Nonlinear Representations. *International Conference on Machine Learning*, *32*, 1476–1484. https://doi.org/10.1007/s11222-007-9033-z

Darliansyah, A., Wandabwa, H. M., & B, M. A. N. (2020). *Intelligent Systems, Technologies and Applications* (Vol. 910). https://doi.org/10.1007/978-981-13-6095-4

Data, T., Using, A., Framework, H., Science, C., & Gurgaon, M. (2016). *Twitter Data Analysis Using FLUME & HIVE on Hadoop FrameWork*. (February), 119–123.

DEITY. (2014). *Digital India Report*. (October 2016). Retrieved from http://deity.gov.in/sites/upload_files/dit/files/Digital India.pdf

Deng, Y., Lei, H., Li, X., & Lin, Y. (2018). An improved deep neural network model for job matching. *2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018*, 106–112. https://doi.org/10.1109/ICAIBD.2018.8396176

Dijkman, R., & Ipeirotis, P. (2015). Using Twitter to predict sales: a case study. *ArXiv Preprint ArXiv: 1503.04599.*, 1–14. Retrieved from http://arxiv.org/abs/1503.04599

Duan, J. W., Chen, Y. H., Liu, T., & Ding, X. (2015). Mining Intention-Related Products on Online Q&A Community. *Journal of Computer Science and Technology*, *30*(5), 1054–1062. https://doi.org/10.1007/s11390-015-1581-7.

Elrhoul, M. (2015, July 30). Research: Four ways brands can build customer service relationships on Twitter. Twitter. Retrieved from https://blog.twitter.com/2015/researchfour-ways-brands-can-build-customer-service-relationships-on-twitter.

Gautam, G., & Yadav, D. (2014). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. *2014 7th International Conference on Contemporary Computing, IC3 2014*, (October 2015), 437–442. https://doi.org/10.1109/IC3.2014.6897213

Gupta, V., & Lehal, G. S. (2014). A Survey of Text Mining Techniques and Applications - Volume 1, No. 1, August 2009 - JETWI. *Journal of Emerging Technologies in Web Intelligence*, *1*(1), 60–76. Retrieved from http://www.jetwi.us/index.php?m=content&c=index&a=show&catid=165&id=969

Haque, A., & Hoque, M. M. (n.d.). *Bangla Word Sense Disambiguation System using Dictionary Based Approach*.

Hoffmann, P. (2005). *Doctorado en ciencias sociales Profesor : Alexis De Greiff Presentado Por : Lisandro Soto Flechas*. (October), 347–354.

Horiguchi, Y., Baba, Y., & Kashima, H. (2017). Predicting Fuel Consumption and Flight Delays for Low-Cost Airlines. *Proceedings of the 31th Conference on Artificial Intelligence (AAAI 2017)*, 4686–4693.

Hu, G., Bhargava, P., Fuhrmann, S., Ellinger, S., & Spasojevic, N. (2017). Analyzing Users' Sentiment Towards Popular Consumer Industries and Brands on Twitter. *IEEE International Conference on Data Mining Workshops, ICDMW*, *2017-Novem*, 381–388. https://doi.org/10.1109/ICDMW.2017.55

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? the mediating role of online review sentiments on product sales. *Decision Support Systems*, *57*(1), 42–53. https://doi.org/10.1016/j.dss.2013.07.009

In, I., & Other, O. W. (2016). *a Re Y Ou*. *90230*(310), 761–767. https://doi.org/10.1016/j.jsps.2009.10.001

Karmaker Santu, S. K., Sondhi, P., & Zhai, C. (2016). *Generative Feature Language Models for Mining Implicit Features from Customer Reviews*. (October 2017), 929–938. https://doi.org/10.1145/2983323.2983729

Khan, F., Borah, S., & Pradhan, A. (2016). Mining Consumption Intent from Social Data : A Survey. *International Journal of Computer Applications*, (September), 975–8887. Retrieved from https://www.researchgate.net/profile/Samarjeet_Borah4/publication/308525949_Mining_Consumption_Intent_from_Social_Data_A_Survey/links/57e6523f08aed7fe4668dcf0/Mining-Consumption-Intent-from-Social-Data-A-Survey.pdf%0Ahttps://www.researchgate.net/profile/S

Kim, Y. (2014). *Convolutional Neural Networks for Sentence Classification*. Retrieved from http://arxiv.org/abs/1408.5882

Lacic, E., Kowald, D., & Lex, E. (2016). *High Enough? Explaining and Predicting Traveler Satisfaction Using Airline Review*. Retrieved from http://arxiv.org/abs/1604.00942

Leung, C. W. K., & Chan, S. C. F. (2008). Sentiment Analysis of Product Reviews. *Encyclopedia of Data Warehousing and Mining, 2nd Edition*, (February), 1794–1799. Retrieved from http://www.igi-global.com/chapter/sentiment-analysis-product-reviews/11061

Liou, J. J. H., & Tzeng, G. H. (2010). A Dominance-based Rough Set Approach to customer behavior in the airline market. *Information Sciences*, *180*(11), 2230–2238. https://doi.org/10.1016/j.ins.2010.01.025

Liu, B. (2010). Sentiment Analysis and Subjectivity in: Handbook of Natural Language Processing, Second Edition. *Handbook of Natural Language Processing, Second Edition*, *2*, 568.

Matsunawa, T., Nojima, S., & Kotani, T. (2016). Automatic layout feature extraction for lithography hotspot detection based on deep neural network. *Design-Process-Technology Co-Optimization for Manufacturability X*, *9781*, 97810H. https://doi.org/10.1117/12.2217746

Mikolov, T., Yih, W.-T., & Zweig, G. (2013). *O'Sullivan - Classification of Lumbopelvic Pain Disorders.pdf*. (June), 9–14. https://doi.org/10.3109/10826089109058901

Misopoulos, F., Mitic, M., Kapoulas, A., & Karapiperis, C. (2014). Uncovering customer service experiences with Twitter: The case of airline industry. *Management Decision*, *52*(4), 705–723. https://doi.org/10.1108/MD-03-2012-0235

Oliver Breen, J. (2012). *Mining Twitter for Airline Consumer Sentiment*. https://doi.org/10.1016/B978-0-12-386979-1.00008-6

Omar, M. S., Njeru, A., Paracha, S., Wannous, M., & Yi, S. (2017). Mining tweets for education reforms. *Proceedings of the 2017 IEEE International Conference on Applied System Innovation: Applied System Innovation for Modern Technology, ICASI 2017*, (May), 416–419. https://doi.org/10.1109/ICASI.2017.7988441

Pang, B., & Lee, L. (2008). LR..references..Opinion Mining and Sentiment Analysis. In *Foundations and Trends® in Information Retrieval* (Vol. 2). https://doi.org/10.1561/1500000011

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up? Sentiment Classification using Machine Learning Techniques*. Retrieved from http://arxiv.org/abs/cs/0205070

Paper, C., & Inspired, N. (2017). *Online Social Media-based Sentiment Analysis for US Airline companies Online Social Media-based Sentiment Analysis for US Airline companies*. (April).

Paquette, H. (2013). Les insuffisances mitrales isch??miques: principes de leur traitement chirurgical. *Revue Medicale de Liege*, *50*(7), 295–297. https://doi.org/10.18276/miz.2016.46-05

Plant, O. H. (2016). *Social media and sales: Determining the predictive power of sentiment analysis towards car sales*. 1–17.

S. Gharehchopogh, F., & Khalifehlou, Z. (2011). Analysis and evaluation of unstructured data: Text mining versus natural language processing. In *2011 5th International Conference on Application of Information and Communication Technologies, AICT 2011*. https://doi.org/10.1109/ICAICT.2011.6111017

Savoy, O. (2012). Feature Selection in Sentiment Analysis. *In Proc. CORIA*, 273–284. Retrieved from http://www.asso-aria.org/coria/2012/273.pdf

Sebastiani, F. (2001). *Machine Learning in Automated Text Categorization*. https://doi.org/10.1145/505282.505283

Severyn, A., & Moschitti, A. (2015). *UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification*. (SemEval), 464–469. https://doi.org/10.18653/v1/s15-2079

Socher, R. (2016). *Deep Learning for Sentiment Analysis - Invited Talk*. 36–36. https://doi.org/10.18653/v1/w16-0408

Stavrianou, A., Andritsos, P., & Nicoloyannis, N. (2007). Overview and semantic issues of text mining. *ACM SIGMOD Record*, *36*(3), 23. https://doi.org/10.1145/1324185.1324190

Sundermeyer, M., Schlueter, R., & Ney, H. (2012). LSTM Neural Networks for Language Modeling. *Proceedings of INTERSPEECH*, 194–197. https://doi.org/10.1109/ISIT.2008.4595144

Tang, D., Wei, F., Qin, B., Liu, T., & Zhou, M. (2015). *Coooolll: A Deep Learning System for Twitter Sentiment Classification*. (SemEval), 208–212. https://doi.org/10.3115/v1/s14-2033

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2006). *Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification*. *2*, 1–54. https://doi.org/10.1561/2200000006

Tripathy, A., & Rath, S. K. (2016). Classification of Sentiment of Reviews using Supervised Machine Learning Techniques. *International Journal of Rough Sets and Data Analysis*, *4*(1), 56–74. https://doi.org/10.4018/ijrsda.2017010104

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning. *Journal of Artificial Intelligence Research*, *37*, 141–188. Retrieved from https://www.jair.org/media/2934/live-2934-4846-jair.pdf

VijayGaikwad, S., Chaugule, A., & Patil, P. (2014). Text Mining Methods and Techniques. *International Journal of Computer Applications*, *85*(17), 42–45. https://doi.org/10.5120/14937-3507

Vinodhini, G. (2012). *SA2 - Copy*. *2*(6).

Vo, D. T., & Zhang, Y. (2015). Target-dependent twitter sentiment classification with rich automatic features. *IJCAI International Joint Conference on Artificial Intelligence*, *2015-Janua*(Ijcai), 1347–1353.

Wan, Y., & Gao, Q. (2016). An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis. *Proceedings - 15th IEEE International Conference on Data Mining Workshop, ICDMW 2015*, (March), 1318–1325. https://doi.org/10.1109/ICDMW.2015.7

Wang, Z. (2017). *The Evaluation of Ensemble Sentiment Classification Approach on Airline Services Using Twitter The evaluation of Ensemble sentiment Classification approach on airline services using Twitter* . https://doi.org/10.21427/D7190M

Wehrmann, J., Becker, W., Cagnini, H. E. L., & Barros, R. C. (2017). A character-based convolutional neural network for language-agnostic Twitter sentiment analysis. *Proceedings of the International Joint Conference on Neural Networks*, *2017-May*(May), 2384–2391. https://doi.org/10.1109/IJCNN.2017.7966145

Záruba, F. (1971). Problémy keratinizace. C. Predběžné závěry. *Cesko-Slovenska Dermatologie*, *46*(5), 223–229.

Zhang, M., Zhang, Y., & Vo, D. (2016). Gated Neural Networks for Targeted Sentiment Analysis. *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)*, 3087–3093.