Emotion and Sentiment Detection in Unstructured Social Data

by

Maha AlGhalibi

Approved Dissertation thesis for the partial fulfilment of the requirements for a Doctor of Natural Sciences (Dr. rer. Nat)

Fachbereich 4: Informatik

Universität Koblenz-Landau

Chair of PHD Board: Prof. Dr. Ralf Lämmel

Chair of PHD Commission: Prof. Dr. Petra Schubert

Examiner and Supervisor: Prof. Dr.-Ing. Habil. Kai Lawonn

Further Examiner: Prof. Dr. Maria Wimmer

Date of the doctoral viva: December 22. 2021

Acknowledgements

I would greatly like to acknowledge the sincere guidance, the warm valuable support, besides the great soul and aspirations of my supervisor, J.-Prof. Dr.-Ing. habil. Kai Lawonn.

I gratefully like to express my gratitude and love feelings to my family for their sacrifice, encouragement, and support. Special appreciation to the staff members in the Institute for Computational Visualistics, the University of Koblenz-Landau for their valuable cooperation.

Thanks are also expressed to my colleagues in the field who provided beneficial interactions during the research years.

Finally, thanks are due to the Institute for Computational Visualistics, University of Koblenz-Landau, Germany which provided the opportunity for this work to be developed and finalized, and to the Ministry of Higher Education and Scientific Research, Iraq, for the financial support.

Dissemination

Journal Articles

- Maha Alghalibi, Adil Al-Azzawi, Kai Lawonn, "Deep Tweets Analyzer Model for Twitter Mood Visualization and Prediction Based Deep Learning Approach," International Journal of Computer and Communication Engineering vol. 8, no. 1, pp. 1-17, 2019.
- Maha Alghalibi, Adil Al-Azzawi, Kai Lawonn, "Deep Attention Learning Mechanisms for Social Media Sentiment Image Revelation," International Journal of Computer and Communication Engineering vol. 9, no. 1, pp. 1-17, 2020.

Conference Articles

- Maha Al-Ghalibi and Kai Lawonn "Topic aspects-based generative mixture model for movie recommendation system using deep convolutional network", Proc. SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019), 114333K (31 January 2020); https://doi.org/10.1117/12.2556294
- Maha Al-Ghalibi, Adil Al-Azzawi, and Kai Lawonn "NLP based sentiment analysis for Twitter's opinion mining and visualization", Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018), 110412A (15 March 2019).
- Maha Al-Ghalibi, Adil Al-Azzawi. (2018). "Time series social network visualization based on dimension reduction" International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision,publish by Skala-UNION Agency,CSRN 2803.

Abstract

Social media provides a powerful way for people to share opinions and sentiments about a specific topic, allowing others to benefit from these thoughts and feelings. This procedure generates a huge amount of unstructured data, such as texts, images, and references that are constantly increasing through daily comments to related discussions. However, the vast amount of unstructured data presents risks to the information-extraction process, and so decision making becomes highly challenging. This is because data overload may cause the loss of useful data due to its inappropriate presentation and its accumulation. To this extent, this thesis contributed to the field of analyzing and detecting feelings in images and texts. And that by extracting the feelings and opinions hidden in a huge collection of image data and texts on social networks After that, these feelings are classified into positive, negative, or neutral, according to the features of the classified data. The process of extracting these feelings greatly helps in decision-making processes on various topics as will be explained in the first chapter of the thesis. A system has been built that can classify the feelings inherent in the images and texts on social media sites, such as people's opinions about products and companies, personal posts, and general messages. This thesis begins by introducing a new method of reducing the dimension of text data based on data-mining approaches and then examines the sentiment based on neural and deep neural network classification algorithms. Subsequently, in contrast to sentiment analysis research in text datasets, we examine sentiment expression and polarity classification within and across image datasets by building deep neural networks based on the attention mechanism.

Kurzfassung

Soziale Medien bieten eine leistungsstarke Möglichkeit für Menschen, Meinungen und Gefühle zu einem bestimmten Thema auszutauschen, sodass andere von diesen Gedanken und Gefühlen profitieren können. Dieses Verfahren erzeugt eine riesige Menge an unstrukturierten Daten, wie Texte, Bilder und Verweise, die durch täglich anwachsende Kommentare zu verwandten Diskussionen ständig zunimmt. Die riesige Menge an unstrukturierten Daten stellt jedoch ein Risiko für den Prozess der Informationsextraktion dar, sodass die Entscheidungsfindung zu einer großen Herausforderung wird. Dies liegt daran, dass die Datenflut zu einem Verlust von nützlichen Daten aufgrund ihrer unangemessenen Darstellung und ihrer Anhäufung führen kann. Insofern leistet diese Arbeit einen Beitrag zum Gebiet der Sentimentanalyse und des Opinion Mining, das darauf abzielt, Emotionen und Meinungen aus riesigen Text- und Bilddatensätzen zu extrahieren. Das ultimative Ziel ist es, jeden Text oder jedes Bild als Ausdruck einer positiven, negativen oder neutralen Emotion zu klassifizieren, um bei der Entscheidungsfindung zu helfen. Sentiment- und Meinungsklassifikatoren wurden für Text- und Bilddatensätze aus sozialen Medien entwickelt, z. B. für Firmen- oder Produktbewertungen, Blogbeiträge und sogar Twitter-Nachrichten. In dieser Arbeit wird zunächst eine neue Methode zur Reduktion der Dimension von Textdaten auf Basis von Data-Mining-Ansätzen vorgestellt und anschließend das Sentiment auf Basis von neuronalen und Deep Neural Network-Klassifikationsalgorithmen untersucht. Anschließend untersuchen wir im Gegensatz zur Sentiment-Analyseforschung in Textdatensätzen die Sentiment-Ausdrucks- und Polaritätsklassifikation innerhalb und über Bilddatensätze hinweg, indem wir tiefe neuronale Netze auf Basis des Aufmerksamkeitsmechanismus aufbauen.

Table of contents

1 Introduction **1**

1.1 Sentiment Analysis in unstructured social Big Data 1
1.2 Motivations 3
1.3 Problem Statement 4
1.4 Contributions 6
1.5 Thesis Overview 8

2 Literature Review **11**

2.1 Literature Survey Areas 11
2.1.1 Data mining approaches for sentiment analysis 12
2.1.2 Higher dimensionality reduction 13
2.1.3 Sentiment analysis in social media texts 16
2.1.4 Visualizing emotions in social media images 20
2.1.5 Machine learning 28

3 Fundamentals **31**

3.1 Data Mining and Machine Learning 31
3.1.1 Supervised learning 31
3.1.2 Unsupervised learning 32
3.2 Machine Learning and Sentiment Analysis 33
3.3 Data Mining Dimensionality Reduction and Feature Extraction Models 35
3.3.1 Singular value decomposition (SVD) 35
3.3.2 Principal component analysis (PCA) 43
3.3.3 Mutual information (MI) 43
3.4 Data Mining Prediction and Classification Models. 44
3.4.1 Artificial neural network prediction model 44
3.4.2 Deep learning prediction model using convolutional neural networks 55
3.4.3 Prediction and classification model using deep learning long short-term memory networks and one-dimensional convolutional networks 59

3.4.4 Prediction model based attention learning mechanisms 61

4 NLP-based Sentiment Analysis and Reviewer's Opinion Mining Prediction 65

4.1 Sentiment in Social Media Text 65

4.2 First Model: NLP-Based Opinion-Mining Prediction Using Dimensionality Reduction and a Residual Neural Network Classification Algorithm 66

4.2.1 First step: Dataset pre-processing based o natural language processing 67

4.2.2 Second step: Analysis of the reviewer's sentiment dataset 68

4.2.3 Third step: Dimensionality reduction feature extraction and selection 68

4.2.4 Fourth step: prediction and classification 71

4.3 Second Model: Users' Mood Prediction Based on Deep Mood Prediction Using a One-Dimensional Convolutional-Neural-Network-Based Deep LSTMs Approach 75

4.3.1 First step: Dataset pre-processing based on natural language processing 77

4.3.2 Second step: Twitter sentiment dataset analysis and visualization 78

4.3.3 Deep-learning prediction and classification 79

4.4 Sentiment Analysis Dataset 82

4.4.1 U.S. airlines dataset 82

4.4.2 Amazon review dataset 83

4.5 Evaluation Metric 83

4.6 First Model Experimental Results Using the U.S. Airlines Dataset . . . 84

4.6.1 Twitter airlines dataset statistical observation 84

4.6.2 NLP-based tweet opinion prediction and classification performance results 86

4.6.3 Comparison with other prediction approaches 92

4.7 First Model Experimental Results Using the Amazon Review's Dataset . 92

4.7.1 Amazon review's statistical observation 92

4.7.2 NLP-based Amazon review's dataset prediction and classification performance results 93

4.7.3 Comparison with other approaches 98

4.8 Second Model Experimental Results Using the U.S. Airlines Dataset . . 99

4.8.1 Deep Twitter opinion-mining prediction model 101

4.8.2 Comparison with other approaches 104

4.9 Summary 105

5 Social Media Image Sentiment Analysis Based Deep Attention Learning Mechanisms 107

5.1 Emotion in Social Media Image 107
5.2 Proposed System 108
5.3 System Overview and Contribution 110
5.4 Sentiment Analysis Dataset 113
5.5 Experimental Results 114
5.5.1 Training experimental results 114
5.5.2 Evaluation criteria 116
5.5.3 Testing performance and experimental results 116
5.6 Summary 120

6 Conclusion and Future Works 121

6.1 Addressing the Research Questions 121
6.2 Summary 123
6.3 Future Work 124

References 125

Internet-resources 137

Appendix Appendix A 139

Appendix Appendix B 143

Abbreviations 144

Chapter 1

Introduction

This chapter aims to highlight the problem area's environment, motivations, and identification of the problem of sentiment analysis and emotion detection in the field of unstructured social big data. It explain the problem statement through analyzing and detecting the emotion and opinion from an unstructured dataset involving texts and images. The research objectives and deliverables are then unveiled. This chapter further articulates the methods that were followed to achieve the research objectives, including the approaches and techniques utilized and the contributions made to the literature. Finally, this chapter concludes with the thesis overview.

1.1 Sentiment Analysis in unstructured social Big Data

Organization sites, clouds, infrastructures, and various websites contain large amounts of data and most of this data is irregular data from texts, images, and videos that the more it accumulates, the more difficult it is to interpret and benefit from it. With this thesis, we will clarify how to benefit from this social big data by analyzing it and extract sentiment to help in decisions making procedures on various topics.

Big Data

Big data is growing and becoming highly important in different areas. In simple terms, it deals with the combination of data from different sources and recognizes patterns in the data, which can be utilized for various objectives, such as developing marketing, medical, and educational research [GH15]. Scientific and technological development has resulted in a massive amount of data that is now being produced in everyday life. Companies and governments have begun to understand the importance of using this data for their growth. As a result, the investigation of big data has gained prominence among academics in various areas of research [EL13, MSC13] and has also received interest in non-academic fields [Loh12]. The idea of big data includes the collection of data obtained from various sources. In Preparing this information and using the results, big data is stored in huge databases requiring complex processing and visualization methods for extracting useful information and knowledge that cannot be handled by traditional and regular data-processing software [FHL14]. Big data can be defined as huge amounts of data that are of high volume and velocity, which come from a variety of sources and which need new methods of processing to allow enhanced decision making and insight innovation [BL12]. The terms "volume" and "amount" here indicate the complexity of datasets and rather than their size. "Variety" highlights the various types of structured or unstructured data, such as text, image, audio, and video. "Velocity" explains which data is capable of being analyzed. Data with such features can be manifested in public comments, individual's opinions, internet access history, private communications, and health records. This thesis aims to obtain insights from huge and complex collections of data and handle the information collected over the internet. Moreover, the extraction of knowledge from huge amounts of data published by people from different platforms and geographic areas is described. In this chapter, an introduction to unstructured big data from social media and some of its applications in various fields, including emotion detection and decision making, is presented.

Social Media

Social media platforms and microblogging services, such as Twitter, Flickr, and Facebook, are frequently utilized by users to access and share information about important events, news, and other topics. These different tools of expressions allow people to present and share their opinions in an easy way [JZSC09]. Emotion analysis or opinion mining relates to the techniques from areas like natural language processing (NLP), information retrieval (IR), and machine learning (ML) to recognize and extract useful knowledge from image and textual datasets [PLO08]. The most popular emotion analysis and detection tasks are the automatic classifications of images or sentences into emotion categories, such as positive, negative, and neutral. These emotion classes represent individual's opinions on a specific topic. Emotion analysis applied to social media platforms has received increasing attention from the research community for it has great importance in many areas such as marketing, sports, and politics. Some research makes use of certain social phenomena, such as stock prices, recommendation systems, and political elections [BMZ11]. The feelings expressed in these can be utilized to judge the emotion indirectly [OBRS10]. This thesis focuses on the data analysis and sentiment detection on Twitter and Flickr. We use Twitter and Flickr because they are the most widely-known and used services. Therefore, it is more likely that they produce large amounts of openly available public data.

Sentiment Analysis

When social media and the web were first used, the content was presented and published by the owners of sites and blogs connected with regular information sources, such as news media and companies. The content was primarily about "truths" which were real emotions captured in relation to special events or topics. In the 2000s, the growth of the Web era began, with the introduction of blogs, online social networks, and microblogging services. This represented a great breakthrough by enabling users to share textual and visual content easily. This huge step enabled massive amounts of useful information, i.e., individual's opinions, to be available on social media, which opened a new door for developers in understanding and explaining the underlying feelings related to this high-dimensional data. Sentiment analysis is a process that takes data, whether it is text, image, audio, or video as an input and returns an "output" that shows whether the data is positive or negative.

Why sentiment analysis is needed

Due to the large amount of data that today has increased rapidly over a variety of social media platforms and companies, it is impossible to analyze this data manually with no error or bias. Each party, whether customers or companies, needs vision to improve decision-making processes, but the two sides are ignorant of the best way to obtain it. This is where sentiment analysis comes into play, as the analysis can be automated, and decisions can be made based on a large amount of data rather than a simple intuition that is not always true.

What is sentiment analysis used for?

Sentiment analysis enables the gathering of quick insights from huge amounts of text and image data. Two examples where sentiment analysis can be helpful are presented below.

1. Company growth: When looking at the momentum in customer opinions, sentiment analysis can be used to draw conclusions based on customer feedback, whether text or image, through checking if feelings are negative or positive. Likewise, customer reviews that have a strong positive feeling can be examined in order to discern, for example, why these customers love this company and thus to help the company to focus on what it can do to increase the number of its promoters.

2. Stock Trading: Stock-trading companies that search the internet for news. Can use sentiment algorithms to detect specific companies that feel positive in news articles. This could reveal a huge financial opportunity, as it could lead to the purchase of more company shares. Access to this type of data may allow traders to make decisions before the market has enough time to respond.

1.2 Motivations

This research focuses on knowledge extraction and emotion detection based on data mining, artificial intelligence, ML, and social big data visualization. The increase in the amount of data, especially on social media platforms, has led to the emergence of multiple challenges. One of these inevitable consequences is that data is collected by many storage devices in a way that influences our methods when it comes to dealing with information; for instance, most information, such as date and time, is stored without applying any refinement or filtering. In addition to this consequence, there is a vast amount of data that has been produced by social media platforms and companies, which makes the problem of using data even worse when decision making is challenging. In other words, data overload may cause the loss of useful data due to its inappropriate presentation. Another challenge is how to extract knowledge from big chunks of data. Visual analytics is the implementation of several techniques to illustrate the relationships within the big data and presents data in a way that makes it easy to understand and interpret [KQM13]. The motivation behind most studies on visual analytics has been to devise various approaches and architectures to develop essential models in the hope of building a reliable visual analytics machine. The motivations behind the current research can be summarized as follows:

1. The observation of huge amounts of data on social media through opinion mining, predicting a person's mood, and visualizing the results: This observation inevitably leads to the elicitation of meaningful decisions. One example application of using this approach is in the prediction of customers' behavior in e-commerce, and this helps other customers to make a decision based on previous behavior systems [POG99].

2. The study of the underlying meaning of posted images: In other words, what are the intentions behind the posted photographs and what emotions do they carry? This is of benefit in many areas, such as health care and video gaming [MSJ18, SS15b].

These motivations reflect the need for dealing with the problems of data overload, the mess, and the loss of useful information problems that have arisen. The themes of this thesis are two-fold: First, the ability to extract meaningful knowledge from text and, secondly, how to perform a similar task with respect to images. Both of these aspects have become serious issues affecting the ability to extract useful knowledge.

1.3 Problem Statement

The tremendous development in the internet and the web has not only provided huge amounts of data containing individual's opinions and feelings that are digitally stored, but has also offered the opportunity to understand and detect these feelings and sentiments of the public by analyzing these huge amounts of large-scale data. However an important point to is that more data, leads to increased difficulty in extracting useful information. Research has revealed that Twitter users generate 21 million tweets per hour [Sim15], while flickr users generate 3.5 million new images everyday [Jef13]. Resarch has also shown that, because of this huge growth in data, more than half of online clients suffer from frustration while, for example, shopping online, selecting a top hospital, choosing a good hotel, or finding comfortable airline to book tickets with. The customers frustration arises from being unable to make the appropriate decision about purchasing a product or choosing a specific airline, and this is because of the vast number of opinions presented online and the need for time and resources to analyze them and make a decision. Moreover, companies face the same problem, as the same survey indicated that three-quarters of 2,100 companies had no clear sense of what their most important customers think of them and that 31% found it difficult to measure customer opinion [Ste12]. According to the above, the most important challenge is to determine feeling and opinion as accurately as possible. Therefore, the most critical problems related to the identification of opinion from texts are the high dimensions and the scattered matrix that affect the accuracy of its classification [YIAH14]. This thesis also discusses the problem of defining the feelings embedded in the images that people share on social media platforms on a particular topic. The issue here lies in the tremendous diversity of these images, despite their depicting the same emotion of, for example, positivity and in the need to focus on the features of the images, especially the objects, to get the highest accuracy in the classification [GG15]. Based on the above problems, the research questions of this thesis can be summarized as follows:

1. How can information about sentiment be extracted from texts that contain huge numbers of words to help in complex decision making?

2. How can the process of sentiment detection described in answering the first research question be improved by increasing performance and accuracy while maintaining the premise of huge amounts of data?

3. How can emotion be inferred from a given image?

4. Is it possible to provide automated systems that can understand the sentiment and emotions of data available on social networking platforms?

To answer these fundamental questions, and to meet the need to develop corresponding systems capable of matching them numerous methods, approaches, and techniques proposed and implemented to achieve and solve the required visual analytics problem in the field of unstructured big data. Devising such models and algorithms involves analyzing and detecting the emotion and opinion in an unstructured dataset that includes texts and images. Each research question has previously been discussed in three separate publications, which constitute Chapters 4 and 5.

1.4 Contributions

The main goal of this work is to create three systems that answer the research questions in Section 1.5. The first system, which addresses the first research question, is based on NLP, data-mining dimension reduction, and classification methods to provide text analysis and prediction for detecting the sentiment in textual data. The second system deals with the second research question, aiming to increase the accuracy of the first approach by employing deep neural networks. The third system is based on deep neural networks and attention mechanisms for detecting the emotion in images and satisfies the third research question. All these three systems take the sentiments and feelings expressed within the unstructured data available on social media platforms as input. Upon testing, it is demonstrated that the models presented provide significant accuracy in classifying tweets, reviews and images taken from Twitter, Amazon and Flickr. The thesis thereby encapsulates the main four objectives detailed below:

1. To build a system of opinion mining and the prediction of individuals' moods based on data extracted from social media platforms through NLP sentiment analysis of text-feature extraction and dimension reduction, in addition to ML for classification.
2. To improve the prediction of individuals' moods based on data on social media platforms through deep learning (DL) and neural networks.
3. To provide a comparison between the first two approaches and compare the two with similar approaches from other studies.
4. To build a system that is able to infer positive and negative emotions using visual emotion and sentiment analysis, through deep neural networks and attention mechanisms.

Hence, the research work addressed by this thesis, and, as mentioned above, the issue can be considered to be simply a single problem involving the multiple facets of detection, classification, and quantification of sentiment, emotion, and opinion in text and images in any form. Figure 1.1 illustrates the general structure of the proposed system, in which the main data sources used in this thesis are as follows: Thereby, the main sources used in this thesis are:

1. English text readily collected from Twitter [144].
2. Amazon products reviews dataset [145].
3. A dataset that records the polarity of images from Flickr [143].

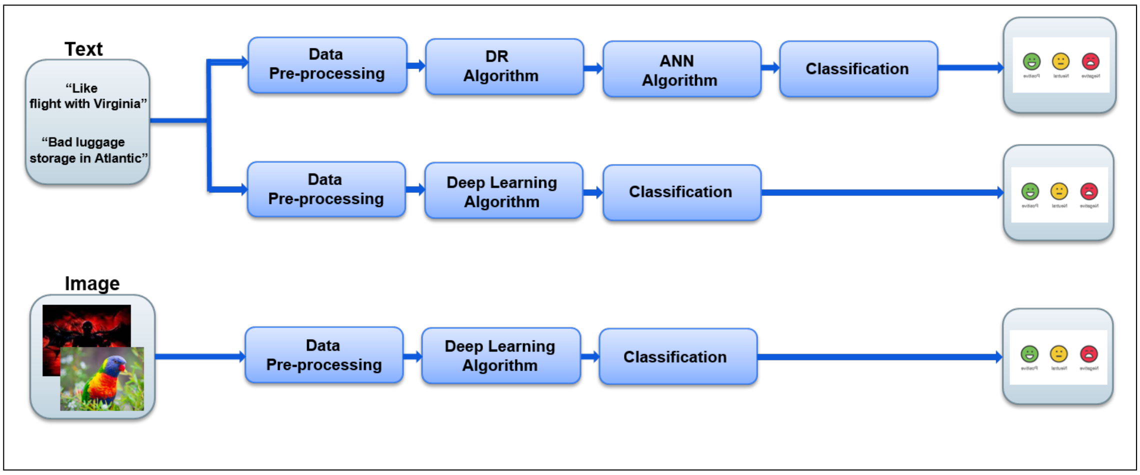


Fig. 1.1 : General structure of the proposed systems.

These three datasets were considered in order to provide information that has utility in a variety of ways, in particular sentiment, emotion and opinion mining. A method is proposed to classify text and images into the different emotion categories of negative, positive, and neutral. The achievement of this research project is embodied in several contributions offered by the underlying four deliverables:

1. The first contribution is to devise an intelligent opinion-mining system capable of analyzing and visualizing users' opinions on Twitter [AGAAL19]. As further explained in Chapter 4, to achieve such a deliverable, text-features extraction, classification, and dimensionality-reduction techniques are all resorted using NLP based sentiment analysis. The methodology applies the following steps:

    - In the pre-processing step, automated dimensionality-reduction and feature-selection processes are proposed based on singular value decomposition (SVD) and mutual information (MI). This new composite method has obtained better results in the classification process compared with other dimensionality-reduction approaches and different classification methods.
    - A successfully devised system built to automatically (with no manual effort) provide an efficient training set for the ML classifiers used, consisting of a set large enough to involve labeled tweets from all emotion categories.
    - A large assortment bag of words in English is created, which consists of words keen to express a particular emotion along with the intensity of that emotion.

    In fact, such a method of application upgrades the utilized back propagation neural network (BPNN) in the prediction and classification of text features with higher accuracy than some other available approaches, like principal component analysis (PCA) and SVD, even with the whole feature space. Moreover, a comparison with different classification methods was conducted. The pertinent model thereby demonstrates better results regarding accuracy and efficiency compared with other recent techniques. The U.S. Airlines Sentiment Analysis Twitter dataset [144] and Amazon reviews dataset [145] is applied in this case.

2. The second contribution is to design, implement, and validate an analogous system capable of classifying, extracting, and predicting individuals' moods from Twitter using a DL neural network approach [AAAL19]. To realize such a deliverable, the pertinent mechanism of the utilized methodology involves almost the same above-mentioned framework steps of the first contribution: Collect the data, pre-process it, NLP for analyze the data, and predict the sentiment by classification methods. in the process of predicting individuals' moods from Twitter is alternatively based on using the DL model. A Twitter dataset of individual's Sentiments about U.S. Airlines was also utilized here [144]. This contribution fully explained in Chapter 5.

3. The final contribution is the introduction of a DL system model to analyze and visualize emotions in social media images through the classification and recognition of image-embedded sentiment patterns [AAAL20]. Through utilizing an adequate technique for DL, (the attention mechanism), this study introduces a deliverable capable of extracting the implied sentimental status (embedded emotional responses) for each social media image as Happy and Sad, which are equivalent to Highly Positive mood and Highly Negative mood, respectively. The utilized methodology hinges on accomplishing two stages: The first is the system's design and implementation, and the second is its validation by applying it to social media images. Thereby, the deliverable here is to introduce a system with a deep attention network mechanism (DANM) to achieve the higher-level outcome of social media sentiment image analysis, in addition to the sentiment classification. This leads to organizing the DANM system such that it utilizes an appropriate technique of ML, which is the convolutional neural network (CNN), to produce the required image feature map. Flickr and Twitter datasets are utilized [143], while the pertinent algorithm is developed using the Matlab 2017a/b platform.

1.5 Thesis Overview

This thesis contains six chapters. Chapter 1 provides an introduction to the environment of the problem area and to the research questions, research motivations, and contributions. Chapter 2 introduces a review of the most relevant research and approaches to sentiment analysis and social media, while Chapter 3 explains the background and the theory on which thesis is based. Chapter 4 describes details of the first deliverable of this thesis, the intelligent opinion-mining system [AGAAL19], and the second deliverable, the Prediction of individuals' mood from Twitter [AAAL19]. Chapter 5 presents the third deliverable: Visualizing emotions in social media images [AAAL20]. Chapter 6 is the thesis summary, which outlines the uniqueness of the research in its deliverables and contributions by comparing its results with those of other contemporary works, while offering a perspective for future work.

Chapter 2

Literature Review

This chapter aims to present a concise literature survey of other works in the research problem field of visual analytics of unstructured big data in social media. It focuses on the problems handled by this thesis, namely, intelligent opinion mining, mood prediction, and visual emotions and sentiments detection. This chapter, therefore, introduces a summary of the most relevant works among the presented literature survey areas.

2.1 Literature Survey Areas

Solving serious problems or taking everyday decisions cannot be accomplished by people in isolation. For instance, people regularly consult the opinions of others when they want to elect a specific candidate. People often read what others write about a specific product before they buy it and when planning a trip, opinions and suggestions about the best areas are often sought by asking friends. Currently, after a period of great technological development, it has become much easier to obtain the opinions and suggestions of millions of people. These opinions and emotions are available on the internet due to the emergence of platforms for communication between people. These are called social media networks. The total time the world spends on the internet and social networks is 22% of all time. Of these users, 65% are adults [BM01]. The increase in the use of discussion and social networking platforms on the internet has led to a tremendous increase in the amount of data and information about various subjects, such as marketing, medicine, education, tourism, and politics. With the increasing generation of data from various platforms, several problems have emerged, including data overload, information loss, and difficulty in decision-making, because large amounts of data require time, effort, and tools to interpret and extract useful information. One of the most commonly used methods of interpreting data today is analyzing emotions and extracting opinions. Emotional analysis techniques have appeared in the past decade to analyze the answers and opinions of people on a specific topic and help make decisions according to the outcome of the analysis process [BM01]. Sentiment analysis and emotion detection methods can be implemented on different data types, such as audio, image with or without text, and video. A common approach to deal with the task of sentiment analysis is the use of data mining algorithms and classification to detect emotions from the data available on social media platforms.

The survey referred to in this Section involves works encompassing several different approaches, with various intrinsic schemes. This has been synopsized in the following five areas, listed across Subsections 2.2.1 to 2.2.5.

2.1.1 Data mining approaches for sentiment analysis

Detecting emotions in social media data based on mining techniques is one of the most recent and most vital fields of applications in structured and unstructured big data analytics. As such, several works have been published in this area. Amongst these is the research presented by Liu and Zhang [LZ12]. It a survey on opinion mining and sentiment analysis. At the beginning, they explain the problem of revealing opinion. Then, they explain the technical issues that needed to be addressed. The various mining tasks that have been studied are described later in the research literature they present. Next, they discuss the issue of identifying unwanted opinions. Finally, they present their research as a tool to assess the quality of online reviews.

Patel et al. [PPB15] also presented their work, which concerns another survey to explore opinions and sentiment analysis, providing another viewpoint. The study shows that the Internet and various web resources contain a large amount of irregular data and information that needs interpretation to extract useful information and avoid losing it. This study showed that using machine learning directly is insufficient to obtain this useful information. The researchers assert that the increased demand for knowledge of people's opinions about political events, product sales, and campaigns led to an emphasis on studying sentiment analysis and opinion classification. They conclude that to obtain an opinion, the enormous amount of data must be analyzed, important lines that express opinions must be examined, and the polarity of the opinions must be determined. The big challenge is in analyzing and identifying the feelings hidden in the irregular data. This scientific paper is a summary that covers the problem of exploring emotions and most of the techniques used.

From a different perspective, Gomez and Caceres [GC17]. focused on using data-mining techniques to build a feature-extraction system. They then used it to classify human emotions, such as happiness and hope or sadness and anger, when a song was played. In this paper, data-mining algorithms, such as Multi-Label K-Nearest Neighbors and, Random K-Label sets, are used to clarify the relationship between music and human feelings.

Aggarwal et al. [AWO10] presented work concerning data-mining. They studied the problems encountered in the process of extracting opinions and feelings from large, unstructured data. They summarized these problems as four important points that a researcher should consider when working on a project to search for opinions and feelings by relying on data-mining methods. These important points are clustering, classification, pattern mining, and outlier analysis. This paper discusses extensively a wide range of techniques that have been used to solve the problems of mining various types of big data.

Mahindrakar and Hanumanthappa [MH13] published an article about data mining in healthcare. This study emphasizes that obtaining knowledge from big medical data requires nothing more than changing the state of the data from low-level to high-level information. Their article presents a survey of the pertinent techniques and algorithms and the main limitations and challenges in this field. They conclude that the state-of the-art need is for devising algorithms with very high accuracy for medical diagnosis. For the area of data analytics, the work provides techniques in unstructured social networks for intelligent medical diagnosis.

Finally, Mohata [MD15] published an article about web-data mining for computational intelligence, knowledge discovery, and decision-making. It reviews the possible techniques and implementations for handling big data. He considers that visualization is a tool that, has been shown to be effective for obtaining insight into big data. This includes visualization-based data-discovery logistics (such as Apache Hadoop and other technologies). Focusing on using visualization techniques to interpret big data has helped companies greatly in exploring and understanding this data easily and using it to make decisions regarding their commercial and electronic businesses.

2.1.2 Higher dimensionality reduction

There are many works in this area, such as the one published by Ljungberg [Lju17]. The researcher deals with the problems of high dimensions of text data, represented in the form of a bag of words. He analyzed the main components of a collection of texts, applied the two-dimensional reduction methods (PCA, LSA), and compared the accuracy of these two methods when classifying data. He found that using the features obtained using the PCA algorithm with the classification algorithm produced much better results than those features acquired by using the latent semantic analysis (LSA) algorithm. Thus, the PCA algorithm is efficient for dimension reduction problems in text, but it is not the only option.

In addition, Yousefpour et al. present a novel feature-reduction method in sentiment analysis [YIAH14]. All research and reports have shown the tremendous increase in the amount of data and in the world since the inception of the internet and the world wide web. The data and information on the web, such as discussions, reports, and results, in turn, have opened the door for sharing and presenting opinions and feelings about topics raised on social media. The researchers believed that with the increase of this data, its dimensions have increased to a large extent, making it difficult to interpret and necessitating special tools to extract useful information from it. The first process that researchers applied was to reduce the dimensions to get rid of unnecessary features and choose the most effective features without losing accuracy. For this purpose, they proposed a new pathway to reduce the features using standard deviation based on minimizing the dispersion of the features. They used three common classifiers: Naive Bayes, Maximum Entropy, and Support Vector Machine. Comparing the results of their research, with other reduction algorithms from previous work, they found that their proposed system led to a significant improvement in detecting and classifying feelings.

Cheng and Chen [CC19] contributed study on mining and identifying sentiments and opinions from textual data. Due to the tremendous development in electronics, people can use mobile devices everywhere and anytime without restrictions, which facilitates their access to social networking sites. This has led to an increase in complex and unstructured textual data that needs interpretation for it to become beneficial in many areas, the most important of which is assistance in decisions making. Therefore, they proposed a new method for extracting opinions from high-dimensional data based on the extraction of additional features to improve and increase accuracy and reduce processing time. This method contains four stages to reduce the dimensions and obtain a high accuracy in classifying feelings, based on using SVD and PCA algorithms. The stages are (1) preprocessing the text data classification stage, (2) extracting more features to improve the efficiency of the classification process, (3) performing SVD and PCA to reduce the dimension of text data, and (4) purposing five systems based on varying features, with or without stemming, and comparing them. The proposed system achieved better results in terms of accuracy and processing time.

Vinodhini and Chandrasekaran [VC14] used the PCA algorithm to mine for opinions and extract sentiments from text data and showed how to use it in e-commerce applications. Statistical studies have shown that e-commerce has developed very quickly over the decades. This development has led to an increase in the volume of product reviews on trade websites. The researchers here used machine-learning methods to classify these reviews as negative or positive. They investigated implementing a hybrid combination of machine-learning approaches (based on bagging and Bayesian boosting) for opinion and sentiment classification along with PCA as a feature dimension-reduction technique. When comparing the results with other methods like support vector machine (SVM) and logistic regression, the proposed method for reducing the dimensions proved more accurate in classifying using a dataset of product reviews.

Recently, in 2020, Madasu and Elango published their study of efficient feature-selection techniques for sentiment analysis [ME20]. In this scientific paper, the researchers focused on classifying the emotions expressed in text as positive or negative, depending on feelings-analysis methods. It has been proven that feature selection is the crucial process in machine learning to extract feelings and opinions from a large piece of text. This work is an investigation of the performance of different methods of feature selection to extract feelings and determine their type. Term Frequency-Inverse Document Frequency (TF-IDF) was used to reduce the dimension and extract the feature for the classification step. Then, an emotion-exploration process was performed using a set of classification algorithms, such as Logistic Regression (LR), SVMs, DT, and NB. It is also observed that FS methods using composite classifiers obtain high performance compared to neural networks.

Another paper in this area was written by Sorzano et al. [SVM14] introduce a comprehensive survey on dimensionality-reduction methods. The researchers note the significant increase in the volume of data in scientific and practical life, which may lead to the loss of useful information when not properly handled. The first way to resolve the difficulty of the increasing, large volume of data is to remove unwanted and repetitive data, and this method is called dimensional reduction. This study included several methodologies, each based on different models but all pursuing the same goal. This goal is to reduce the complexity of big data and present the same information in a more understandable way. The methods of dimensional reduction that were explained in this survey are Self-Organizing Maps, standard PCA, robust PCA, sparse PCA, kernel PCA, and so on.

Finally, Mohamed [Moh20] presented three well-known algorithms (PCA, SVD, and NMF), which were used to reduce reducing the dimensions of data from Arabic texts. The results obtained from testing the three methods showed that the PCA method produced more accurate results. PCA is able to easily recognize the structures underlying textual data for both Arabic and English documents. From the viewpoint of the researcher, the advantage of PCA is that it dictates that the main component vectors are orthogonal to each other, which was not achieved in the non-negative matrix factorization (NMF) or SVD algorithms. The type and size of data sets and reprocessing techniques may still influence the effects of dimensional-reduction algorithms. Nevertheless, this paper provides readers with some useful references for future research in the compilation of Arabic texts.

A summary of some approaches that were presented before and used dimension reduction methods with text data will be illustrated in Table 2.1.

Table 2.1 : Summary of the previous approaches used dimension reduction methods.

Authors	Summary	Modality
Ljungberg [Lju17]	Comparison between implement two methods for reducing the dimension of the bags of words. PCA vs. LSA.	collection of texts stored as bags of words
Yousefpour et al. [YIAH14]	"They used a novel method for reducing features with standard deviation based on more feature variance in the feature space."	Book and music reviews dataset
Cheng and Chen [CC19]	Proposed a sentimental text mining method based on DR methods to enhance accuracy and reduce processing time. Reducing the dimension of text data done by using SVD and PCA algorithm.	Blogs Facebook forums
Vinodhini and Chandrasekaran [VC14]	Applying a hybrid combination of machine learning approaches (bagging and bayesian boosting based) for opinion classification tied with PCA to reduce the amount of features.	Products reviews
Madasu and Elango [ME20]	TF-IDF has been used as the feature extraction technique for creating feature vocabulary and reduce the dimension.	Reviews polarity data
Sorzano et al. [SVM14]	Comprehensive survey on dimensionality reduction methods. The methods of dimensional reduction that were explained in this survey are (Self-Organizing Maps, standard PCA, robust PCA, sparse PCA, kernel PCA).	Labs text data.
Mohamed et al. [Moh20]	Presented three well-known algorithms (PCA, SVD, and NMF), which were used for the purpose of reducing the dimensions of data from Arabic texts.	Arabic text data

The problem of high-dimensional texts in this thesis will be solved by using a new method based on the SVD algorithm developed based on MI, to provide the features that have the greatest impact on the classification result. Besides, the experimentation result of our new method will be compared with re-implemented to the existing PCA and SVD on the same dataset. A full explanation of this method is detailed in Chapters 3 and 4.

2.1.3 Sentiment analysis in social media texts

Many related works have been introduced in this area; amongst these is that presented in 2009 by Go et al. [GBH09] Their work considers sentiment analysis in Twitter data in terms of the classification of feelings by using distant supervision. On this basis, they introduced a new approach automatically to categorize feelings in Twitter data as positive or negative, which is useful for consumers who want to check the quality of a product from the opinions of people before buying it, a process that many perform. It is also useful for companies to check how satisfied people are with their brands. The Twitter data the researchers relied on consisted of text messages and emoji that were used as noise. This type of data is widely available and can be obtained through automated means. Their results demonstrate that utilizing a machine-learning technique Naive Bayes, Maximum Entropy, and SVM showed 80% higher accuracy when implanted with emotional data. The main contribution here is using tweets with emojis for distant supervised learning.

Another work in this area that of Jiang et al. [JYZ+11], who categorized feelings and sentiments in Twitter tweets based on targets. The problem that the researchers considered was categorizing and identifying target-dependent sentiments in tweets. Thus, the researchers used a set of tweets that contained a target as inputs. They designed a three-step system to extract emotions from tweets, depending on the target, to obtain more features for accurate classification. A first step, the authors examined if the tweet was subjective or neutral about the target. Then, they determined the polarity of the tweet, whether it was negative or positive about the target and, by relying on graph-based optimization, identified the relevant tweets to enhance the performance. The SVM algorithm was used in each of the first two steps as a classifier, and the SVM Light 6 algorithm was used in the third step. When comparing the results obtained with the result of other studies, the proposed system added a clear and significant improvement to the target-dependent emotion-classification process.

Another contribution in this area is the work done by Mohammad et al. [MSK17]. In this paper, the authors focus on revealing the stance of the person who published the tweet: Whether he feels satisfied, dissatisfied, or neutral regarding a specific purpose or target. This purpose may be a person, organization, product, government policy, and so on. Often, people share their stance on the above topics through online posts on multiple platforms. Automatic stance detection has many advantages for information retrieval, text summarization, and textual detection. The task of stance detection is formulated as follows: Take the text of a tweet and a target and use automatic natural language systems to determine whether the person with the tweet is satisfied with the specified target or dissatisfied with the specified target. Stance detection is similar to an emotion analysis, but it has a slight and fundamental difference from it. The emotion analysis process simply determines whether a used text is positive, negative, or neutral. However, stance detection determines satisfaction with a target. In this study, the researchers conducted a careful analysis of the data and conducted several experiments to clarify the differences between feelings and stance. They show that emotion features are not useful in detecting stance as they are in predicting feelings.

Soelistio et al. [SS15a] designed a model to analyze and detect sentiments and opinion polarity from digital political newspapers by utilizing a Naive Bayes classifier approach. The system relies on three important variables: “who” is speaking, “to whom”, and “what” is said. Depending on these variables, the probability of feelings is determined and the values of these three variables are updated according to what the system has learned from the training data. The system produced impressive results when tested, and it can be used to solve other problems in sentiment analysis.

Sharma et al. [SD13a] present a sentiment-detection system utilizing boosted SVM. The development idea here is to integrate SVM with boosting to create a powerful classifier that depends on selecting the most useful features for the base learners at each boosting step. This study assumes that several SVMs can be combined and a boosting approach can be utilized to train each SVM. The sentiments were extracted based on information gained. The model was tested on movie reviews and hotel reviews, and the results obtained indicated that the model has succeeded in improving the performance of SVM. For better sentiment classification, the authors focused on boosting SVM and input feature selection. The results show that this model succeeded in enhancing the performance of SVM for sentiments analysis when the hotel and movies reviews were used.

Furthermore, Davidov et al. [DTR10] introduce a system capable of the automatic identification and detection of different sentiment types when using short text fragments from Twitter data. This system is a supervised classification based on using hashtags and emoji as training labels. The researchers built a set of feature vectors for each emotion tag in the Twitter data. This step enables the system to estimate dependencies through a variety of sentiment types. The large volume of Twitter data make it possible to identify and determine dozens of types of emotions without any strenuous training processes. The assistance of the different features in sentiment classification was evaluated and showed that the presented framework successfully identified the emotional types of untagged sentences. The quality of classifying feelings was further approved by human judges. Combining different sentiment types expressed by emoji and Twitter hashtags was also examined.

Tan et al.[TWX] presented study on text sentiment analysis. They used Amazon product reviews and at the ratings of the products given by the Amazon users. For the classification step, the authors used traditional learning algorithms and deep neural network machines. The traditional methods used by the researchers were Support Vector Machines, Naive Bayes analysis, and K-Nearest Neighbor, and they used the RNN method,a deep-learning algorithm. They found the best accuracy when they used RNN. They obtained the most accurate results when using deep-learning algorithms, compared to traditional methods.

Another work in this area is that of Shrestha and Nasoz [SN19], who presented a new model to extract sentiments from Amazon reviews depending on the product rating posted by consumers. The researchers relied on their system for converting review texts into feature vectors, utilizing paragraph vectors. They then grouped these vectors by product to train a recurrent neural networks with the gated recurrent unit. The above system showed satisfactory results in the accuracy of the classification, which was rated as 81.82% accurate.

Finally in this area, Rane and Kumar [RK18] introduced a study of opinions classification for airline companies based on what consumers shared on Twitter. Their work was based on converting texts into word vectors, based on the principle of-Doc2Vec, to train seven types of classification models and compare them to find the best in terms of classification accuracy. These models were, in order, Decision Tree, Random Forest, SVM, K-Nearest Neighbors, Logistic Regression, Gaussian Naive Bayes, and AdaBoos.

A summary of some approaches that were presented before for texts sentiments analysis methods will be illustrated in Table 2.2.

Table 2.2 : Summary of the previous approaches for text sentiment analysis methods.

Authors	**Summary**	**Modality**
Go et al. [GBH09]	Present a system to classify the feelings as positive or negative based on distant supervision and by using (Naive Bayes, Maximum Entropy, and SVM classification methods).	Tweets text with emoji
Jiang et al. [JYZ+11]	Categorizing and identifying target-dependent sentiments in tweets by using SVM and SVM Light 6 algorithms.	Tweets text with target feature
Mohammad et al. [MSK17]	Predict the stance of the person who published the tweet on whether he was a favor of, against or neutral to a specific target based on automatic natural language systems.	Tweets text with target or destination feature
Soelistio et al. [SS15a]	Design a model to analyze and detect the sentiments and opinion polarity from digital political newspapers by utilizing a naive Bayes classifier approach.	Political newspapers text data
Sharma et al. [SD13a]	Detect the sentiment by integrate SVM with Boosting to create a powerful classifier by combined several SVMs.	Twitter data
Davidov et al.[DTR10]	Introduce a system capable of automatic identification and detection of different sentiment types by estimate dependencies between different sentiment types when using short text fragments from Twitter data.	Tweets , emojis, and Hashtags data
Tan et al.[TWX]	Presented their work about text sentiment analysis by using SVM, KNN, Naive Bayes, and RNN.	Amazon reviews and products ratings
Shrestha and Nasoz [SN19]	Presenting a new model to extract sentiment from Amazon reviews by using RNN and GRN.	Amazon reviews and product ratings data
Rane and Kumar [RK18]	Detect opinion from twitter text by using SVM, KNN, Logistic Regression, Gaussian Naïve Bayes, and AdaBoost algorithms.	Tweets about airlines

In this thesis, the sentiments are extracted from the texts in two ways. The first way is based on the principle of dimension reduction, and then using residual neural network for the classification task. The second way is dependent on the use of the deep learning approach. A full explanation of these methods is detailed in Chapters 3 and 4.

2.1.4 Visualizing emotions in social media images

Visualizing emotions in social media images represents one of the most recent and most vital fields of applications in structured and unstructured big data analytics. As such, several works have been published in this area. Amongst these is the one presented by Islam and Zhang [IZ16]. They devised a method for visually judging social images by analyzing the feelings involved in these images as shown in the Figure 2.1, relying on the learning transfer strategy. In this scientific paper, the researchers proposed a unique idea using the deep network technique to analyze and explore emotions. The proposed system is shown in Figure 2.2 and Figure 2.3 To work with the deep network system, they configured it using hyperparameters to avoid the problem of overfitting. The extensive experiments on a set of Twitter data in the form of images proved that the proposed system gave more reliable and accurate results than other systems in the same field. The example of Positive and Negative images shown in Figure 2.4 and Figure 2.5. Finally, this work helped develop an in-depth understanding of related topics, such as analyzing the feelings of an image by classifying and identifying embedded emotion patterns. This work did not provide any metrics for evaluating the performance of the algorithms used.

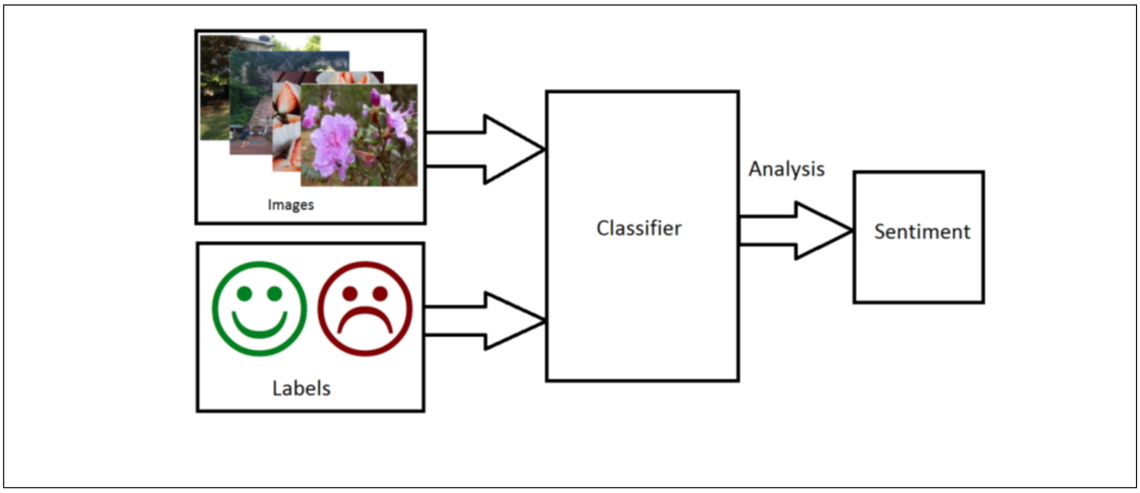


Fig. 2.1 : Diagram of a generic visual sentiment analysis framework [IZ16].

From another perspective, Jin et al. [JWLH19] published a significant work in the field of emotional visualization. They utilized a distinguished approach to learning applied on a 3D morphable face model as shown in Figure 2.6. Facial expressions were analyzed to explore this type of data, and the researchers relied on the principle of learning a transformable, 3D face model. More specifically, various facial expressions from a 3D face database were utilized to build an NMF part-based morphable 3D face model as shown in Figure 2.7. This can iteratively reconstruct a 3D face with an expression from an input image. The features that were used as inputs to the system for emotion analysis and visualization were displacement maps. Next, they used support-vector regression for emotion analysis and visualized the result relied on using 2D emotion space (VA space) mapping to share high-level information of emotion status about the test subject's from the video or image. By relying on the 3D face database, the researchers reconstructed the 3D face for the inputted video, specifically for each frame of the video. Important

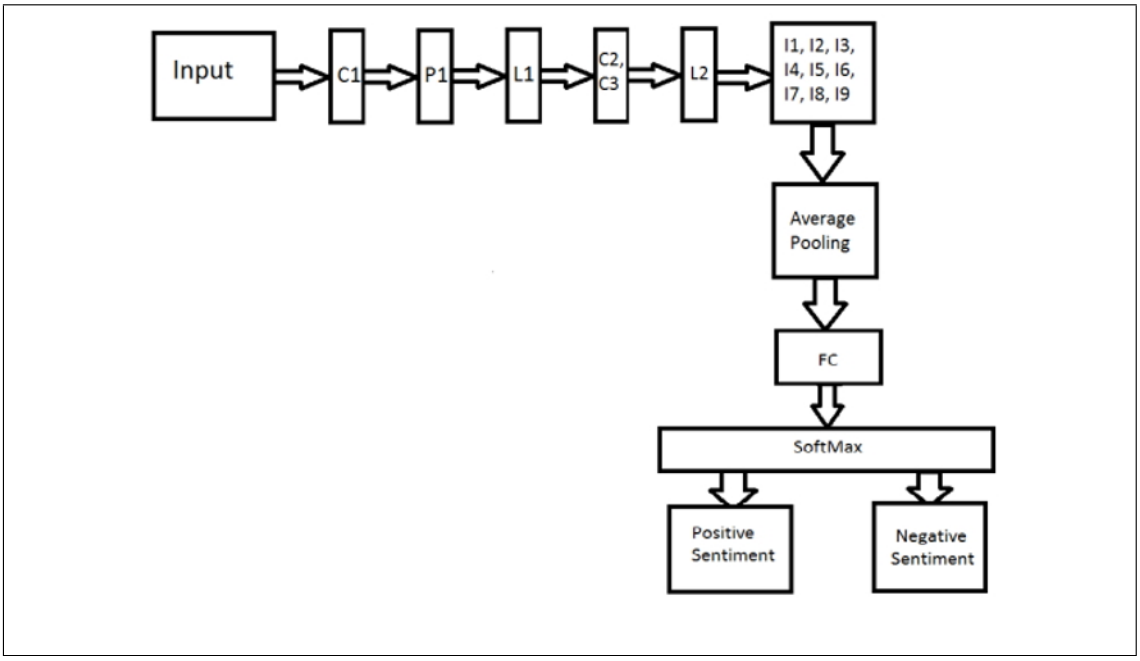


Fig. 2.2 : Illustration of proposed visual sentiment analysis framework [IZ16].

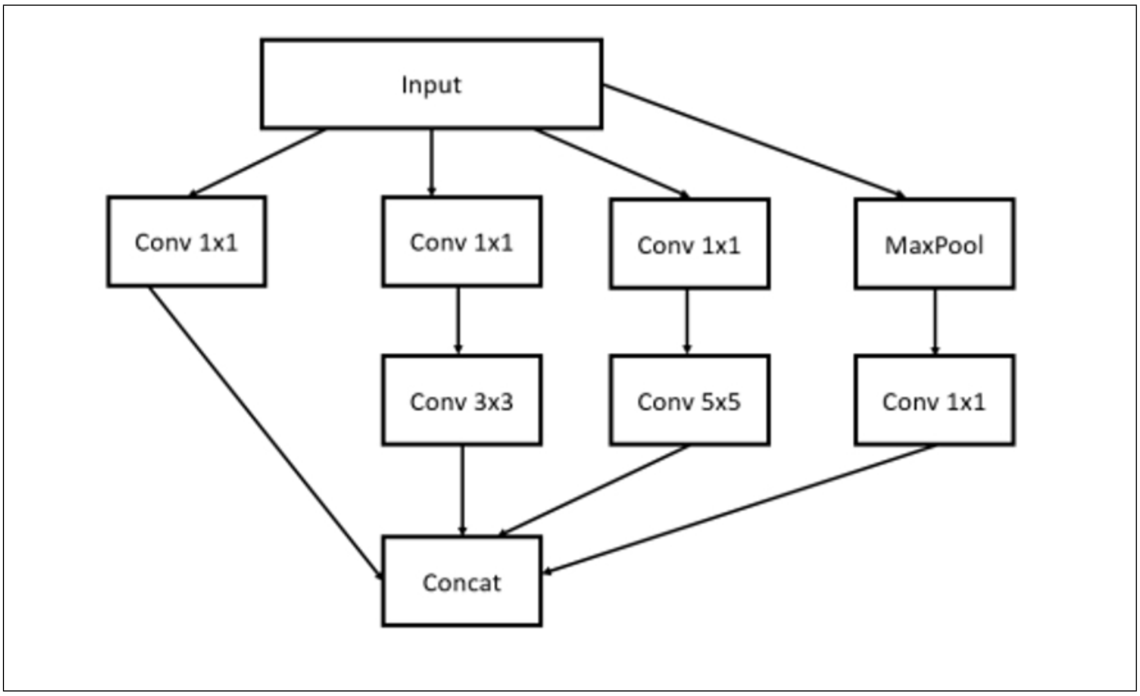


Fig. 2.3 : Illustration of an Inception module [IZ16].

features were obtained, including coefficient vectors and displacement maps as shown in Figure 2.8. These features were later utilized to train the support vector regression (SVR) algorithm in the training step. Thus, the process of extracting emotions in this research relies on the trained SVR system and the morphable face model

A robust image sentiment analysis was presented by You et al. [YLJY15]. Progressively trained and domain-transferred deep networks were used. The work focused on

Fig. 2.4 : Example of positive images in twitter dataset from training set [IZ16].

Fig. 2.5 : Example of negative images in twitter dataset from training set [IZ16].

the prediction of sentiment from visual content by creating a system based on developing machine-learning algorithms to analyze sentiment via both visual and textual features. Using this development, they obtained powerful visual features for challenging tasks, such as the task of analyzing feelings and extracting them from image data. Because of the large volume of data obtained from poorly labeled data, they used visual models as well as textual data for training to extract powerful features for sentiment analysis. In this paper, the researchers construct and develop a multi-modality regression system. The experimental results indicate that the proposed multi-modal regression model outperformed both the latest textual and visual sentiment-analysis models and the fusion models. The results obtained with the CNN algorithm were better than those of competing algorithms.

A complementary work is "Robust image sentiment analysis" by You et al. [YJ12]. The contribution of this system is the development of robust algorithms from computer vision (CNN). The main proposed inputs to this system are the visual features from the whole image or video. The authors built a sentiment-analysis machine by using an attention mechanism to discover the relevant local regions. The proposed system proved

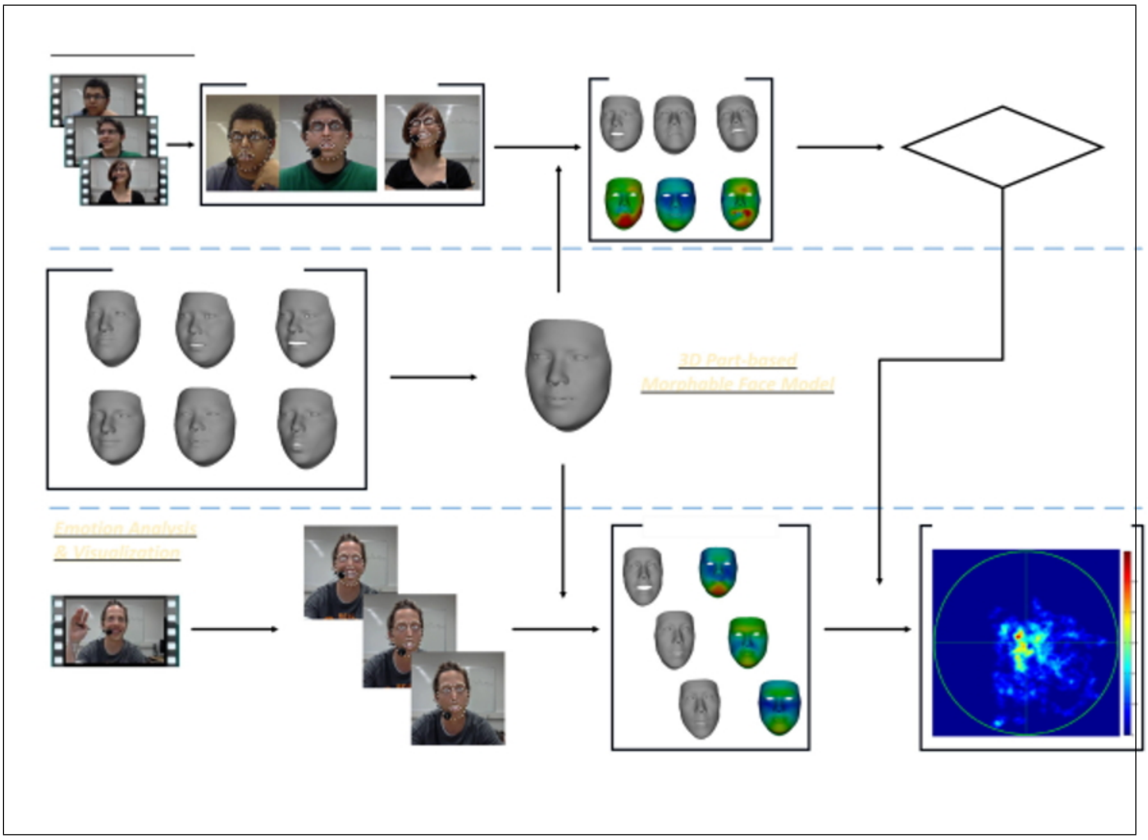


Fig. 2.6 : 3D morphable face model [JWLH19].

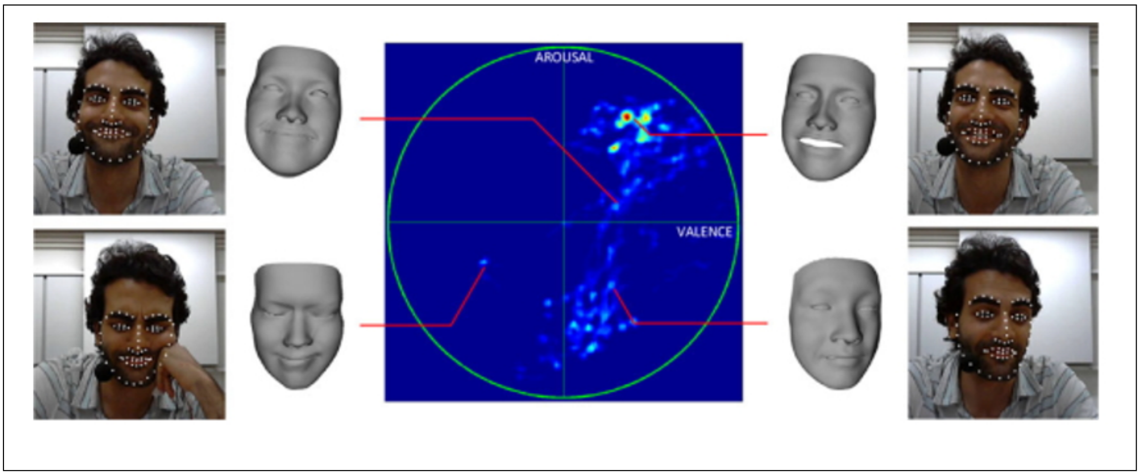


Fig. 2.7 : Interactive 3D emotion query and VA space [JWLH19].

its efficiency and ability to automatically detect sentimental local regions of video or images.

Image emotion identification based on visual sentiment analysis was used by Kanishcheva and Angelova [KA15]. The work presented an approach for analyzing sentiments and emotions in an image as shown in Figure 2.9, utilizing SentiWordNet as an external linguistic resource of sentimental words. The contribution of this work is to

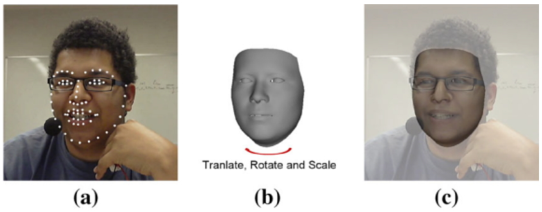


Fig. 2.8 : Feature point detection and initial alignment [JWLH19].

compute image sentiment scores using external resources. Furthermore, they performed algorithms that evaluated the emotions and polarity in a set of image tags.

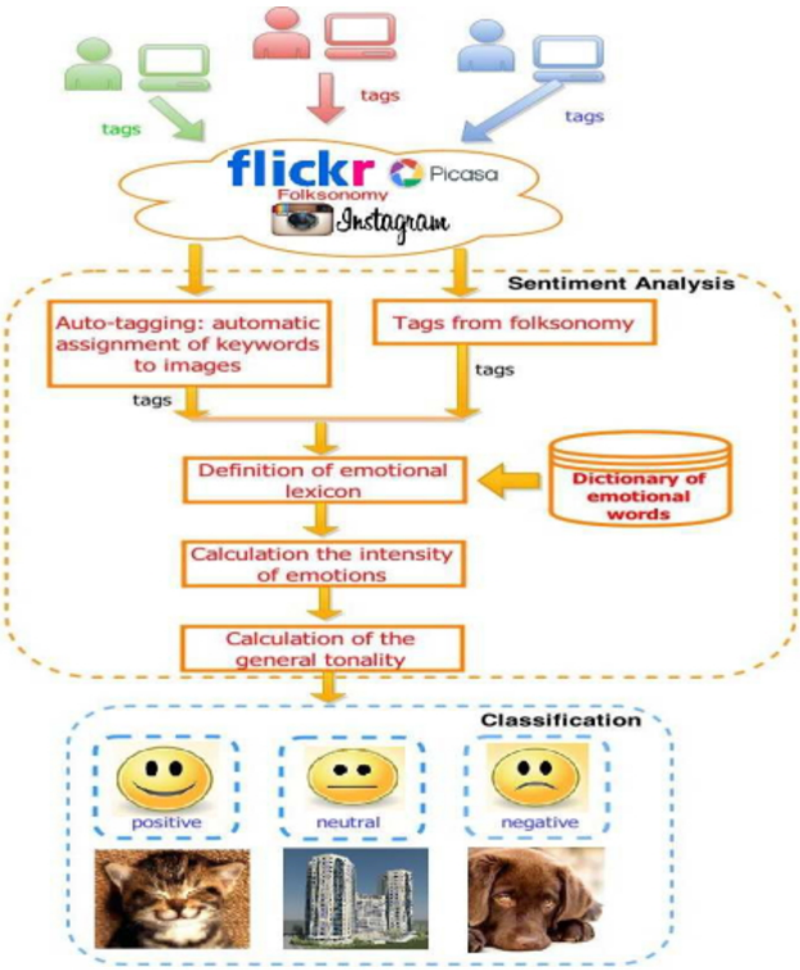


Fig. 2.9 : General scheme of the sentiwordnet approach [KA15].

A new method for sentiment analysis was proposed by Siersdorfer and Hare [SMDH10]. Their method is based on the combination of image emotion expressed in the metadata and the visual content in Flickr photographs. This work contributed to an important achievement in emotion classification. The researchers used of SentiWordNet dictionary to invigorate and extract the numerical values of the proposed system from the available metadata by focusing on the portfolio of visual words and the color distribution of the images. To take advantage of these features, the SVM light-classification algorithm was used.

Another work in this section is that of Borth et al. [BJC+13]. They presented a system to predict sentiment from visual content. They proposed a systematic, data-driven approach to create a large-scale sentiment ontology basd on psychology and web-crawled folksonomies. Moreover, they utilized SentiBank, a detector library based on the created ontology, to establish a novel mid-level representation to bridge the effective gap. Lastly, they released the concept ontology, dataset, and ANP detector library with an SVM classifier to stimulate research in this direction.

The contribution of Yuan et al. [YMYL13] has been effective in the field of visual-content analysis. They introduced a new algorithm to predict the feelings inherent in an image and applied their analysis to mid-level features. To boost their system's forecasting performance using Linear SVM, they used an asymmetric approach to deal with unbalanced data.

They depended on emotions based on the face and presented Eigen's face-based emotion-detection system. This is a simple but effective tool in detecting extremely varied facial expressions, able to deal with an image containing faces and obtain a good increase in accuracy compared to the score based on mid-level features only. This proposed system used visual content alon and did not address the textual content of the data.

Finally, Jindal and Singh's [JS15a] work in emotion prediction relied on the use of visual-content features in addition to textual-content features for data. They made significant progress using this technology by creating an emotion-classification system for images using convolutional neural network technologies. The proposed system was tested on a large body of Flickr data to identify objects and extract features for the classification step. Experiments showed the superiority of the proposed system using CNN, with impressive results compared with other systems.

A summary of some approaches that were presented before for images sentiments analysis methods will be illustrated in Table 2.3.

Table 2.3 : Summary of the previous approaches for emotions in social media images.

Authors	**Summary**	**Modality**
Islam and Zhang [IZ16]	Proposed novel framework to predict sentiment by using hyper-parameters learnt from a very deep (CNN) to initialize the network model in order to prevent overfitting.	Images from Twitter dataset
Jin et al. [JWLH19]	Emotional visualization approach. Various facial expressions from a 3D face database were utilized to build an (NMF) part-based morphable 3D face model to train a support vector regression (SVR).	Frames from videos
You et al. [YLJY15]	Prediction of sentiment from visual content by creating a system based on developing a cross-modality consistency regression model (CNN) to analyze sentiment via both visual and textual features.	Flickr images related to the 2012 United States presidential election
You et al. [YJ12]	Sentiment analysis machine by using an attention mechanism and CNN to discover the relevant local regions.	Visual features from the whole image or video
Kanishcheva and Angelova [KA15]	Approach for analysis emotions in the image utilizing SentiWordNet linguistic resource.	Flickr and Instagram image dataset
Siersdorfer and Hare [SMDH10]	System combine between image emotion expressed in image metadata and their visual content. And predict the emotion by SVM light algorithm.	Flickr image dataset
Borth et al. [BJC+13]	Utilize the SentiBank, a detector library based on the created ontology to build emotion prediction system based on SVM algorithm.	Image Twitter data
Yuan et al. [YMYL13]	Novel image sentiment prediction algorithm based on mid-level attributes and using Linear SVM.	Image Twitter data
Jindal and Singh [JS15a]	New image emotion detection framework based on CNN for object recognition.	Flickr images dataset

In this thesis, the sentiments are extracted from the images by using CNN based on an attention mechanism to get more focused on image features. A full explanation of these methods are detailed in Chapters 3 and 5.

2.1.5 Machine learning

A significant versions of machine-learning methods is the one presented by Lu et al. [LZN20] in their work classifying audio data. They introduced a new method called CABCNN. This new method relies on simple classifiers based on the attention mechanism as a selector. The use of this system demonstrated a significant reduction in the number of parameters needed by the classifier and thus helped reduce complexity. By reducing the complexity of the system, it becomes faster to train classifiers and obtain high-precision performance.

due to the rapid developments in a wide range of research areas, such as pattern recognition, classification, and signal processing. The use of neural networks and machine learning is one of the best methods for solving problems related to recognition and classification Lim et al. [LJL16]. In this work, the SER system was introduced based on CNN and RNN networks, without using manual features, to extract feelings and emotions from the speech database. The results obtained using SER systems based on neural networks showed better accuracy than traditional methods.

An appreciable work in the area of machine learning was presented by Choromanska [Cho14]. In this thesis, the author discusses a set of improved techniques for treating convex and non-convex learning problems using a machine-learning approach. The researcher focused on the learning approach based on the principle of reduction, where a single problem is divided into a group of smaller problems to find a simpler and faster solution. The author presented an improvement system based on the linked quadratic assignment. This was the first piece of research presented in this thesis and concerned the principle of reduction by relying on machine learning. It was demonstrated that the bound technique can be used in both supervised and unsupervised learning. In this thesis, Choromanska demonstrated the linear convergence rate of batch and quasi-random variables capable of solving various learning problems. This depends on the number of classes that must be counted. The second piece of research in this thesis an online multi-class classification problem, for which the researcher developed a practical, tractable logarithmic time method using a decision-tree algorithm. In the third part of the thesis, she discusses the online K-means clustering problem. She introduces the first online clustering algorithm with approximation concerning the K-means clustering objective.

From a different angle, Suero et al. [SGM$^+$19] explain the use of deep neural networks in the identification and classification of different types of music. These rely on the visual representation of the spectrum of the signal frequencies because they vary with time and as an input to the neural network. However, other techniques also use different features of music to classify and identify genres. Here, the researchers propose a unique model for a deep network that combines a CNN with a simple, multi-layered neural network to classify different types of music. Since other features are considered in the multi-layered network, the combined deep neural network showed better resolution than either single models in the experiments.

Sharma and Dey [SD12] compared different machine-learning methods used to analyze and classify sentiment and feelings. The results were tested using people's opinion data for a series of movies. The main objective of the research was to make a comparison of the methods used to select the features and classify them in terms of accuracy. The best results were for the Gain Ratio algorithm for feature selection and the SVM algorithm for emotion classification.

Kumar et al. [KGR+20] compared multiple sentiment-analysis techniques based on machine-learning methods used to analyze and classify sentiment and feelings. The results were tested using 900 users from Facebook along with the users' age and gender information. The main objective of the research was to compare the accuracy of methods used to classify emotions. Different learning methods were used in this research, and the results showed the superiority of the CNN and SVM algorithms in terms of classification accuracy.

Renault [Ren19] focused on using pre-processing methods and machine-learning techniques to evaluate the performance of a huge dataset: Too many messages posted on the StockTwits platform. To significantly improve the accuracy of the sentiment classification, the researcher added bigrams and emojis. In this scientific paper, the researcher discuss the effect of several factors on the accuracy of the sentiment-analysis system. The factors discussed are as follows:

1. Data size.

2. The complexity of the classification algorithms.

3. The effect of both investor sentiment and stock returns

The classification algorithms tested in this paper were Multinomial Naive Bayes, Maximum Entropy, Support Vector Machine, Random Forest, and Multilayer Perceptron.

A summary of some approaches that were presented before and used machine learning algorithms in the field of sentiments analysis and emotion detection will be illustrated in Table 2.4.

Table 2.4 : Summary of the previous approaches used machine-learning methods.

Authors	**Summary**	**Modality**
Lu et al. [LZN20]	They introduced a new method called (CABCNN) for classification and detection.	Audio database
Lim et al.[LJL16]	SER system was introduced based on CNN and RNN networks.	Speech database
Choromanska [Cho14]	"They solved the most difficult learning problems by developing two sophisticated machine learning tools".	Text dataset
Suero et al.[SGM+19]	Deep network that combines a CNN with a simple, multi-layered neural network for classification.	Music genre data
Sharma and Dey [SD12]	Comparison of the methods used to select the features and different machine learning methods to classify them in terms of accuracy.	Movies reviews dataset
Kumar et al. [KGR+20]	Comparison is presented between multiple sentiment analysis techniques based on machine learning methods.	Facebook dataset
Renault [Ren19]	discussed the effect of (Data size, the complexity of classification algorithms, and effect of both investor sentiment and stock returns) on the accuracy of the sentiment analysis system.	StockTwits dataset

In conclusion, the major part of the above-mentioned works in the five areas have individually, collectively, and comprehensively delineated and, elucidated the various sides of the research topic named visual analytics of unstructured big data in social media. The research project presented in this thesis is simply a humble attempt to highlight other hidden, relevant aspects of the topic. It represents a step in the direction traced by these earlier international efforts.

Chapter 3

Fundamentals

In this chapter, the background theory of different approaches for emotion and sentiment detection in unstructured social data is discussed. These include data mining methodologies for dimensionality reduction and feature selection, backpropagation neural networks (BNN), and convolution neural network (CNN) for sentiment analysis when opinion mining tweets, prediction, visualization, and deep-attention learning mechanisms in social media image sentiment classification and identification.

3.1 Data Mining and Machine Learning

Data-mining-based machine-learning approaches are computer-based approaches that automatically improve in the knowledge through experience [AZ12b]. They have become ubiquitous in recent years, in some areas rivaling or even surpassing human expertise [HK06]. Example applications include spam detection, face detection, object recognition, pattern recognition, speech recognition and translation [Gia08].
Machine learning aims to process data recognize or identify patterns and extract the information to enable a better decision to be made. It has been improving for many years, especially as approaches that work with big data become more prevalent [DD08].
This overview describes some of the most common machine-learning algorithms in use today. The discussion has broken into two sections, each with a specific theme [MA05]:

- Supervised learning: Next generation techniques: trees, networks, and rules.
- Unsupervised learning: Classical techniques: Statistics, neighborhoods, and clustering.

3.1.1 Supervised learning

Supervised learning is one the most important approaches in machine learning. It has been designed to use labelled and analyzed data based on a teacher or helper approach.

is illustrated in Figure 3.1, the observed training data (labeled) is used by a supervised learning algorithm to amylaze and generate learning rules. Then, the decider, which is called the “classifier,” uses the generated rules to “classify” the (unlabeled)test data (unlabelled) based on what has been learned so far [Gia08, HK06, MA05].

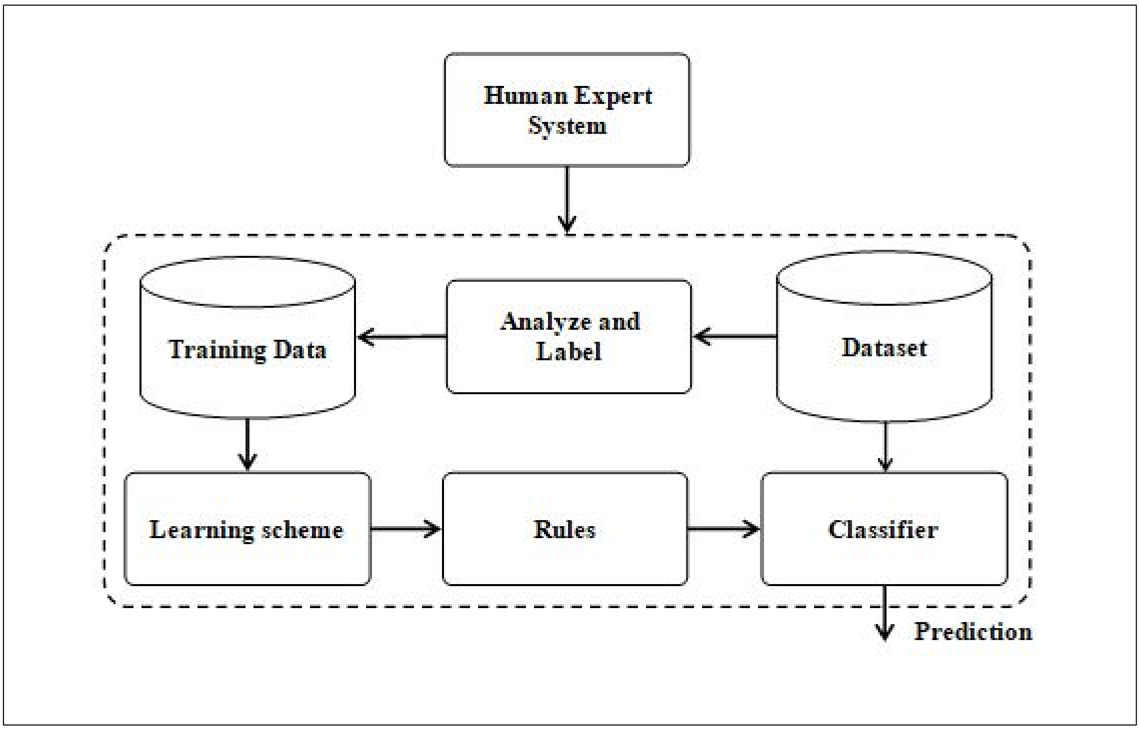


Fig. 3.1 : Supervised learning system with labeled data [MA05].

3.1.2 Unsupervised learning

Unlike the supervised learning approach, unsupervised learning in machine learning is based on a different strategy which called “clustering”. An unsupervised learning approach does not decide which data point is “good” and which is “bad” but groups the “similar” datapoints in one class and those considered the “other” in a different one. An unsupervised learning approach mainly works directly without a training phase on unlabeled data, and it is sometimes used to create a labeled dataset to be used for a supervised learning approach, as shown in Figure 3.2 [LL14, MA05].

This section focuses on supervised learning. The task is to predict the target variably given data X. The function $f : X \to y$ is generally unknown and needs to be estimated based on training examples [CPS98].

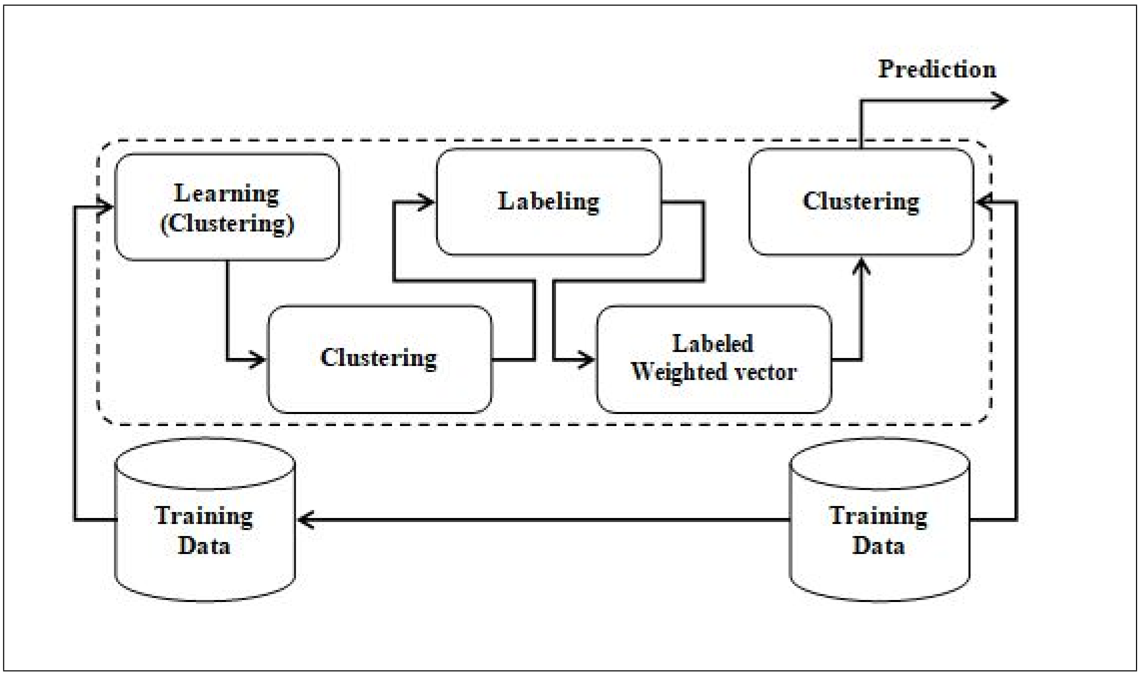


Fig. 3.2 : Unsupervised learning system with unlabeled data [MA05].

3.2 Machine Learning and Sentiment Analysis

The learning task in machine learning can be in several forms, such as supervised, unsupervised, or semi-supervised learning. If the computer is presented with input example, including its desired outputs (which are provided by a teacher in this case) it is an example of supervised learning. A popular example of supervised learning is filtering input messages as spam or ham message. In the case of email filtering, a supervised learning algorithm is presented with email messages that are labeled as example of spam or not spam. Therefor machine learning implies a computer system (program) that is able to predict and correctly label a new input message as either spam or not spam [SEZS00]. Text mining using machine learning initially processed many documents that have been gathered. In other words, the text mining using machine-learning and the tool that is mainly used to extract information or features form the documents and process them [CPS98]. The main stage of the text mining using machine learning is the text analysis or pre-processing step. In this step, various techniques are repeatedly used until relevant information is extracted from the processed documents [CPS98, LL14]. Machine-learning approaches or tools organize the documents or data structures from the database only once. In contrast at text-mining approach using machine learning extracts some information from the semi-structured and structured datasets, such as e-mails, text, and HTML files [SEZS00]. However, using machine-learning tools and approaches is the best option to organize and handle online data [SEZS00]. A high-level general approach for text mining using machine learning is illustrated in Figure 3.3. In general, opinion mining follows the general approach that is shown in a high-level abstraction in Figure 3.3. Opinion mining using machine learning consists of many tasks and functions, including the following:

- Opinion clustering.
- Opinion concept or entity extraction.
- Opinion summarization.
- Opinion classification.

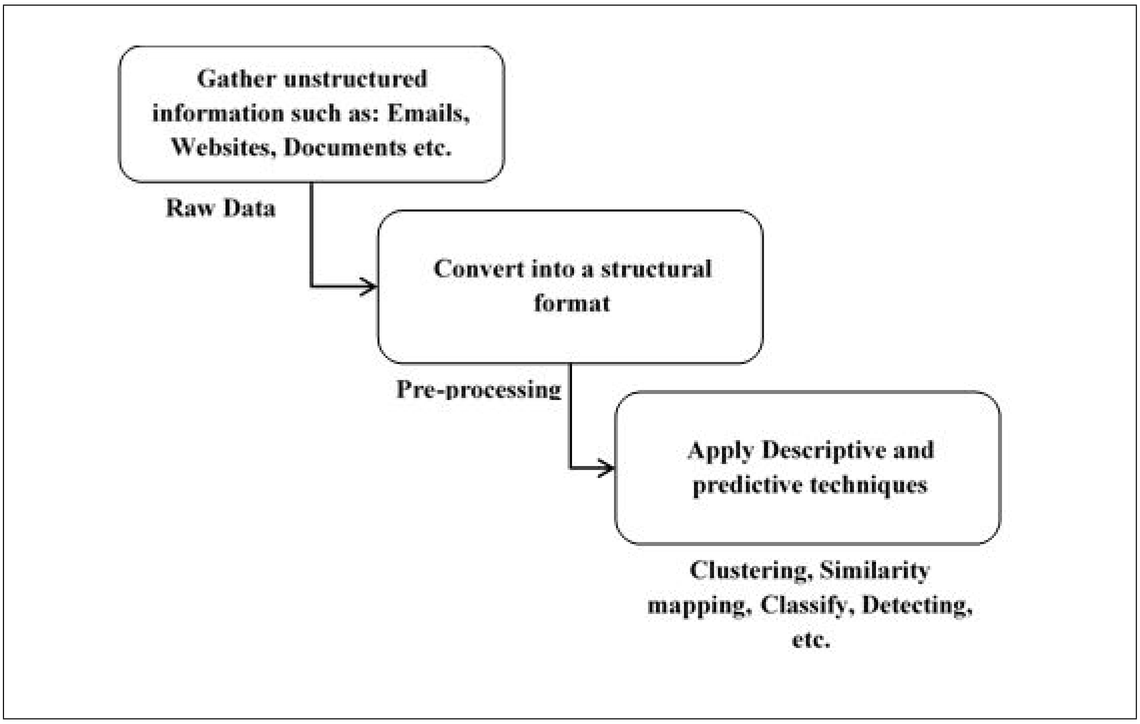


Fig. 3.3 : A high level of text mining general approach using machine learning [CPS98].

Opinion clustering Opinion clustering, or document clustering, is a procedure that is used to analyze the clusters in textual documents. In other words, it is a machine-learning program that is able to perform topic extraction, document association, and fast information recovery [AZ12a]. Opinion clustering can also be designed in two types: Online and offline text-clustering system. An online clustering system is generally controlled by the efficient problem as compared to the offline one [Swa16].
Opinion concept of entity extraction Concept or entity extraction is defined as a subtask of machine learning that aims to organize the elements in a text. In other words, it is a process that reorganizes text elements into pre-defined categories, which presents the names of the person, organization, area, and so on [AZ12a].
Opinion summarization Text summarization or automatic summarization is a text-processing procedure that reduces a document from the data into a synopsis form. This

process contains three main points from of the real document [Swa16]. Text summarization plays a vital role in text processing using machine learning, for which the growth of online text or data is rapid. Examples of text summarization technologies a typical search engine (such as Google), Facebook, and Twitter [Agg15].

3.3 Data Mining Dimensionality Reduction and Feature Extraction Models

In data-mining and machine-learning application, such as sentiment-data analysis, different important techniques are mainly such as dimensionality data reduction and feature extraction. One of the most mathematical applications is the dimensionality reduction of data features. Dimensionality reduction and feature extraction are defined as generalizations of the inverse data matrix [Moo20]. One of the most widely used types of data mining dimensionality reduction and feature extraction is SVD [BH11, CEHM14, DL18], which was described by [FZ09, HB14, Rak97]. Earlier, in 1903 [Wat04] a key concept of the matrix pseudoinverse using the SVD of integral operations was introduced. At the time, Fredholm [Fre03] referred to a matrix that was dependent on the generalization aspect term of the pseudoinverse without being based on a further specification. The feature-extraction approach in machine learning is defined as an approach that is able to provide a set of information, called “extracted features,” to the classifier in supervised or unsupervised learning to classify or cluster them. In brief, the process of feature extraction is based on reducing the cost of training and clustering by providing a set or sub-set of features for the candidate data points to provide a better indication for them and reduce the dimensionality of the problem (datapoints) [Agg16].

3.3.1 Singular value decomposition (SVD)

SVD is the most most widespread unsupervised data-mining algorithm. It is such a significant algorithm that is mainly used for higher dimensionality data (feature space) projection. It is also, one of the most appropriate mapping tools for mapping higher dimensionality data space or (vector space) or to another dimension. Moreover, (SVD) is the most useful method for analyzing and mapping data (feature vector space) in one dimension (one vector space) onto another space such as a higher dimensionality space (with different dimensions) [KFD+07].

Most linear equations simulation systems rely on (SVD) to analyze and map data space. singular value decomposition allows the linear equations simulation systems to represent or extract any metric that may be simply omitted. In general, the SVD approach is based on the metric representation, “an approximate matrix representation” [Hui04]. Of course,

the fewer the dimensions that are chosen, the less accurate will be the approximation. singular value decomposition is a technical use of a number application that includes the analysis technique of two-way variables (tables) evaluation. Although, SVD in experimental design, empirical fitting of any function, and regression [Hui04].
However, SVD is designed based on defining a set of numbers that connects rows and columns together. It is used to reduce the original dimensionality of a matrix m by extracting an approximation, a small number "rows" and "column" and connecting them through other components [Abn07]. For example, X is a matrix of data that has the dimensionality $m \times n$ and the rank of the matrix X is r. It can be assumed that rank of the data matrix is defined as the largest component "number" in the either rows or columns. In this case, the component "number" in the matrix that is a non-zero linear combination in the vector 0 (a set of independent "rows" and "column") my be chosen. Then, a matrix components, such as U, Σ, and V may be extracted as shown in Figure 3.4 [Alp09].

1. The column-orthonormal matrix $x \times r$ is extracted from each column and represents the unit vector and the dot products of any two columns, which are 0.

2. V is another component that is extracted that also has an $n \times r$ column-orthonormal matrix in a transposed form, called V, so it is the rows of V^T that are orthonormal.

3. Finally, the singular values of X, called the "diagonal matrix Σ" combine all the elements not on the main diagonal that are 0.

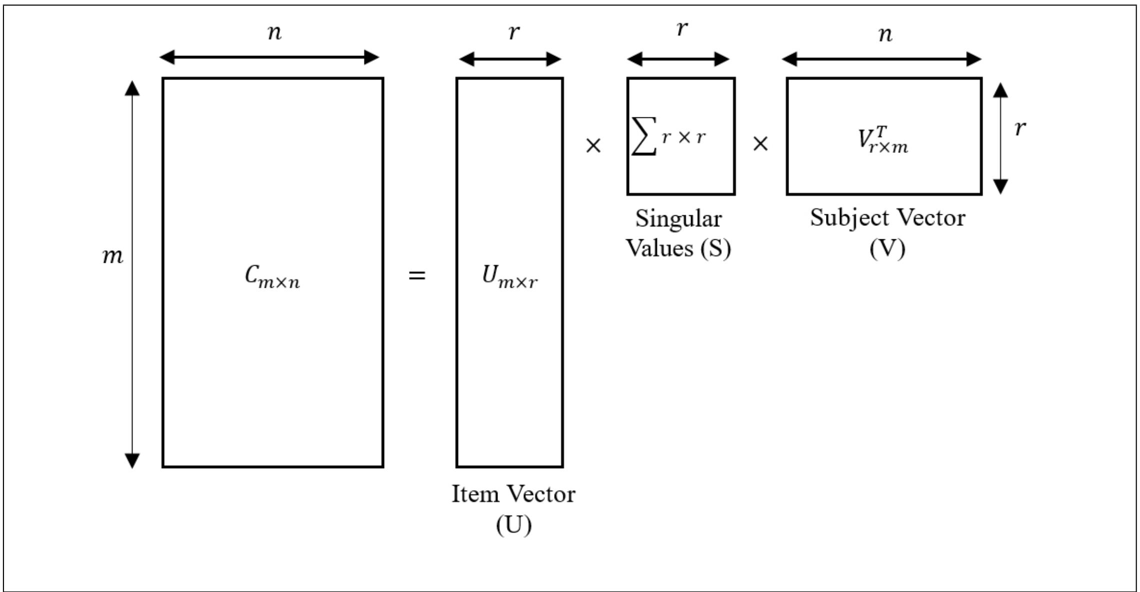


Fig. 3.4 : The form of SVD [Abn07].

singular value decomposition is a form of Eigenvalue or Eigen-vector mechanics, which is a similar process of finding the singular value (Eigen-vector). However, it is used to find the corresponding singular vectors (Eigen-vector) that are mainly yielded by matrix decomposition term. This term is a more general and flexible matrix decomposition factorization. The terms singular vector and Eigen vector used interchangeably. The SVD of matrix A can be written as shown in Equation (3.1) [Alp09].

$$A = U\Sigma V^T \tag{3.1}$$

Here, U is the orthogonal $m \times m$ matrix and U is the column of Eigen-vectors that is illustrated by AA^T. In another case, V is a $n \times n$ matrix that represents the orthogonality of the columns of V. However, S is the diagonal Eigenvalues (entities), also called the diagonal sigma values $\sigma_1, ..., \sigma_2$ which are computed based on the orthogonal matrices AA^T and A^TA that calculated by the square roots of the non-zero Eigenvalues. In this case, both are the singular value of the matrix A, which fills the first rank place r on the main diagonal orthogonality matrix S. However, S is an r matrix that is defined and written as Equation (3.2) demonstrates [Abn07].

$$AA^T = (USV^T)(VSV^T) = USS^TU^T \tag{3.2}$$

similarly, A^TA can be written as Equation (3.3) shows [Abn07].

$$A^TA = (USV^T)^T(USV^T) = USS^TU^T \tag{3.3}$$

By relying on Equation (3.2), the U "Eigen-vectors" matrix AA^T and the "Eigen-value matrix" SS^T are defined as the $m \times m$ matrix with the Eigenvalues. However, based on the same method and using Equation (3.3), the S^TS has exactly the property $\lambda_1 = \sigma_1^2, ..., \lambda_r = \sigma_r^2$. which is also is defined as the $n \times n$ matrix [Abn07, SSBD14]. Based on the same method and using Equation (3.3). U is defined as the A^TA, where the SS^T is the Eigenvalue matrix that is placed in the middle and defined as the $m \times m$ matrix with the Eigenvalues $\lambda_1 = \sigma_1^2, ..., \lambda_r = \sigma_r^2$ [Abn07, SSBD14].

Singular value decomposition Inverted Matrix Pseudo-Inverse Computation

Mathematically, for any square matrix A of size $n \times n$, the matrix inverse exists if the matrix A has a non-singular rank. In other words, the matrix A with a non-singular rank is shown below in Equation (3.4) [BIG03]:

$$rank(A) = n \tag{3.4}$$

In this case, may be assumed that the inverse matrix A is A^{-1}, which is equivalent to the squared matrix with full rank that can be inverted as shown in the following formula that is defined in Equation (3.5) [BIG03]:

$$AA^{-1} = A^{-1}A = I_n \tag{3.5}$$

This means only the squared matrix with full rank can be inverted. For a general rectangular matrix A with size $n \times n$ with the deficient rank matrix (matrix-inverse) is mainly used as a generalization mathematical approach for matrix inverse [BWZ16]. The Moore-Penrose matrix inverse is one of the most widely used pseudo-inverses, and it is mathematically defined by Equation (3.6) [BWZ16]:

$$A = V_A \Sigma_A^{-1} U_A^T \tag{3.6}$$

In this case, let A be any matrix with size $m \times n$ with the rank p matrix, then:

$$AA^T = U_A \Sigma_A \underbrace{V_A^T V_A}_{=I_p} \Sigma_A U_A^T = \underbrace{U_A}_{m\times p} \underbrace{U_A^T}_{p\times m} \tag{3.7}$$

Where, AA^T presents an orthogonal projection. In general, for any matrix B the main formula of the Moore-Penrose matrix inverse is defined by Equation (3.8) [BWZ16]:

$$AA^T = U_A U_A^T B \tag{3.8}$$

Generally, the pseudo inverse of the matrix A can be written as Equation (3.9) [Hou58]:

$$A^+ = A^+ A A^+ \tag{3.9}$$

Then, the pseudo inverse of matrix A is defined in Equation (3.10) and (3.11) [Hou58]:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{the pseudoinverse is} \quad A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \tag{3.10}$$

Indeed

$$AA^+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{3.11}$$

And thus,

$$AA^+A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = A \tag{3.12}$$

Similarly,

$$A^+A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{3.13}$$

$$A^{+}AA^{+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} = A^{+} \tag{3.14}$$

For

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} A^{+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \tag{3.15}$$

For

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} A^{+} = \begin{bmatrix} \frac{1}{5} & \frac{1}{2} \\ 0 & 0 \end{bmatrix} \tag{3.16}$$

Then the denominators are

$$5 = 1^2 + 2^2 \tag{3.17}$$

Non-inverted matrix pseudo-inverse computation In this part we discuss such a different situation where the matrix cannot be inverted because it is singular. In this case, we use the SVD to get the pseudoinverse of this matrix. To compute the (SVD) and the pseudoinverse of the non-inverted matrix, first, need to define such a complex matrix (non-inverted) where it is defined as a matrix A that has a complex dimension such as $m \times n$. In this case, assume that the convenient matrix decomposition should be defined. We assume that the matrix dimension is $m \times n$.
In this case, let A be defined as any matrix with $m \times n$ that contains complex elements. In this case, the matrix A can be decomposed as the following formula that is described in Equation (3.18) [Mor60]:

$$A = PJQ^{*} \tag{3.18}$$

Where here P and Q are the unitary matrix and also J is defined as a matrix with $m \times n$ dimensions which is called a (bidiagonal matrix) and it is mathematically described in the following form Equation (3.19):

$$J = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \alpha_2 & \beta_2 & 0 & & & \vdots \\ \vdots & & \ddots & & & & \vdots \\ \vdots & & & \ddots & & & \vdots \\ \vdots & & & & \ddots & & \vdots \\ \vdots & & & & & \ddots & \beta_{n-1} \\ 0 & \dots & \dots & \dots & \dots & \dots & \alpha_n \end{bmatrix} \tag{3.19}$$

In this case, to proof that, let assume that $A = A^{(1)}$ and also let's assume that $A^{(\frac{3}{2})}$, $A^{(3)}, \dots,$ $A^{(n)}$, $A^{(\frac{(n+1)}{2})}$ which will be defined as the following formula in Equation

(3.20) based on Householder transformation [AB09, GK65, Gra76].

$$A^{(\frac{(k+1)}{2})} = P^{(k)} A^{(k)} \tag{3.20}$$

Where $k = 1,\ 2, \ldots,\ n$, and also be defined as in the following Equation (3.21):

$$A^{(\frac{(k+1)}{2})} = P^{(k)} A^{(k)} \tag{3.21}$$

Where $k = 1, 2, \ldots, n-1$. Also, $P^{(k)}$ is Hermitian and also unitary matrices which is defined in the following form that is described in Equation (3.22) [CKNP17]:

$$P^{(k)} = I - 2x^{(k)} x^{(k)}, \quad x^{(k)} \times x^{(k)} = 1 \tag{3.22}$$

Then, the unitary transformation $P^{(k)}$ is determine as it is defined by the Equation (3.23) [CKNP17, Gra76]:

$$\alpha_{i,k}^{(\frac{k+1}{2})} = 0 \quad \text{were} \quad i = k+1, \ldots,\ m \tag{3.23}$$

And then $A^{(k+1)}$ has the mathematical form Equation (3.24):

$$A^{(k+1)} = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & \alpha_2 & \beta_2 & 0 & & & & & \vdots \\ \vdots & 0 & \ddots & & & & & & \vdots \\ \vdots & & & \ddots & & & & & \vdots \\ \vdots & & & \alpha_k & \beta_k & & & & \vdots \\ \vdots & & & & x & x & & & \vdots \\ \vdots & & & & x & x & \ddots & & \vdots \\ \vdots & & & & & & & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ddots \end{bmatrix} \tag{3.24}$$

So, in this case, to distribute the x elements, the new set has been defined in Equation (3.25) [CKNP17]:

$$x_i^{(k)} = 0, \quad \text{were} \quad i = 1, 2, \cdots, k-1 \tag{3.25}$$

In this situation, since $P^{(k)}$ is a unitary transformation matrix, then the length is preserved and called consequently as it is defined in Equation (3.26) [CKNP17]:

$$|\alpha_k|^2 = \sum_{i=k}^{m} |\alpha_{i,k}^{(k)}|^2 \tag{3.26}$$

Although, since $P^{(k)}$ is a Hermitian, then

$$P^{(k)}A^{(\frac{k+1}{2})} = A^{(k)} \tag{3.27}$$

That is equivalent to [CKNP17]:

$$(1-2|x_k^{(k)}|^2)\alpha_k = \alpha_{k,k}^{(k)} - 2x_i^{(k)}x_k^{(k)}\alpha_k = \alpha_{k,k}^{(k)} \quad \text{where} \quad i = k+1, \cdots, m \tag{3.28}$$

By based on Equations (3.26), (3.27), and (3.28) the possible $x^{(k)}$ is defined. In this case, as a summarization in Equation (3.29) [CKNP17]:

$$A^{(\frac{k+1}{2})} = A^{(k)} - x^{(k)}.2(x^{(k)} \times A^{(k)}) \tag{3.29}$$

With the following [CKNP17, Gra76]:

$$s_k = \left(\sum_{i=k}^{m} |\alpha_{i,k}^{(k)}|^2\right)^{\frac{1}{2}} \tag{3.30}$$

$$\alpha_k = -s_k\left(\frac{\alpha_{i,k}^{(k)}}{|\alpha_{i,k}^{(k)}|}\right) \tag{3.31}$$

$$x_k^{(k)} = \left[\frac{1}{2}\left(1+\frac{|\alpha_{i,k}^{(k)}|}{s_k}\right)\right]^{\frac{1}{2}} \tag{3.32}$$

$$c_k = \left(2s_k\frac{\alpha_{i,k}^{(k)}}{|\alpha_{i,k}^{(k)}|}x_k^{(k)}\right)^{-1} \tag{3.33}$$

$$x_i^{(k)} = c_k\alpha_{i,k}^{(k)} \quad \text{for} \quad i > k \tag{3.34}$$

Then the final formula will be described in the Equation (3.35) [CKNP17]:

$$A^{(k+1)} = A^{(\frac{k+1}{2})} - 2(A^{(k+\frac{1}{2})}y^{(k)}).y^{(k)} \tag{3.35}$$

In case of the non-invertible matrix is defined as a matrix that has one side (left or right is invertible). In other word, no-square matrix of full rank is a matrix that has several one-side inverses. For instance [CKNP17, Gra76]:

- A is matrix with size $m \times n$ where $m > n$ in this case, we have a left inverse that is define as the following Equation (3.36) shows:

$$\underbrace{(A^T A)^{-1} A^T}_{A^{-1}_{left}} A = I_n \tag{3.36}$$

- A is matrix with size $m \times n$ where $m < n$ in this case, we have a right inverse that is define as the following Equation (3.37) shows:

$$I_m = \underbrace{(A^T A)^{-1} A^T A}_{A^{-1}_{left}} \tag{3.37}$$

In this case, the left inverted side can be used to determine the least norm solution. That means it is originally used as a least-square formula for the regression matrix, which has any even-one-inverted-side. Although, the Pseudo-inverse approach using SVD can be used for both sides either left or right to find the existing invested side. For such an example, let's assume and consider that an inverted matrix A is defined as is given in Equations (3.38)-(3.44) [CKNP17]:

$$A^{(2\times 3)} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \tag{3.38}$$

So, in this case, $m < n$, and for this reason we have right inverse, which is defined as [CKNP17, Gra76]:

$$A^{-1}_{\text{right}} = A^T (AA^T)^{-1} \tag{3.39}$$

Based on that and by computing the matrix component, we have [CKNP17]:

$$AA^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix} \tag{3.40}$$

$$(AA^T)^{-1} = \begin{bmatrix} 14 & 32 \\ 32 & 77 \end{bmatrix}^{-1} = \frac{1}{54} \begin{bmatrix} 77 & -32 \\ -32 & 14 \end{bmatrix} \tag{3.41}$$

$$A^T (AA^T)^{-1} = \frac{1}{54} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \times \begin{bmatrix} 77 & -32 \\ -32 & 14 \end{bmatrix} \tag{3.42}$$

$$= \frac{1}{18}\begin{bmatrix} -17 & 8 \\ -2 & 2 \\ 13 & -4 \end{bmatrix} = A^{-1}_{right} \tag{3.43}$$

While the left side does not exist because:

$$A^T A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 17 & 22 & 27 \\ 22 & 29 & 36 \\ 27 & 36 & 45 \end{bmatrix} \tag{3.44}$$

Which shows that the singular matrix cannot be inverted. SVD algorithm steps are described in the Algorithm refsvd [GPW10, RPG06].

3.3.2 Principal component analysis (PCA)

PCA is a procedure that is used for feature mining “extraction”. Data used in intrusion detection problems are high-dimensionality data. It is desirable to omit and reduce the dimensionality of the data for easy exploration and analysis [GPW10, RPG06]. The PCA is often used for this purpose, as shown by Algorithm A.2, which is described in Appendix A.

3.3.3 Mutual information (MI)

One of the most important statistical approaches in machine learning is mutual information (MI). Mutual information measures and extracts the relationship between two random variables in the data that are sampled simultaneously. Essentially, MI measures the variables randomness. In another words, it measures how much information can be extracted from random variables which provides an indication of how much each is related to other. For instance, the MI I $(A : B)$ measures the information about A that is liked by [ZB16a]. For example, if A and B are independent, then A includes no information about B, MI is zero. Moreover, if we assume that both A and B are similar in this case A is carrying all the information that B has. In another word, all the information in carried by variable A and is shared with variable B. More specifically, knowing A reveals nothing new about B, so their MI is zero. In this case, If A and B are the same variables, then all the information is carried by the variable A which is shared with variable B: knowing A reveals nothing new about B, the MI is the same as the information carried by A or B alone, namely the entropy of A. In a particular sense, MI which is one of the most important technique that is using to measure the specific amount of variables that are similar or can be retrieve by of the another [ZB16b]. The official definition of the MI between two random variables is defined as a joint distribution between two random variables, the first one is X and second is Y as demonstrated by Algorithm A.3 [ZB16b] which is described in Appendix A.

3.4 Data Mining Prediction and Classification Models.

Data mining has many different prediction and classification models, such as artificial neural networks (ANN), and a deep learning model (DL). essentially, it is mainly based on running the main approach of infer or predicts the learning function on the training data and the desired data points, called the "prediction function" or training task. The infer function is used later on the test data to predict the target data point, which is called the classification task [VPZ13, SHK$^+$14].

3.4.1 Artificial neural network prediction model

The biological neural network, which is the main natural example of a neural network, has approximately 10^4 neurons. These neurons are connected via 100^6 intersections (a trillion connections and intersections) in the human brain. The neural networks' (NNs) neurons process information and data. The neurons are connected by synapses; the synapses have a set of variables, such as weight. Therefor, the NN is a parallel distributed processing system [Bis06, MKA11, SO03]. Figure 3.5 shows biological neurons.

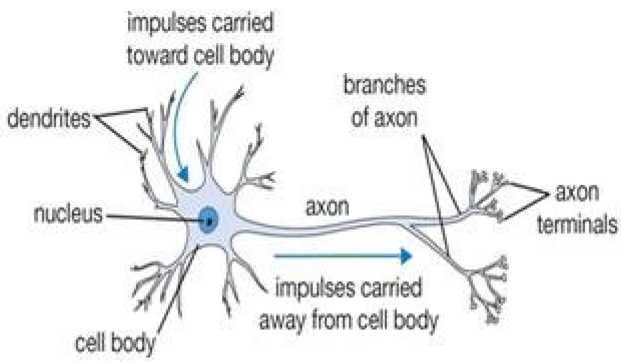


Fig. 3.5 : Biological neuron [Kan11].

Basic concepts of artificial neural networks The main structure of the ANN or the “artificial neural networks” based on analogous to the biological neuron is defined as the process the highly interconnected. In this case, the processing is passing through the weighted links from one neuron to another are the connections between neurons. Each neuron receives several input signals, which are transmitted via the neuron's outgoing connections. The outgoing connection is divided into many branches that transmit the same signal. The neuron is connected by a link, has a specific numerical weight that is associated and called the long-term memory in ANN [Kan11]. An ANN is a computational system designed to at least mimic intelligent behavior. A neural network system has successfully been developed to solve capacity planning, pattern recognition, robotics, business intelligence, and intuitive problems. In the last few years, neural networks have accelerated their progress in computer science, such as in forecasting data analysis, and data mining [Bis06, Kan11]. Figure 3.6 shows the basic structure of a neuron with 'n' input.

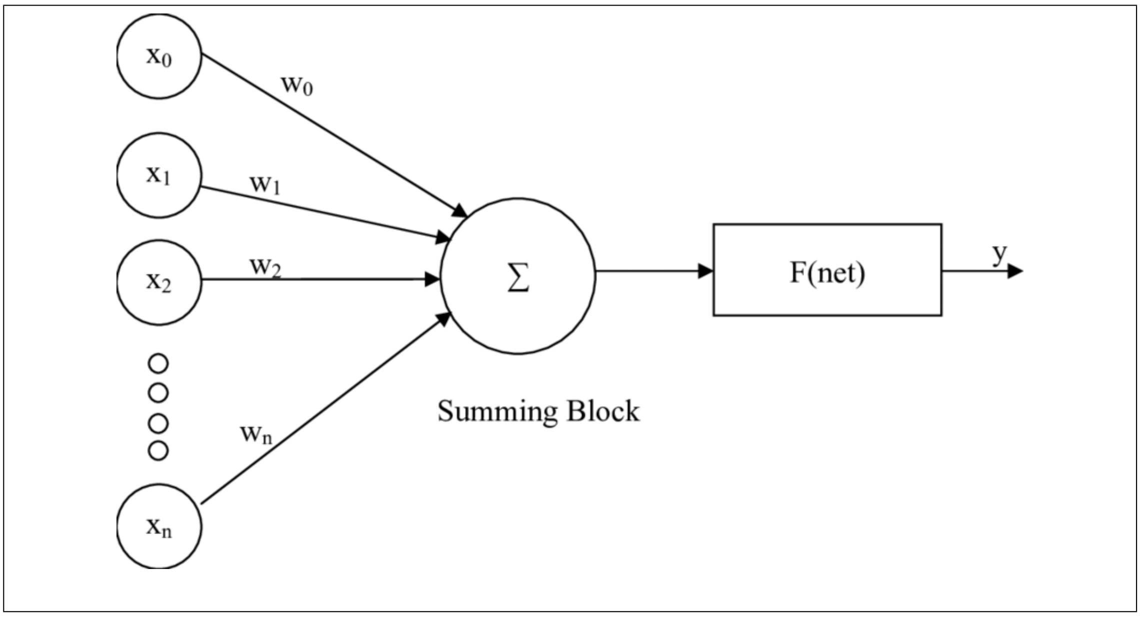


Fig. 3.6 : A basic neuron model [Kan11].

The following equations (3.45) and (3.46) described the function of the neuron [SO03].

$$y_j = f_j(x) \tag{3.45}$$

and

$$x = \sum_{n}^{n} P_i W_{ij} + b_j \tag{3.46}$$

where,

p_i the summation of all signals from each connection

i^{th} the inputs of the system

w_{ij} is the weight

b_j is the Bias

f_j a transfer functions

Artificial neural network model Neurons are the units of information processing that are essential for the main "operations" of the ANN [Bis06]. Figure 3.7, illustrates the main model of nerve cells, which form the basic design of the ANN Model. In this case, there are three fundamental elements of the model neurons [SO03]:

1. A set of connections called "connections links."
2. Each connection link is weighted by a different cost "weight" (w).
3. The signal from each connection (x_j) at each input of synapse j that is connected by a specific neuron (k) is later multiplied by each synaptic weight (w_{kj}).
4. A summation, called "adder," is used to sum the input signals which are weighted by the respective synapses "connection" of the neuron.
5. A prediction function, called "activation function," is used to amplify the output signal of each neuron [Bis06].

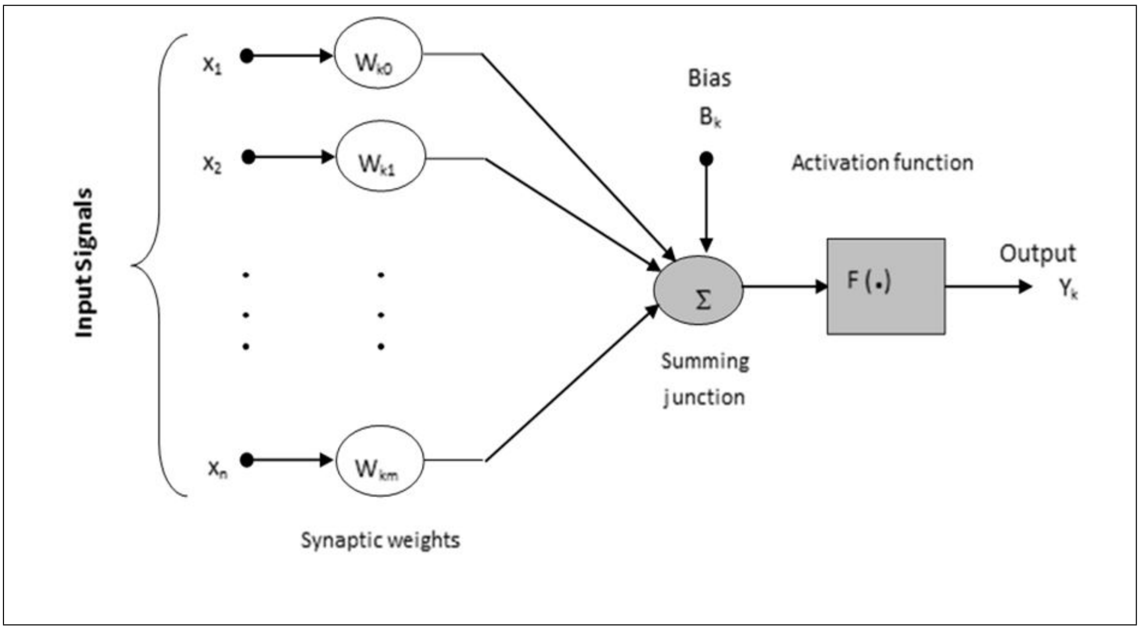


Fig. 3.7 : A neural network model [SO03].

As shown in Figure 3.7, the neuronal model can be externally effected by another term called "Bias," which is denoted by (b_k). Bias (b_k) has different effects on the ANN model. It essentially increase or lower the net input for each neuron, based on its weight [SHK+14]. For each neuron k the mathematical model can be described by Equations (3.47) and (3.48) [Bis06, SO03]:

$$u_k = \langle w, x \rangle = \sum_{i=1}^{n} w_i \cdot x_i \tag{3.47}$$

and

$$y = f[\langle w, x \rangle] = f\left(\sum_{i=0}^{n} w_i \cdot x_i\right) \tag{3.48}$$

where in this case x_i are the input signals, w_{kj} is the synaptic weight of the neuron k, u_k . Fundamentally, combines the output and the input signals. Finally, the activation function $f(u_k)$ activates the function y_k, which is the output signal of the neuron [MKA11, Bis06].

Activation functions

The basic process of artificial neurons (unit) contains a summary of likely incoming neuron signals and produces the output signal by applying the activation function to the network (weighted signal input) [MKA11]. Figures 3.8, 3.9, 3.10 show some of the most common activation functions [Bis06, MKA11, SO03].

1. Sigmoid function

The sigmoid function is the activation function that has been most widely used in machine learning (supervised learning). It is used especially for logistic regression and some other basic neural network designs [Kan11]. As shown in Figure 3.8, the sigmoid function is not the only activation function that is chosen.

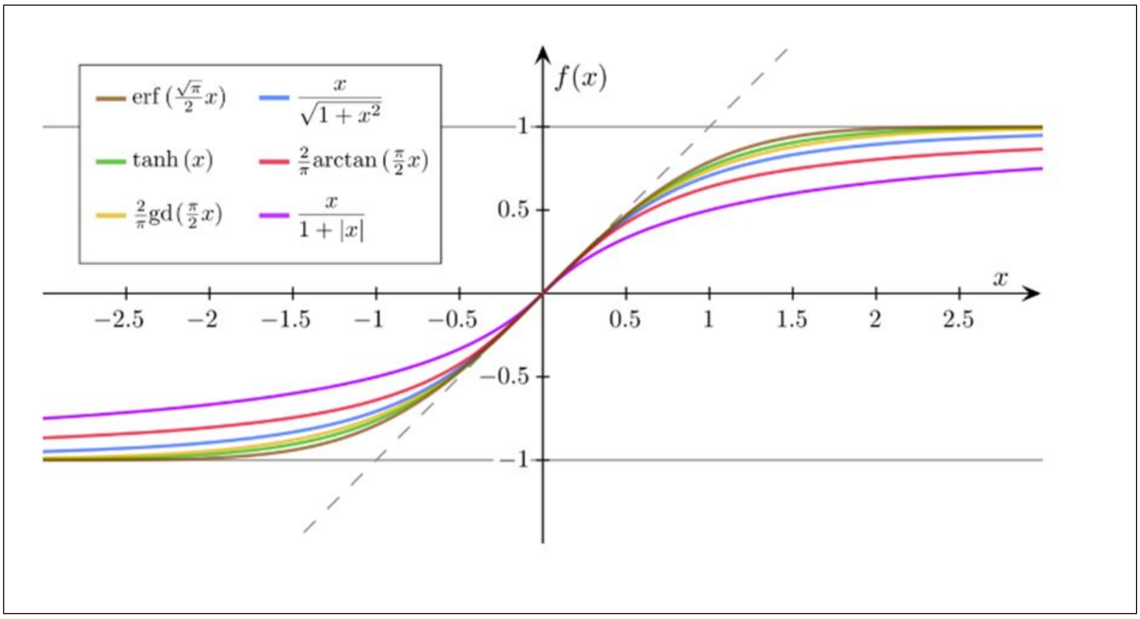


Fig. 3.8 : Different activation functions [Kan11].

The sigmoid function is defined as the Equation (3.49) shows below:

$$f(x) = sigmoid(x) = \frac{1}{1+e^{-x}} \tag{3.49}$$

By considering Equation (3.49), the new formula of the sigmoid function $f(x)$ is defined in Equation (3.50) [SHK+14].

$$f'(x) = f(x)(1 - f(x)) \tag{3.50}$$

In an ANN, the backpropagation process reduces the error by a quarter (at least) at each layer. It essentially has real number as a function domain. It monotonically increase in most cases from 0 to 1 or in other cases from -1 to 1 as shown in Figures 3.9 and 3.10 [MKA11].

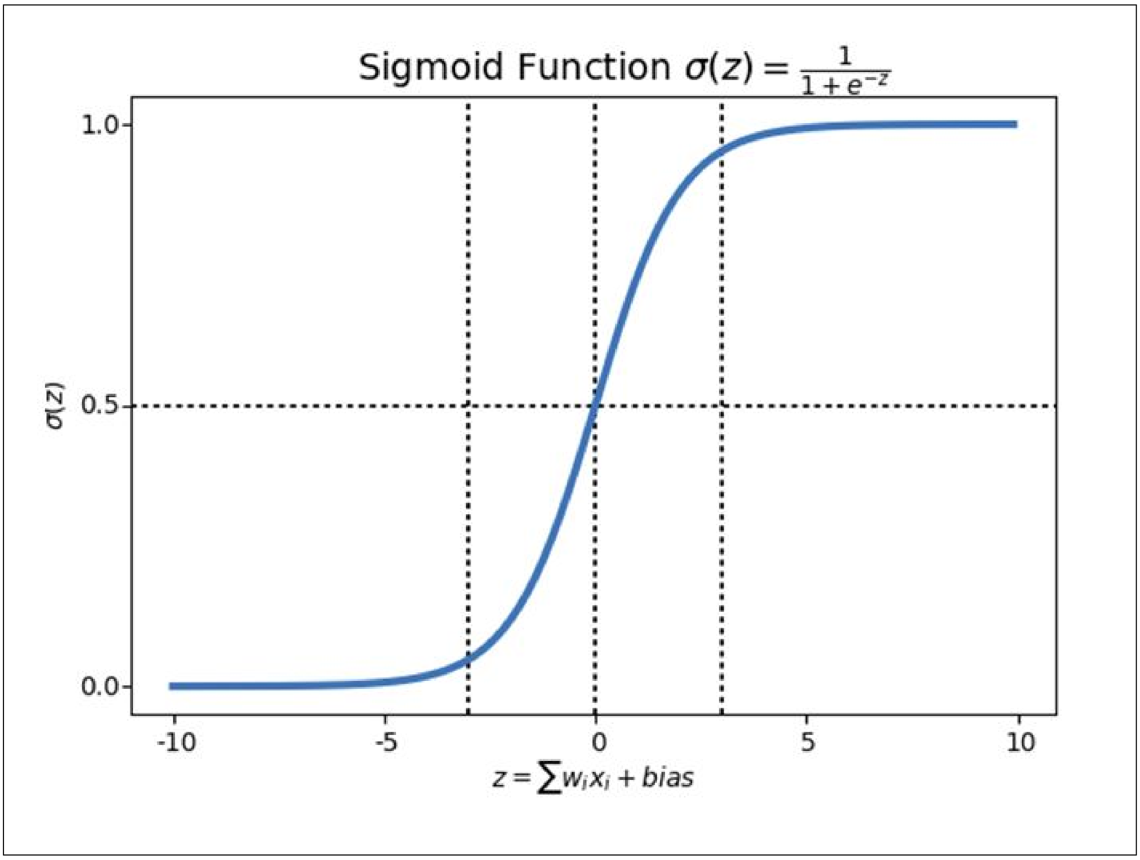


Fig. 3.9 : Logistic curve sigmoid function [MKA11].

2. **Rectified linear units**

Instead of the sigmoid function, most recent ANNs using deep learning use rectified linear units (ReLUs) in the hidden layer [MKA11]. Put simply, ReLU activations are the simplest non-linear activation function where is determined as less than 0 and draw the predict as otherwise [SO03], as Equation (3.51) shows below:

$$f(x) = max(x, 0) \tag{3.51}$$

In the rectified linear unit, when the input is positive the desired output is 1, so the backpropagation error is reduced from the sigmoid function and the training

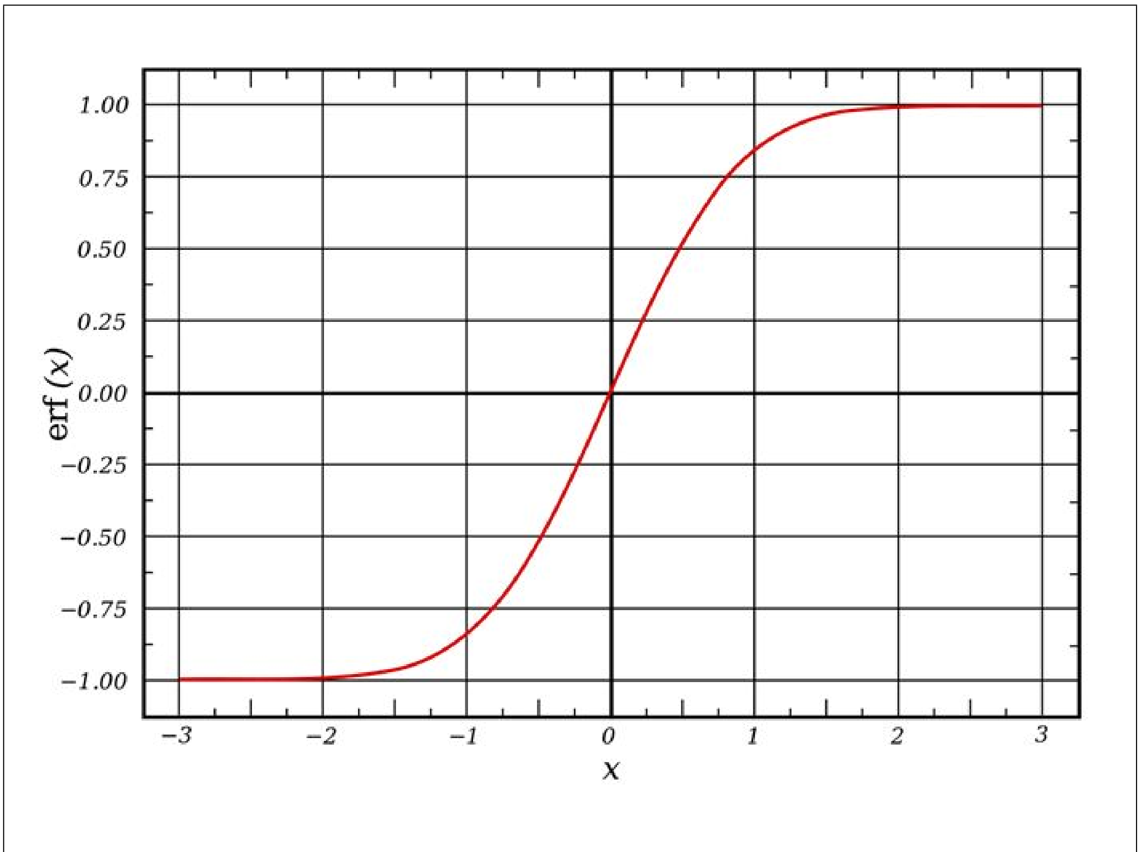


Fig. 3.10 : Error sigmoid function [MKA11].

process is faster for larger networks as shown in Figure 3.11 where x is the input to a neuron [Bis06].

3. Softmax

In some classification problems (binary classification) where the output range is either between $[0,1]$ or $[-1,+1]$, the sigmoid or the ReLUs are the best activation function to choose. In other cases (multiclass classification problems) the sigmoid function and ReLUs are not particularly helpful. In these case, the SoftMax function compresses the outputs of each unit to be between 0 and 1, as a sigmoid function. In another words, each output in the logistic sigmoid function is a total sum of the output. which means that it must be equal to 1 [Kan11]. Mathematically, the SoftMax activation function that is shown in Equation (3.52) is equivalent to a categorical probability distribution that determines which of the classes are true [Kan11].

$$f(x) = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \tag{3.52}$$

Neural network architecture

Typically, an ANN has a different architecture depending on the type of connections between the neurons. Types of connections provide an indication of the behavior of the

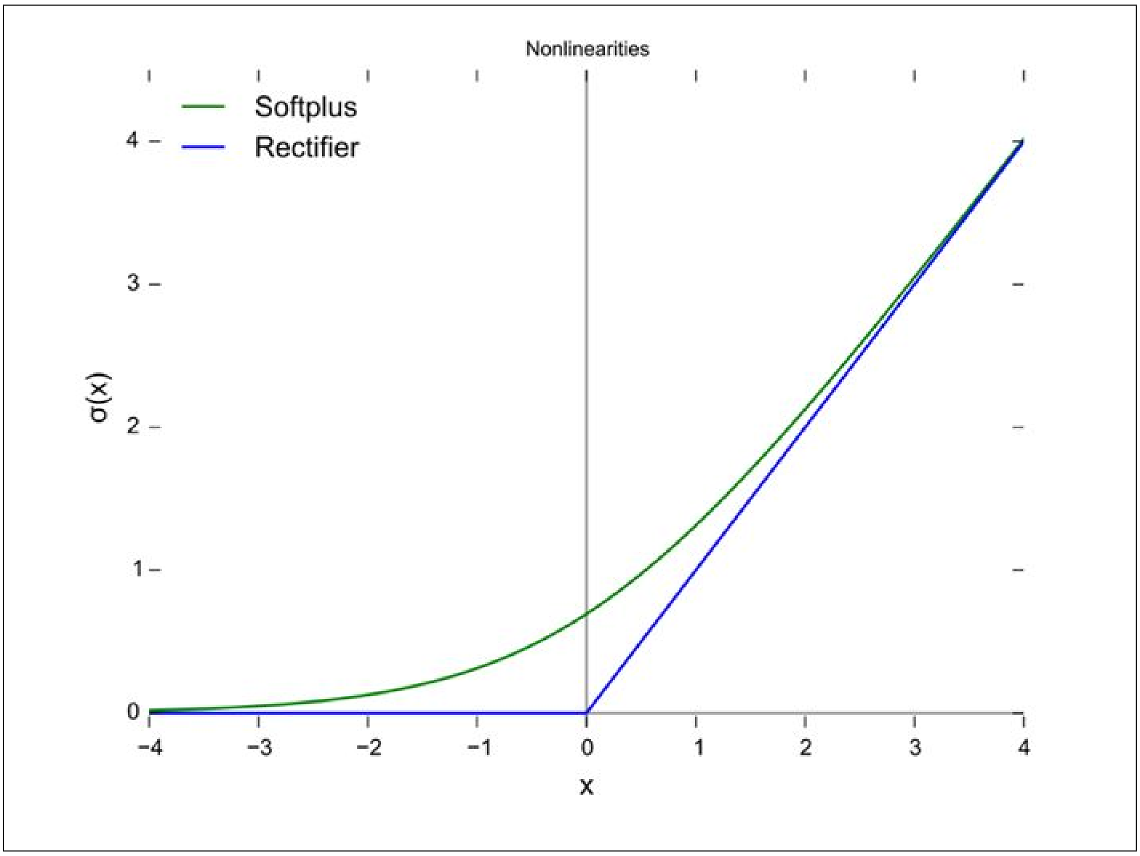


Fig. 3.11 : Rectified linear units: rectifier (blue) and soft plus (green) functions near x = 0 [Bis06].

ANN based on the activation function. In this case, if any of the neuron in the any layer for example the "hidden layer" is linked "connected" to any neuron then in this case each of them "neuron" is linked directly to each one in the output layer [Kan11, Koh82]. A neural network can be classified according to several layers:

Single-layer feed forward networks (SLFFNs) A SLFFN essentially a single layer of connections. Neurons in this type can be distributed as "input units" that pass a signal to other neurons to output as "output units". Figure 3.12 illustrates the typical structure of this model "type" [Kan11, Koh82].

Multilayer feed forward networks (MLFFNs)

The second type of ANN design is the feed forward neural network (FFNN). In this design, the ANN has one input layer, more than one single hidden layer, and also has only one prediction layer or "output layer". Mainly, the main purpose of using a hidden layer is to give the ANN the ability for the hidden layer to extract the higher order of valuable statistics, which are used to fine-tune and adjust all parameters in the ANN. The graph architecture of this model is illustrated in Figure 3.13 which shows the layout of an MLFFN with a single hidden layer [CGGS12, DAYD17].

Neural network training algorithms

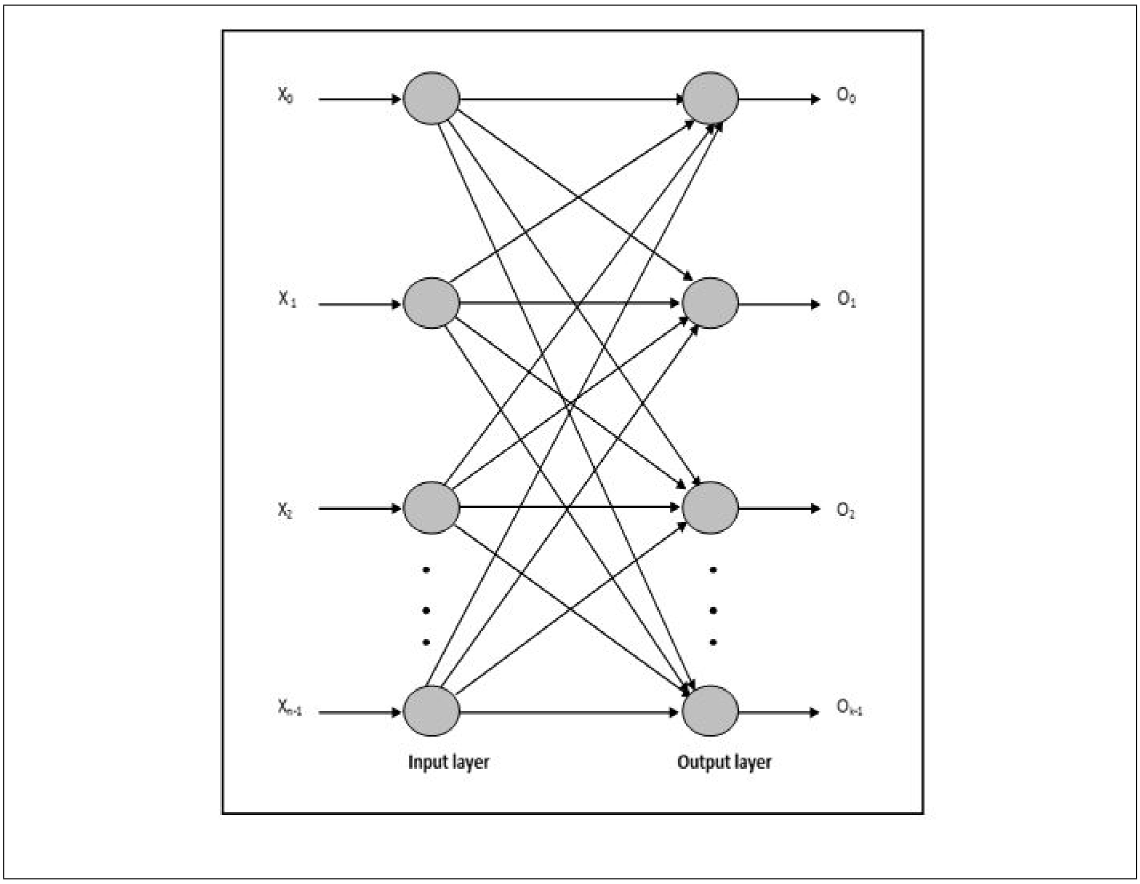


Fig. 3.12 : A single-layer feed forward network [HDE15, TS12].

One of the most interesting features of ANNs is their capability to learn, which means that an ANN has to be trained. The aim of the training is to produce the desired output (or at least a consistent output) when it is applied to a set of inputs to the network [CGGS12, DAYD17]. Training is a process that that uses set weights. Most training algorithms begin by assigning random numbers to the weight matrix. It is then adjusted on the basis of weights [DAYD17, RGC+97]. The most widely used taxonomy is the first one shown below in Figure 3.14:

Three main learning algorithms (systems) have been widely used in different approaches in data mining and training schemes, such as the following:

- Unsupervised learning: In this system, the input patterns is the only component in the training set. The network tries to detect similarities and generates pattern classes [RGC+97].

- Reinforcements learning: Training sets consist of different input patterns; whose values are passed into the network after completion of a sequence, indicating whether the results were correct or not [RGC+97].

- Supervised learning: In this learning system (technique), a set of parameters for the ANN is constructed and computed from training [TGL15].

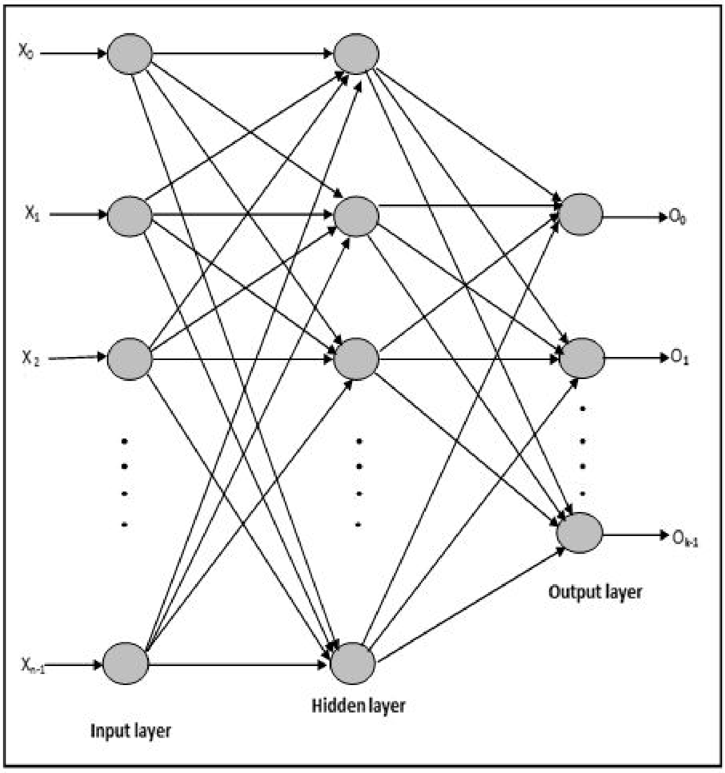


Fig. 3.13 : Fully connected feed forward with one hidden layer and one output layer [CGGS12, DAYD17].

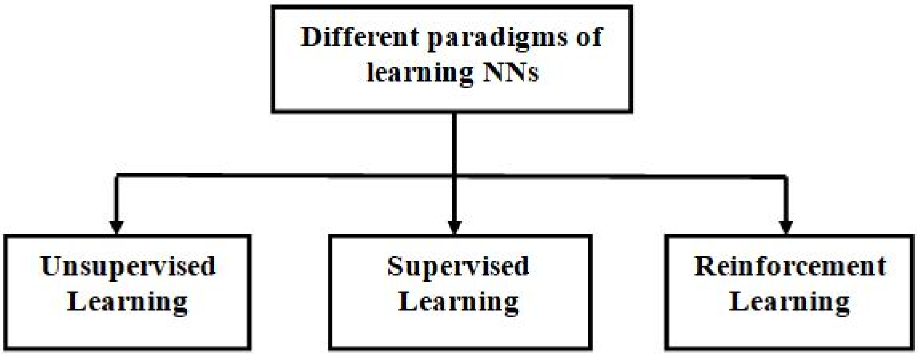


Fig. 3.14 : A neural network category [RGC+97].

Back-propagation

The main design of the backpropagation (PB) learning algorithm (design) uses multi-layer neural network design. This means that the design of the ANN is organized in layers. The output signal sends forward, then the errors of the whole networks are computed and propagated again to the same layer [GRUG17, Kan11, VVW+15]. The backpropagation learning algorithm is mainly used to fine-tune the training parameters, where the error is reduced. The main parts of the backpropagation algorithm are described below [HDE15].

1. Function signals: The function signal represents the incoming signal (input) that comes directly from the input of the network. It is, propagated forward again neuron by neuron along the network and then emerges at the output of the network as an input signal [HDE15].

2. Error signals: Layer by layer, an error from the output neuron of the whole network is computed and propagates backward to the first hidden layer through the whole network [CGGS12].

The estimation computation of the gradient descent vector is the gradient of the difference (error surface) based on the weights. The weights are connected to the inputs of the whole design (input). The backpropagation framework of the ANN architecture is shown in Figure 3.15 [VVW+15].

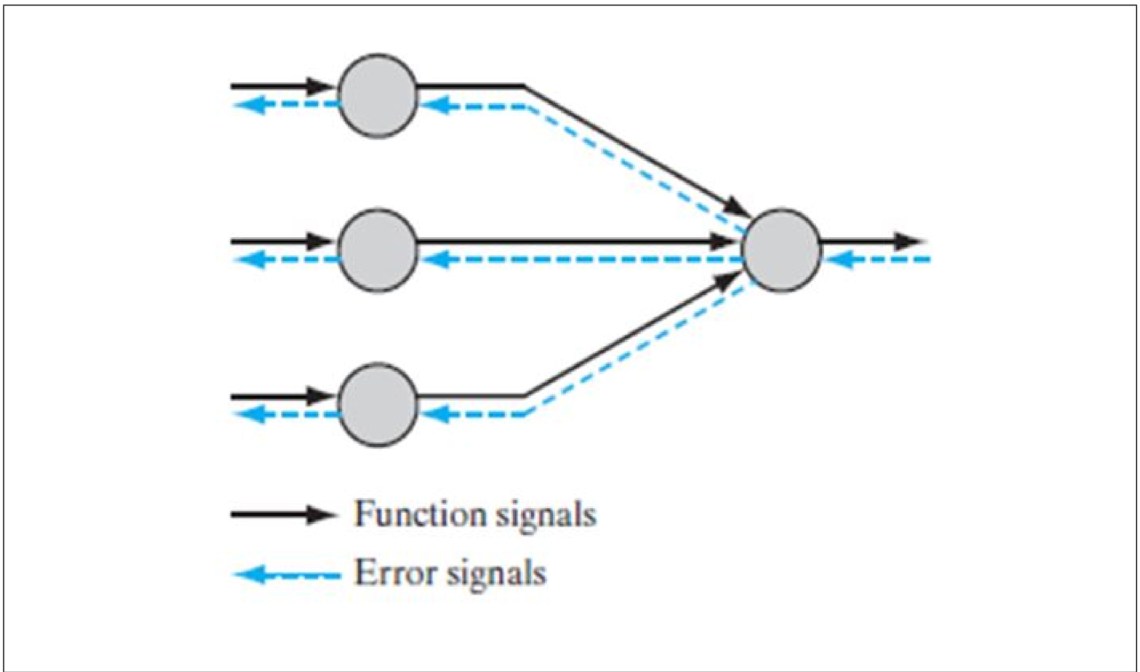


Fig. 3.15 : Illustration of the directions of two basic signal flows in a multilayer perception [VVW+15].

Residual neural network

In a traditional ANN, each layer transfers the parameters to the next layer. More technically, it uses a feed-forward pass to feed the next layer and the next layer for approximately two to three layers. Typically, an ANN is a universal learning function that approximately increases the number of layers that may be added to any structure. However, the limited number of layers is still a major issue in the design of an ANN with improved accuracy. In some cases, an increasing number of layers in the ANN causes the complex learning function to decrease and affects the ability of the universal learning function [GRUG17]. In contrast, DL, uses this situation to increase the number of layers with slim learning function as one critical solution to increase the layer dimensions in the ANN structure. In this case, on the one hand the ANN become more complex and deeper than the original structure of a simple ANN. On the other hand, if there is a desire to further increase the number of layers in the simple ANN structure, we will start at the point of the eventually the overfitting. Therefor,this may show that a deep ANN learns more effectively than regular structure with an overfitting problem [GRUG17]. To overcome these issues, the Residual Neural Network is proposed as a new structural ANN, incorporating the idea of residual connections. Simply, the Residual Neural Network is based on passing the connection output description of the previous layer to the new layers. The Residual neural network tries to omit some connections from the previous layer to the next layer to avoid a full connection and a complex learning function. This is one effective solution to retain the expanded neural network structure without the over fitting issue [CMH+18, GRUG17]. A diagram of an Residual Neural Network (RNN) is presented below in Figure 3.16 [IK19].

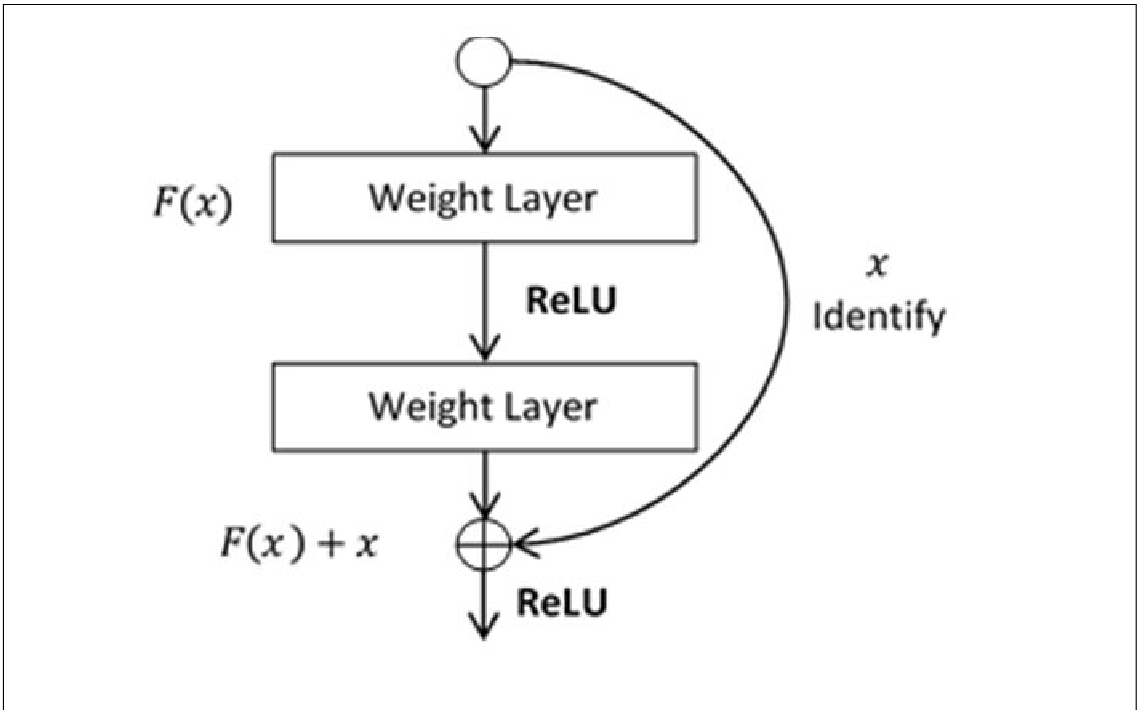


Fig. 3.16 : Residual neural network single residual block diagram [IK19].

From the Residual Neural Network single residual block diagram, it can be assumed that the difference between the input and the output (residual) is define by Equation (3.53) [SMKR13]:

$$R(x) = Output - Input = H(x) - x \tag{3.53}$$

where output is the new weights of the previous layer and input is the original weights of the next layer. By rearranging Equation (3.53), equation (3.54) is produced [SMKR13]:

$$H(x) = R(x) + x \tag{3.54}$$

In this case, the residual block tries to learn the true output $H(x)$. Figure 3.14 above shows that an Residual NN tries to learn residual,

$R(x)$ since it show the identical connection (x) coming from the same input x. In conclusion, the layers in the original ANN try to learn the output $H(x)$ by learning and adjusted weights only, while the residual neural network tries to learn the true output $R(x)$ [SMKR13].

3.4.2 Deep learning prediction model using convolutional neural networks

Recently, the DL approach has shown accurate results in different computer science applications, such as in computer vision and pattern recognition, by achieving high performance results in biomedical image analysis, brain image segmentation [AHMJP12, CCH+14, SMKR13], mitosis detection [LSD15]. The "CNN" is the main key component of DL [SLX+15]. It is essentially a multi-layer perceptron "neurons" network type. It consists of one input layer with multiple neurons, many convolutional layers, and one output layer [LSD15].

Convolutional neural networks architecture

CNN is the key component of the DL approach. The main function of the CNN is denoted by g: it maps data x to another vector (output) denoted by y. The g function, which is the intensive combination, is a simpler form of function f_l. It combines computational blocks or layers; $g = f_1 \circ \cdots \circ f_L$ [SLX+15]. Assuming that the CNN has many input signals that are represented by $x_0 = x$, were $x_1, x_2, ..., x_L$. where the network's output is that comes from the previous output$x_\ell = f_\ell(x_{\ell-1}; w_\ell)$ is computed based on $x_(\ell - 1)$. This is done by applying the function fl with the parameters of wl [KSH12]. Since the data structure has different representations, in this case the network representation is denoted as $x_\ell \in R^{H_\ell \times W_\ell \times D_\ell}$ where data x has a spatial structure, H_ℓ and W_ℓ are spatial coordinates, and D_ℓ is depth of channels. The CNN function f_ℓ mainly acts as an invariant translation function in the main CNN, called "convolutional." It distinguishes between different data classes of desired output in the testing data by assuming that $\hat{y} = f(x)$. In this case, if y is the true label of the data point x then the CNN performs as true label prediction for the testing data point x using a loss function $\ell_y(\hat{y}) \in R$ which assigns a penalty to classification errors [MK20]. A CNN has different components, such as convolution layers that is applying to extract the feature maps, averaging or what is called "max-pooling layers" that is applying to reduce the feature map dimensions, and a finally the fully connected layer which is used for the final prediction layer in the whole network [MK20].

Convolutional layer

The main aim of the convolutional layer is to convolve the output from the previous layer using a set of learnable kernels called "filters," as illustrated in Figure 3.17 [SLX+15]. Figure 3.17 shows a processing example with depth = 1, filter size 3×3, stride = 2, padding = 1. If $\ell = 1$ input image otherwise the output of the previous layer is convolved by filters.

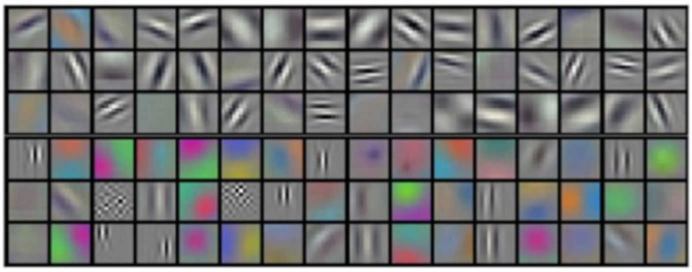


Fig. 3.17 : Sample learned filters, image adapted from [MK20].

Each kernel "filter" is skid through the width or top of the input amount to produce a two-dimentional feature map or "activation map." Each filter "kernel" has the same depth, "dimension" as the input [SLX+15].

The depth filter is the number of filters that are used to the produce the depth of the feature map. These filters detect structures such as edges, corners, blobs, and so on, as shown in Figure 3.17. Also, the stride represents a number of steps that the filter is jumping during the sliding process across the whole input data (image). Also, the zero-padding means that some of the zero-pixels are added to the input data (image) specifically around the border to unify the extracted feature map's size. Assuming a certain dataset with a network structure of N-squared neuron layers, the structure is followed by a convolutional layer with a kernel filter ω with a filter of size $m \times m$.

In this case, the output of the convolutional layer is measured based on the output size estimation of $(N-m+1) \times (N-m+1)(N-m+1) \times (N-m+1)$. However, to estimate the pre-non-linearity input to a unit x, the shared variables are weighted by using the filter components and summed from the previous layer, as shown in Equations (3.55) and (3.60) [MK20].

$$x_{ij}^{\ell} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} y_{(i+a)(j+b)}^{\ell-1} \tag{3.55}$$

Then, the convolutional layer applies its nonlinearity:

$$y_{ij}^{\ell} = \sigma(x_{ij}^{\ell}) \tag{3.56}$$

Pooling layer

The pooling layer is the second component of the CNN. The most popular pooling operators' "layers" are showing in Figure 3.18 and 3.19 [SLX+15]. The main aim of the pooling layer is either select the max value within a Specific small spatial block or to compute the average based on specific parameters such as block size, and stride [KSH12]. Figure 3.18 and 3.19 illustrate a max-pooling operation with 2×2 filters and a stride of two [KSH12].

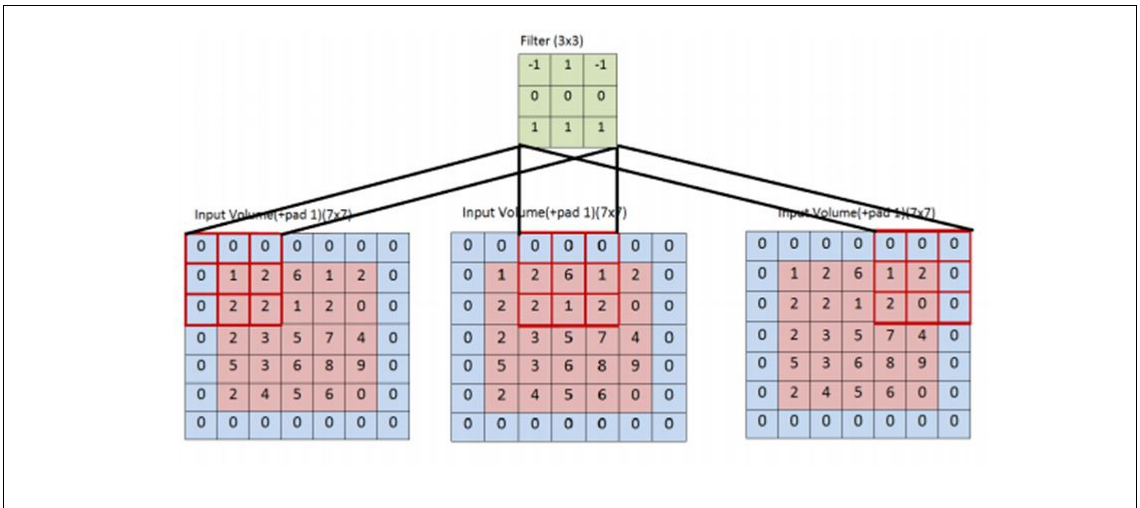


Fig. 3.18 : Sample convolutional layer processing [SLX+15].

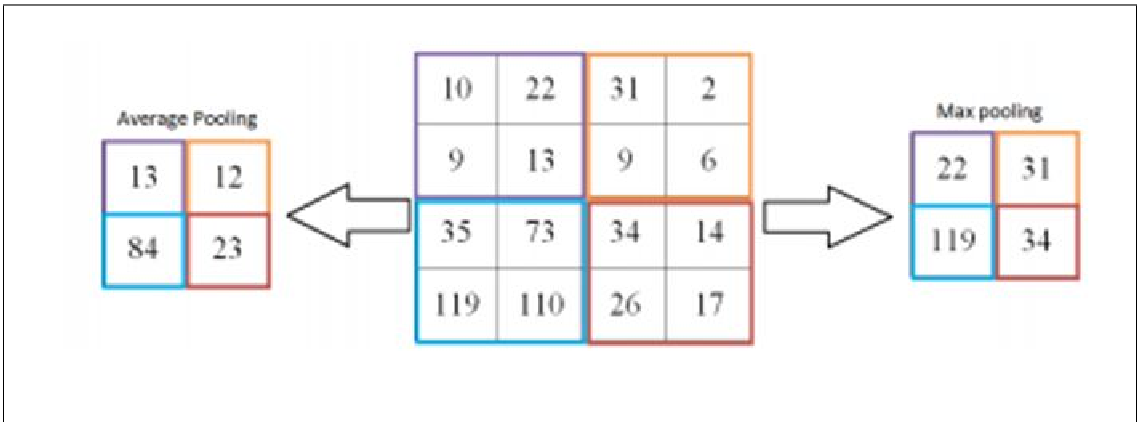


Fig. 3.19 : The average or the maximum pooling layer [SLX+15].

A typical model for machine learning learns the mapping function, which implies that $X \Rightarrow Y$. In this case, it can be represented by a set of possible learned mapping equations, such as that shown in Equation (3.57) [KSH12].

$$OX \Rightarrow f(X, \alpha) \tag{3.57}$$

where α is defined as a set of variables (parameters) that is associated with the function $f(x)$. In this case, A given input of X and a parameter α should be chosen, so that the machine-learning model always gives the approximately the same output [VL15]. The number of classes provides an indication of the classification model's construction. For example, from the training samples if there are two classes, in this case, the goal is defined as the producer to build the final binary classifier. However, based on a small probability value the predictor targets the training pairs values. That causes sometimes a classification error or it is called the misclassifying of the testing samples [BK18]. For instance, if we assume using the feature vector of the classified document X which is represented as an example in the training set, in this case, this feature vector consists of a frequency of distribution (distinct) that has both keywords and Y which is the data labels that is defined by user-category [MK20, VL15].

Fully connected layer

The fully connected layer in the CNN structure is used to update the initial weights as the normal ANN acts, so that the network can successfully predict the desired output [KSH12]. It is based on adjusting the error from each layer, each node in the output layer, the difference between the final predicted value (desired) and correct label (actual responses) is calculated [SLX+15]. In this case, the initial weights are first randomly initialized and are changed during the adjustment process based on the error direction. The change in weights is given in equation (3.58).

$$\Delta w = -\eta \frac{\partial E}{\partial w} \tag{3.58}$$

where η is the learning rate. Then, the output of each neuron j in the main structure is based on the input units $x_i = (x_1, x_2, ..., x_N)$ given by Equation (3.59) [Kou16]:

$$net_j = \sum_{i=1}^{n} w_{ji} x_i + w_{j0} = \sum_{i=0}^{n} w_{ji} x_i = \langle w^j, x \rangle \tag{3.59}$$

where i represents the as the input layer index, and j represents the hidden layer index. Moreover, the term w_{ji} layer j represents the input and hidden weights, as shown in Equation (3.60) [ILW16].

$$y_i = f(net_j) \tag{3.60}$$

The same approach is applied in the hidden units where the neuron output k is derived from the the output layer, as given in Equation (3.61) [ILW16].

$$net_k = \sum_{j=1}^{n} v_{ki} y_j + v_{k0} = \langle v^k, y \rangle \tag{3.61}$$

where k represents the unit index of the output layer, also, p represents number of units in the hidden layer shown in Equation (3.62) [ILW16].

$$z_k = f(net_k) \tag{3.62}$$

3.4.3 Prediction and classification model using deep learning long short-term memory networks and one-dimensional convolutional networks

A new development techniques in DL using CNNs is stacked long short-term memory network "LSTMs" also called "Deep LSTMs." An LSTM is usually called a long short-term memory network. It is a special case of the RNN that is based on composing the state of the transition function f to an output function g. Usually, the RNN model is built and implemented using a multi-layer neural network. The standard RNN functions are directly related to the positional acyclic-based graphs, which are based on a super-source node transition theory. The main structure of the RNN is the same as the graph structure shown in Figure 3.20 [SMKR13]. The basis transition function of the RNN is implemented based on the recursive state representation as shown in Equation (3.63) [AHMJP12]:

$$a(v) = f(a(ch[v]), I(v), v, w_f) \tag{3.63}$$

where w_f and w_g represent the model parameters (synaptic weights) of the networks f and g respectively. The RNN processes the weight transitions based on the graph U to adjust the synaptic weights. In this case, the whole network (graph) encodes the network in both the learning and recall phase: For this reason, it called an RNN [AHMJP12]. Essentially, the feed-forward of the RNN achieves the encoding of the network, in which each node v represents the transition function f, and $a(v)$ represents the state transition of the state v. More specifically, the state $a(v)$ is computed by the transition function of the specific input label $I(v)$. The pseudo-code for the RNN training-based stochastic gradient descent is described in Algorithm 3.1 below [AHMJP12].

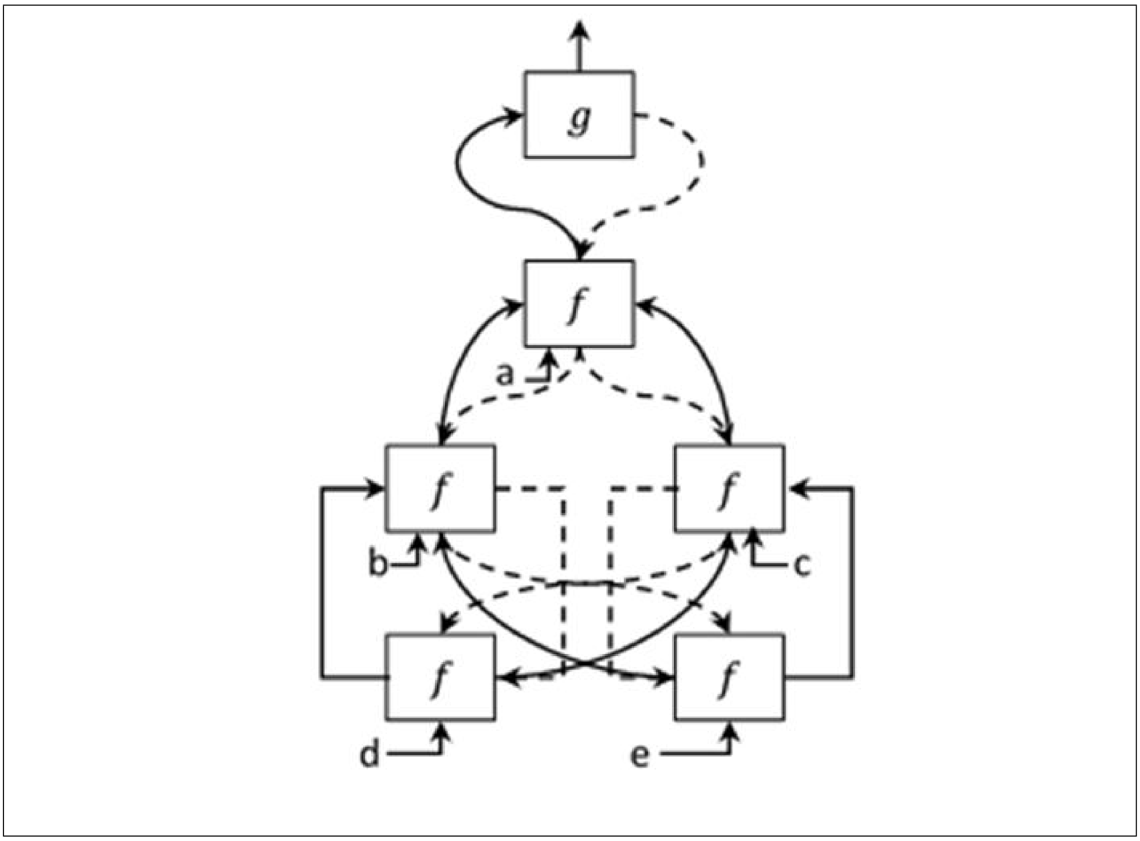


Fig. 3.20 : Recurrent neural network block diagram [SMKR13].

Algorithm 3.1: RNN Training Algorithm

Initialized the model parameters

1. $W \leftarrow random(W^f \cup W^g)$
2. $n \leftarrow 0$
3. $\eta \leftarrow \eta_0$
4. $\phi \leftarrow \phi_0$

6. Repeat
7. $k \leftarrow 1$
8. Repeat
9. randomly select a pattern (U,Y) from training set
10. $g_k \leftarrow S_{Gradients}(U_k,Y_k)$
11. $\hat{g}_k \leftarrow a_k\hat{g}_{k-1} + b_k g_k$
12. $S_k^t \leftarrow a_{k-1}S_{k-1}^t + b_k diag[(g_k - \hat{g}_{k-1})^t(g_k - \hat{g}_{k-1})]1_m^t$
13. Increment the steps $k \leftarrow k+1$
14. **Until** $k = k_{max}$ (all pattern or a small batch in D have been selected).
15. $W_{n+1}^t \leftarrow W_n^t + diag^{-1}[(\phi 1_m + \sqrt{S_{k_{max}-1}})1_m]g_{k_{max}-1}^t$
16. $n \leftarrow n+1$
17. Until performance criterion is met

An LSTM is capable of learning long-term and explicitly avoids the long-term dependency issue (Saha and Senapati n.d.). All RNN networks have a chain design form, repeating modules of a sequence of neurons called states. The first step in the LSTM process is to decide which information should be removed from the cell state. Therefor, the output remains between 0 and 1 for each cell [Gei19].
Stacked LSTMs or deep LSTMs are a special type of LSTM with one-dimensional CNN that is used for challenging sequence prediction problems, such as text meaning prediction. Stacked LSTMs designs (architecture) can be defined as multiple LSTM layers in one LSTM network, similar to CNNs. Each single layer in the LSTM provides a prediction output (sequence output), for example, text prediction output rather than a one single value. Each output is used as a new predicted feature map for the next LSTM layer in the stacked LSTMs network until the final prediction layer of the LSTM network. An example of a stacked or deep LSTM network is shown in Figure 3.21. Each LSTM relies on the one-dimension CNN.

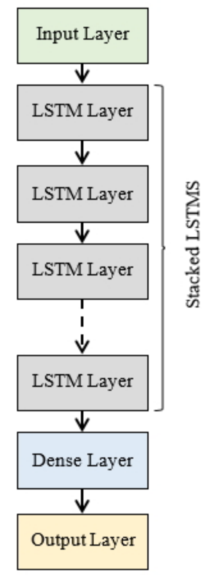


Fig. 3.21 : Stacked long short-term memory design (Architecture).

3.4.4 Prediction model based attention learning mechanisms

"Attention" is defined as a type of action that proceeds directly to the object. In other words, it is defined as a “giving need” which is the ability of a mind to allocate uneven consideration across a field of sensation [TSN18]. Moreover, it helps to focus and bring certain inputs to the core of the attention, at the same time, diminishing or ignoring others [TSN18]. Technically, in a neural network the attention action helps of the credit assignment. The main challenge to that action is long-range dependency. In other words, the prediction is become more impacting and affected by another facts [CLZ17]. The core probability model of the attention network is based on the Markov Assumption [SMGS14] which is aimed to introduce a model that consist different probability numbers as is shown in Equation (3.64) [KDHR17].

$$P(w_1 w_2 ... w_n) \approx \prod_i P(w_i | w_{i-k} ... w_{i-1}) \tag{3.64}$$

In this case, at the high-level demonstration, the attention mechanism approach enables the regular neural network to focus on just a part of the relevant data [KDHR17]. In our design, we use a stack of different blocks to design our CNN models such as a feature extractor layer, convolutional layers, averaging or max-pooling layers in addition to the non-linear transformation layers and normalization layers. As a typical approach, the CNN model uses the convolutional layers based on the sliding window that is cross all the images using different kernels to extract the features map. Technically, the convolutional layer uses the point-to-point inner products to calculate the pixels corresponding area which we called the kernel response area or local respond signal.The local respond signal is the local connection between the kernel and the original data "image area" in which the feature map is collected. In the biological human mechanism, the certain area of the signal response is the visual context, in which all the information is transferred from one cell to another in the human brain [SMGS14]. During the convolutional process (sliding window), kernel size still the same which the filter or the kernel size remains the same. The reason behind that is we want to give the same shared weight to a different area while the image processing is performed, or we called it while the feature map is extracted from different locations. Essentially, shared weights for the fixed window size or what we called "fixed kernel size" reduces the total number of parameters that are used during the convolution process between each pair of hidden layers. That enables the CNN during the convolutional layer to extract similar features from the different areas at the same time based on using the same kernel on the different searching areas. More fundamental features could be indicated based on using a large number of different kernels for a strong feature mining process. For this reason, we design a pre-activation residual process that depends on the output of the attention mechanism. H as is shown below [KDHR17].

$$H_{i,c} = M_{i,c}(x) \times T_{i,c}(x) \tag{3.65}$$

which, i indicates the range of all special area "positions" in the original image, and c represents the index of the original channel since we aim to use a color image in three channels such that $c \in 1, ..., C$. It is essentially based on the idea of the attention model in which the feed-forward inference of the corresponding area can be extracted from the interest area only by depending on the attention backpropagation process as is illustrated below [KDHR17, SMGS14]:

$$\frac{\partial M(x,\theta)T(x,\theta)}{\partial \phi} = M(x,\theta)\frac{\partial T(x,\theta)}{\partial \phi} \tag{3.66}$$

Which, θ indicates the kernel's mask in the attention model and ϕ is the trunk branch parameter. Those two parameters make the convolution process in the convolutional layer work attentionally on the corresponding area only where the main robust features are located. In another word, in the residual learning process, SoftMax enables to construct the identical feature map, in which the output is only the counterpart of the extracted map. To achieve that, we use the following model to design the attention residual process as follows [AAASCH18, AGAA18].

$$H_{i,c}(x) = (1 + M_{i,c}(x)) \times F_{i,c}(x) \tag{3.67}$$

where, the $M(x)$ is the attention indexing that is indicated as a range from [0, 1], and $F(x)$ is the original feature map [AAASCH18]. In this case, as we notice that the extracted feature can be only come from the corresponding labels based on the attention map avoiding the wrong ones or we called the noisy labels based on the updated trunk parameters that are used in equation (3.65). Technically this process is called the attention mechanism in the attention model or in our case we called it the "attention residual learning process" after we ad the attention map to the whole process using the back=propagation process as well in addition to the feed-forward [AAASCH18]. In the previous design (ResNet) model, the residual learning formula is based on the typical formula that $H_{i,c}(x) = x + F_{i,c}(x)$. We can notice that the extracted feature map comes only from the residual branch in addition to the original feature map. In contrast, in our formula, $F_{i,c}(x)$ is indicated by only the corresponding residuals that comes based on the mask branches $M(x)$ which pays more attention to the corresponding labels than the other noisy ones and works as an enhancer for the feature map extraction [AAASCH18].

Chapter 4

NLP-based Sentiment Analysis and Reviewer's Opinion Mining Prediction

This chapter presents the concept of text sentiment analysis and opinion mining. Different DR methods have been used to reduce the dimensions of data to enhance detection accuracy when using ANNs, and we examine how it differs when using DNNs.

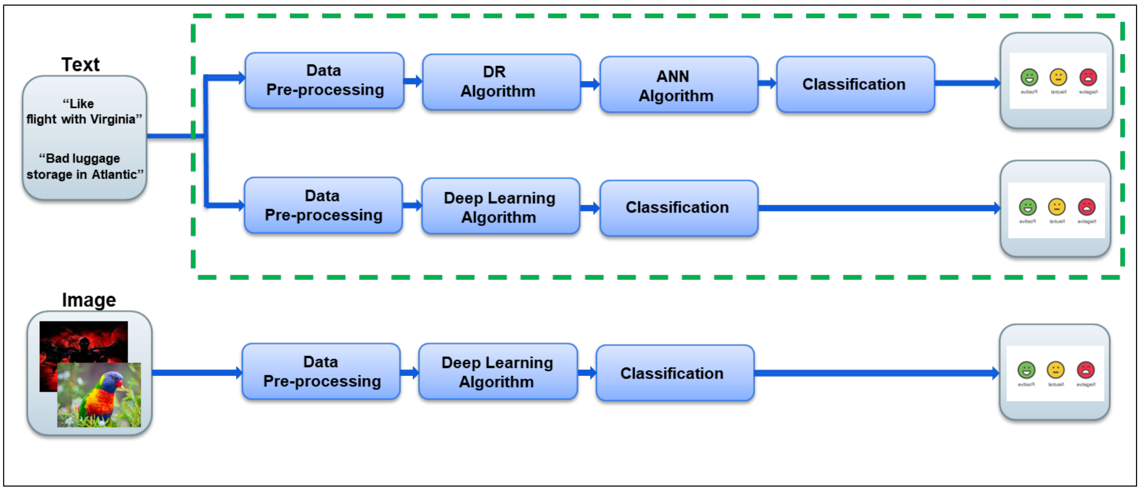


Fig. 4.1 : The highlighted part denotes how chapter 4 contributes to overall workflow of the general proposed system.

4.1 Sentiment in Social Media Text

Nowadays, individuals check different sites, before conducting business, purchasing any item online, or selecting any service or product. Social networking services, in this information era, play a significant role in creating a different social NLP-based sentiment analysis and reviewer's opinion-mining prediction. Media "networks," such as Flickr, Facebook, Twitter, Telegram, YouTube, and Instagram, are now widely considered to be a significant integral part of our modern life.

However, many commercial sites, such as Amazon and eBay, rely on users' reviews (comments) to evaluate their products and services, while other sites are specially designed for users and reviewers to evaluate any product or service. Nevertheless, selecting the best review is still a big challenge. This chapter outlines the proposal for the main approach for text-mining analysis-based on sentiment analysis for detecting the social emotion based on social media text data. The proposal includes the introduction of two models for sentiment detection approaches: The first is NLP-based opinion-mining prediction using dimensionality reduction and a residual neural network classification algorithm, and the second is sentiment analysis and opinion mining using deep-learning approach based on deep LSTMs. Therefore, the first and second contributions stated in Section 1.6 are fulfilled in this chapter. The first approach is to build an opinion-mining and a system to predict an individual's mood based on data extracted from social media platforms through NLP sentiment analysis for text-feature extraction and dimension reduction, in addition to machine learning for classification. In the second approach, the predictin of individual's moods on social media platforms is improved through DL and neural networks.

4.2 First Model: NLP-Based Opinion-Mining Prediction Using Dimensionality Reduction and a Residual Neural Network Classification Algorithm

Social media opinion data contained in online reviews is a good starting point for our approach, because reviewers (people) openly share their "thoughts," known as "opinions" by general public researchers. Here, an opinion-mining approach proposed based on sentiment analysis, as illustrated in Figure 4.2. The proposed system has three main parts: 1) the pre-possessing of the dataset using NLP tools; 2) dataset visualization and statistical observation; and 3) data-mining approaches for opinion-mining prediction and classification. Our first model has four main steps. The first step is pre-processing the reviews dataset by processing each review text. Secondly, a big picture of the reviews' opinion is drawn, based on analyzing the distribution of the classes in the data using one of the statistical approaches (a histogram). Thirdly, we use the data-mining approaches for text data dimensionality deduction and use a feature-extraction and selection model to extract significant features, which are used in the last step, which is the opinion-prediction model using a residual neural network. A sample of the original data and pre-process are presented in Appendix B.

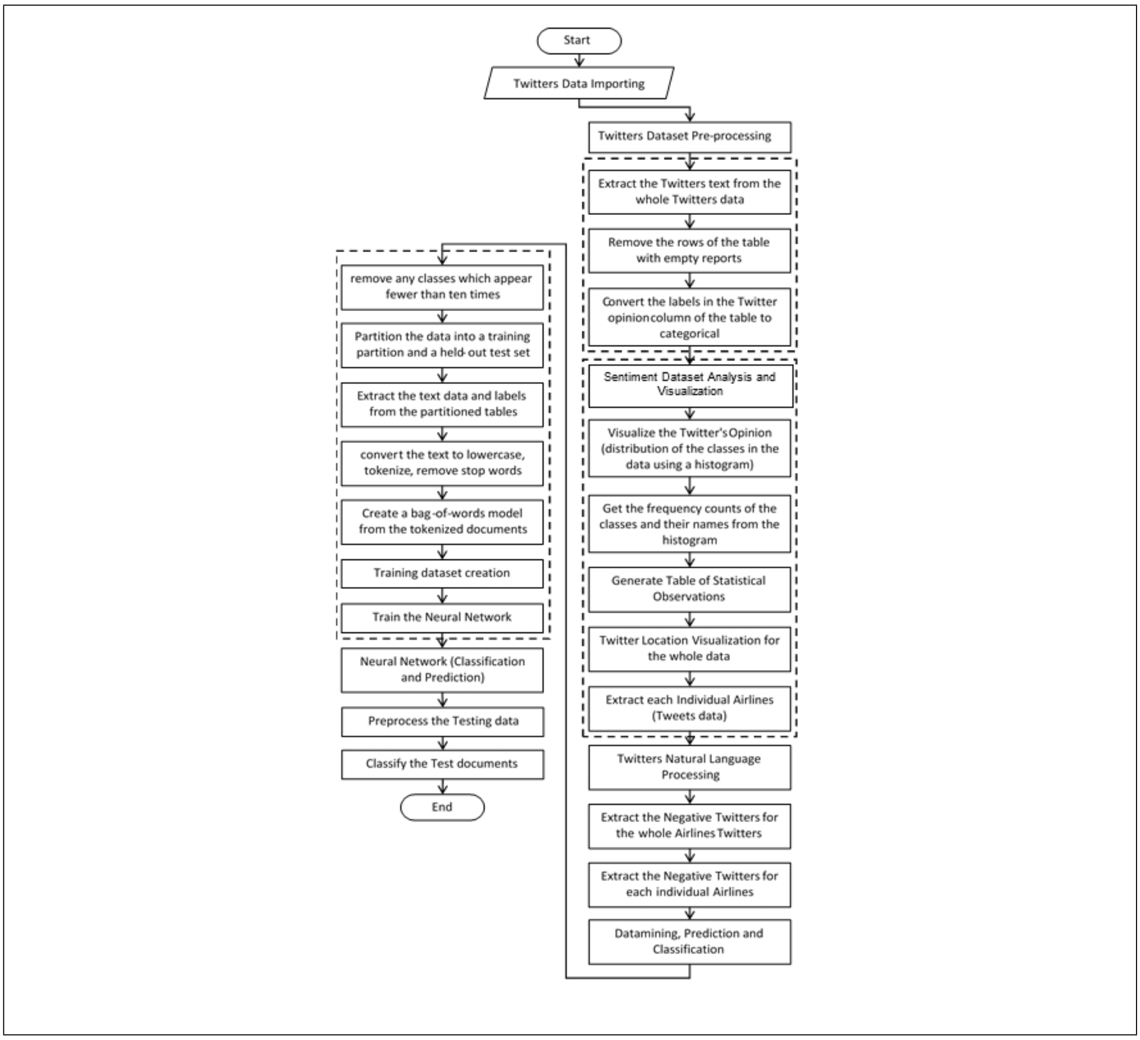


Fig. 4.2 : Opinion-mining propose approach.

4.2.1 First step: Dataset pre-processing based o natural language processing

In this step, some NLP tools are implemented for pre-processing the reviews in terms of extracting and predicting people's opinions. At the beginning, the dataset is processed to extract the row text from the whole dataset. First, the review text is extracted from the whole dataset, and the tables with empty reports are removed. Subsequently, the labels in the Twitter opinion column are converted in to categories. Four different steps are implemented in this case, which are shown in Algorithm 4.1.

Algorithm 4.1: Opinion-Mining Propose Approach

Input: Initialize the **X** "the data Matrix"

Output: Generate a new Dimensional array **C**

1. **Step1:** Erase the punctuation, then make everything in the lower case.
2. **Step2:** Replace hashtag abbreviations with standard language.
3. **Step3:** Tokenizer the text data.
4. **Step4:** Use the tokenized documents to create the bag-of-words dictionary.

Then, from the dictionary, the non-appearance words are removed, based on the condition that any words that appear less than two times are removed, and these words are also removed from the corresponding entries in the labels.

4.2.2 Second step: Analysis of the reviewer's sentiment dataset

In this step, we analyze the review's opinion using the distribution of each class to build the observation histogram. This is a common analytical method that can be performed on a large number of reviews. Sentiment analysis provides another way to observe the data by scoring each word in a particular tweet. It is a convenient method that takes the pulse of the public review. The main steps in the review-data analysis that are implemented in this stage are shown in Algorithm 4.2.

Algorithm 4.2: Reviews Sentiment Dataset Analysis

1. **Step1:** Import the words frequency by counting the classes and their names from the gen the histogram.
2. **Step2:** Generate table of that statistical observations.
3. **Step3:** Visualize the Twitter's Opinion (distribution of the classes in the data using a histogram).
4. **Step4:** Twitter location visualization for the whole data
5. **Step5:** Twitters location visualization (for each individual airlines).
6. **Step6:** Extract the Negative Twitters for the whole airlines Twitters dataset
7. **Step7:** Extract the Negative Twitters for each individual airlines.

4.2.3 Third step: Dimensionality reduction feature extraction and selection

Data preparation Data-mining approaches are used in this step for text opinion-mining prediction and classification. In this task, we use data-mining techniques such as dimensionality reduction and feature selection, as well as a supervised learning classification framework. To implement this, some important steps are implemented, such as follows.

Data partitioning In this stage, the data is partitioned into different folds, a training partition and a held-out test set, by specifying the hold-out percentage at 10%.

Sub-feature selection based on non-negativity In this step, the text features and labels from the partitioned tables are extracted, using a new dimensionality-reduction approach. This has two main aspects: Sub-feature selection and dimensionality reduction. Sub-feature selection utilizes MI, which was discussed in Section 3.4.3, and Algorithm A.3 (Appendix A). The second aspect develops the SVD approach that was discussed in Section 3.4.1, described in Equations (3.1) to (3.48), and illustrated in Algorithm A.1 (Appendix A). Our development of the SVD, which is a dimensionality-reduction approach, in terms of sub-feature selection, is a step to make our dimensionality-reduction algorithm non-biased thresholding. A mathematical model is developed and implemented in this thesis to automatically select a pre-defined threshold that is defined and based on the mutual information (MI) scoring. This determines the sub-feature selection based on the uncertainty features using mutual information (MI) scoring. After data is normalized, the mutual information (MI) is used in measuring the non-negative-value features selection. In another word, it measures the dependency between every two variables in the data. It essentially depends on using x and y and extracts the joint distribution between both variables. In this case, mutual information (MI) measures the specific dependency value that is resented by $I(x,y) = 0$. In another word, the probability of x is illustrated as p by calling both x and y are independent and random variables [VPZ13, ZB16a], as shown in Equation (4.1).

$$P(x,y) = P(x) \times P(y) \tag{4.1}$$

After applying the log function the probability is driven as it shown in Equation (4.2).

$$log\left(\frac{p(x)}{P(x)P(y)}\right) = log(I) = 0 \tag{4.2}$$

That's mean the mutual information (MI) scoring is non-negative values i.e.,

$$I(x,y) \geq 0 \tag{4.3}$$

And the same time the mutual information (MI) scoring is a symmetric as it given in Equation (3).

$$I(x,y) = I(y,x) \tag{4.4}$$

Moreover, the mutual information is equivalently expressed based on the following Equation below after approving that the MI selects the non-negative value.

$$\begin{aligned} I(x,y) = H(x) - H(y) \approx H(x) - H(\frac{y}{x}) \approx H(x) + H(y) - H(x,y) \\ \approx H(x,y) - H(\frac{x}{y}) - H(\frac{y}{x}) \end{aligned} \tag{4.5}$$

In which $H(x)$ and $H(y)$ are both the original entropy, also the $H(\frac{x}{y})$, $H(\frac{y}{x})$ is the expressed and the conditional entropy. Since $I(x,y)$ is a non-negative value, then in this case, the consequently of $H(x) \geq H(\frac{x}{y})$ is expressed as is shown below.

$$I(x;y) = H(y) - H(\frac{y}{x}) \tag{4.6}$$

The approve of that is illustrate in following steps Equation (4.7).

$$\begin{aligned} I(x,y) &= \sum p(x,y) log\Big[\frac{p(x,y)}{p(x)p(y)}\Big] \\ &= \sum p(x,y) log\Big(\frac{p(x,y)}{p(x)}\Big) \qquad (4.7) \\ &- \sum p(x,y) log(p(y)) \\ &= \sum p(x)p(\frac{y}{x}) log(p(\frac{y}{x})) \\ &- \sum p(x,y) log(p(y)) \\ &= \sum p(x)\Big(\sum p(\frac{y}{x}) log(p(\frac{y}{x})\Big) - \sum logp(y)\Big(\sum p(y,x)\Big)) \\ &= -\sum p(x) H\Big(\frac{Y=y}{X=x}\Big) - \sum log(p(y)p(x))r \qquad (4.8) \\ &= -H(\frac{y}{x}) + H(y) = H - H(\frac{y}{x}) \end{aligned}$$

However, if the entropy $H(y)$ in this case measures the uncertainty between each two variables, then $H\Big(\frac{Y=y}{X=x}\Big)$ measures the other criteria, indicating the extent to which variable X does not depend on variable Y. Therfor, it is the measure of the amount of uncertainty that is still reamed by the variable Y after X is known.

Our development for the singular value decomposition

Our a new method for dimensionality reduction is based on the basic SVD algorithm. The main contribution is using the MI to rank our data features then using the non-negativity approach to find the significant value for the automatic threshold selection. Essentially, the developed SVD approach ranks the whole features set with a non-negative diagonal elements based MI, as shown in Algorithm 4.3.

Algorithm 4.3 : Our proposed dimension reduction algorithm (SVD)

Input: Original input matrix X with the whole feature dimensions **X**.
Output: Automated new dimensional data **C**.
Step1: Extract the whole input data siz.e
Step2: Set-up the initial eigenvector ratio by 0.1.
Step3: Calculate the squared decompose data matrix **S** by $X^t.X$.
Step4: Automated optimal new reduction dimensional reduction.
  Step 4.1: Compute the diagonal data matrix of X.
  Step 4.2: Extract the optimal maximum value in the calculated data X.
  Step 4.3: Extract the new data seize.
  Step 4.4: Determine the new data dimension.
    Step 4.4.1: Compute the component vector-based matrix eigen value.
    Step 4.4.2: Extract the diagonal Eigenvalue vector.
    Step 4.4.3: Re-sort the eigenvalues and extract a new eigenvalue vector.
    Step 4.4.4: Extract the last eigenvalue's index (maximum ones).
    Step 4.4.5: Calculate the new matrix U of eigenvalues.
Step 5: Automatically extract the max eigenvalue from the new matrix U.
Step 6: Extract the eigenvalue index.
Step 7: Replaced the extracted indexing by the eigenvalue.
Step 8: Automatically select the new eigenvalue.
Step 9: Calculate the new matrix U (unitary matrices).
Step 10: Compute the new diagonal vector S in a decreasing order of rank X that has the non-negative eigenvalues.
Step 11: Omit the minimum eigenvalues that represent half of the new data vector.
Step 12: Produce the final data matrix V based on the unitary eigenvector matrices V.

4.2.4 Fourth step: prediction and classification

The main processing steps (whole approach) of using NLP sentiment analysis based on the dimensionality reduction and feature selection for the opinion mining of online reviews are described below. First, the whole dataset is partitioned into training and testing datasets. We specify the hold-out percentage at 10% for the testing dataset and the remaining 90% for training. The total numbers of training and testing samples are shown in Table 4.1. Secondly, in terms of applying the NLP and data-mining approach (training and testing) to actual dataset, we extract the reviews and labels from the partitioned tables. The same step is applied to the testing dataset. The text data is extracted and isolated from the whole dataset. Then, in terms of using the text data as an input for

Table 4.1 Whole training and testing dataset.

Whole Dataset	Training Samples (90%)	Testing Samples (10%)
11712 samples	10541 samples	1171 samples

dimensionality reduction and feature selection, the whole text data should be converted to a set of features.

To do this, we use a document-to-sequences approach. To convert the documents into sequences of word vectors (features) for both training and testing data, we calculate the documents length for both training and testing. The training and testing dataset labels are extracted and isolated. In this case, after the text data is extracted for both training and testing, the NLP text pre-processing steps are applied to both. Punctuation is deleted from each text, then upper-case letters are converted to lower case, hashtag abbreviations are replaced with standard language, and the vocabulary list (dictionary) is built by tokenize the whole text data. The next step of the NLP-text processing is creating a bag-of-words model from the tokenized documents. The anonymous function inputted to docfun takes string array input and outputs the elements, which are the histogram of words (HoW) or bag of words (BoW). Then, the words from BoW model are removed, especially those that do not appear more than twice in total. Subsequently, any documents containing no words are removed along with the corresponding entries in the labels. Finally, we convert the documents to sequences of word vectors that will be optimal for use in the dimensionality-reduction and features-selection algorithm, as well as the prediction and classification approach. To convert the training documents into a cell array of sequences, we use the example function doc2sequence, shown at the end of this example. The columns of each sequence are the word vectors. Then, the training dataset is created by converting the sparse matrix to full matrix (features vector construction).

In addition, the labels of each review (y) in the training data set are converted to a number that is used as a numeric label in the prediction and classification approach. In this case, we have a full matrix of features (11712 Tweets and 2976 features). We must then chose, which dimensionality-reduction method to use for this approach: PCA, SVD, developed SVD, or data with out dimensionality-reduction methods. After using the CDF function for uncertainty point detection (UPD), we determine the UPD threshold within which all the features are selected.

Finally, the new feature domain that is selected based on the dimensionality-reduction and feature-selection approach is ready for classification. For more details of the step-by-step implementation of NLP-based sentiment analysis and dimension reduction, see Appendix B.

In a traditional ANN, each layer transfers the parameters into the next layer. More technically, it uses the feed-forward pass to feed to the next layer and directly to the next layer, for approximately two- to-three layers. Typically, an ANN is a universal learning function, which approximately increases the number of layers that would be added to any structure. However, a limited number of layers remains a major issue in the design of the ANN to improve the current accuracy. Therefore, in some cases, the increased number of layers in the ANN causes a complex learning function that drops and affects the ability of the universal learning function [KSH12].

In this sotuation, DL, in contrast, to increase the number of layers, uses learning function as one critical solution to increase the layer dimensions in the ANN structure. In this case, the ANN becomes more complex and deeper than the original structure of the simple ANN. However, if we want to further increase the number of layers in the simple ANN structure, we start at the point of the eventually the over-fitting. Therefore, it might show that the deep ANN is learning better than the regular structure with the over-fitting problem [BK18]. To overcome these issues, a residual neural network is proposed as a new structural ANN with the idea of the residual connection. Simply, the residual neural network is based on the connection of the output description in the previous layer to the new layers. In this case, the residual neural network tries to extend some connection from the prior layer to the next layer to avoid the full connection and the complexity learning function. This is great solution to keep the expanded neural network structure without the over-fitting issue [BK18, Kou16]. The main diagram of the residual neural network is illustrated in Figure 4.3 [Kou16].

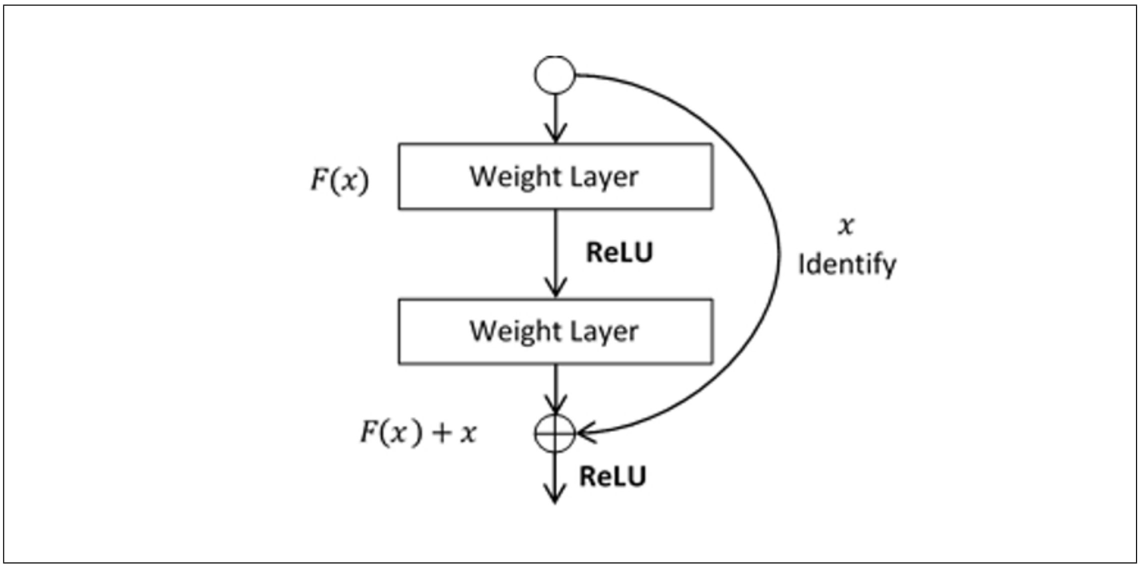


Fig. 4.3 : Residual neural network single residual block diagram [KSH12].

From the residual neural network single residual block diagram, we can assume that the difference between the input and the output (residual) is defined by Equation (4.9) [SMKR13]:

$$R(x) = Output - Input = H(x) - x \tag{4.9}$$

where Output is the new weights of the previous layer and Input is the original weights of the next layer. By rearranging Equation (4.9), we obtain Equation (4.10) [SMKR13].

$$H(x) = R(x) + x \tag{4.10}$$

In this case, the residual block is irritating to acquire overall the correct productivity (output) $H(x)$. Figure 4.3 indicates that the residual neural network tries to learn the residual R(x), as we have the actual identity connection (x) that comes from the same input x. In conclusion, the layers in the original ANN try to learn the output H(x) by learning and adjusted weights only, while the residual neural network tries to learn the true output R(x) [SMKR13].

Conceptually, in the classification stage, we use the residual neural network methodology in our regular expanded neural network to reduce the over-fitting and achieve better accuracy. The input of the training formula consists of examples in the form of feature vectors with a label appointed to them. The aim of the classification algorithm is to learn to assign correct labels to new unseen samples of constant task. A classification formula consists of three parts: A model, a classification module, and a learning module. The learning module builds a model supported by a tagged training set. This model consists of designed by the training module and contains a group of associative mappings (e.g. rules). These mappings, once applied to associate untagged check instance, predict the labels of the checked set. The prediction of the labels of the test set is conducted using the classification module.

Training data pre-processing To pre-process the training data, the punctuation is deleted, the text is converted into lowercase, and then the text tokenized.

Training and classification model The classification model is trained, and then the test documents are classified using this trained model. The main design of our residual neural network for the proposed system has one input layer, three middle (hidden) layers, and one prediction layer (output), as illustrated in Figure 4.4. The learning activation function that is used for each hidden layer is a rectified linear units (ReLUs) as given in Equation (3.51) and discussed in Chapter 3 in the artificial neural network prediction model section, while the activation function for the output layer is the SoftMax that is given in Equation (3.52) in the same section.

The general flowchart of the ANN algorithm that has been used in our approach is illustrated in Figure 4.5.

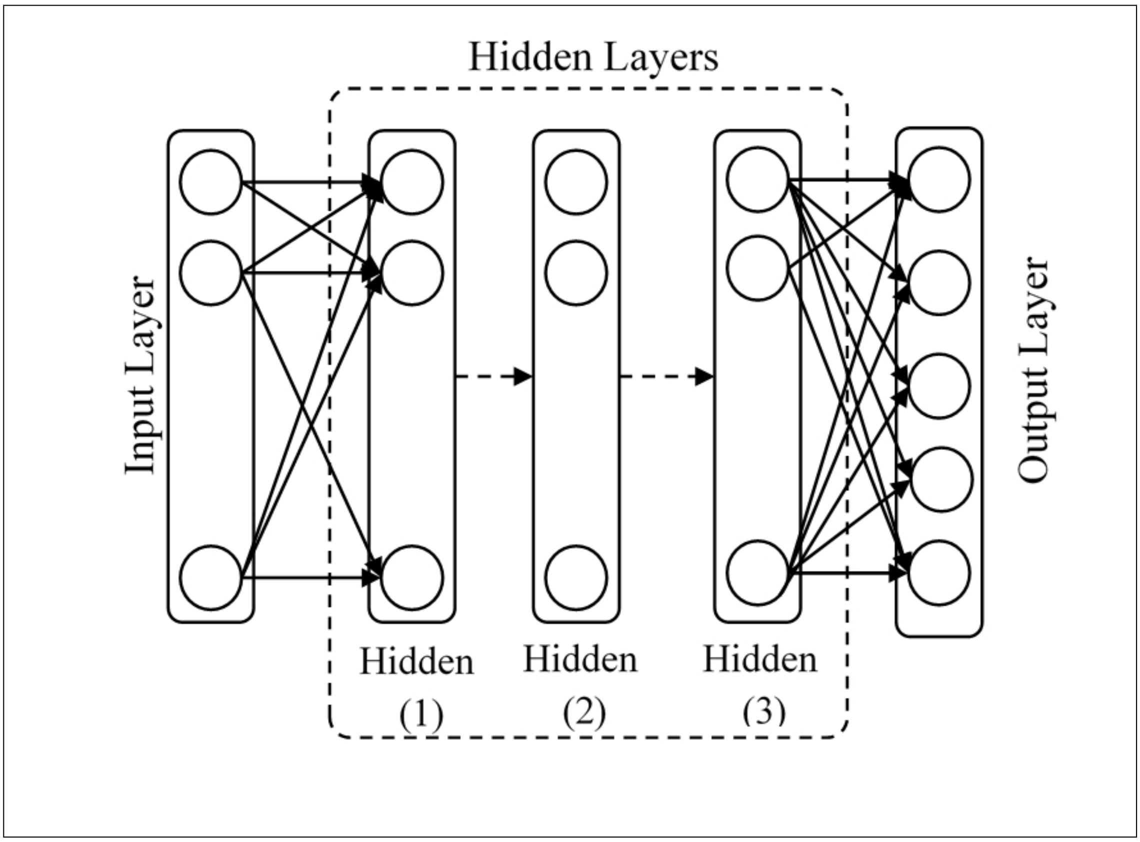


Fig. 4.4 : Residual neural network framework design.

4.3 Second Model: Users' Mood Prediction Based on Deep Mood Prediction Using a One-Dimensional Convolutional-Neural-Network-Based Deep LSTMs Approach

In this section, we propose the opinion-ming approach illustrated in Figure 4.6 based on DL. The proposed system has three main parts: 1) we pre-process the users' reviews (comments) dataset by processing each user's text data; 2) draw a big picture of the user's opinion based on analyzing the distribution of the classes in the data using one of the statistical approaches (a histogram), 3) we design a Deep LSTM network-based one-dimensional convolutional neural network (CNN) for user's opinion-mining prediction and classification. Text mining using DL is initially processes many documents that are gathered. In other words, text mining using DL approaches and tools that are primarily used to extract the information or features from the documents and to process them [Swa16]. The main stage of the text mining using DL approaches is the text analysis or the pre-processing step, in which various techniques are repeatedly used until some relevant information is extracted from the processed documents in this stage [Agg15]. DL approaches or tools organize the document or data structure from the databases once, whereas a text DL approach extracts information from the semi-structured and structured

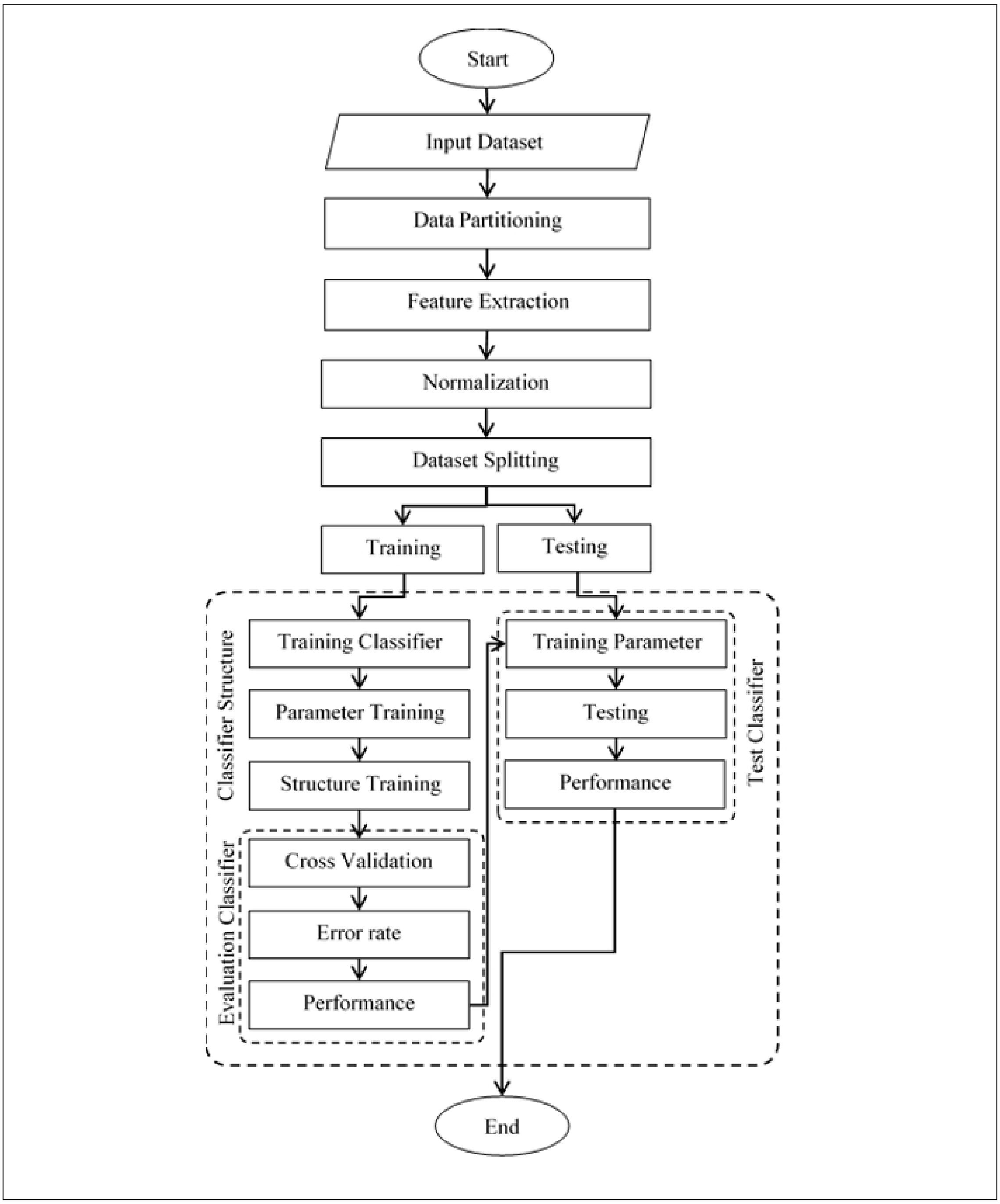


Fig. 4.5 : Standard back-propagation classification approach.

dataset, such as e-mails, text, online reviews, and HTML files. [Moo20]. However, DL tools and approaches are the best options for organizing and handling the online data [CEHM14]. A high-level general approach for text mining using DL is illustrated in Figure 4.7.

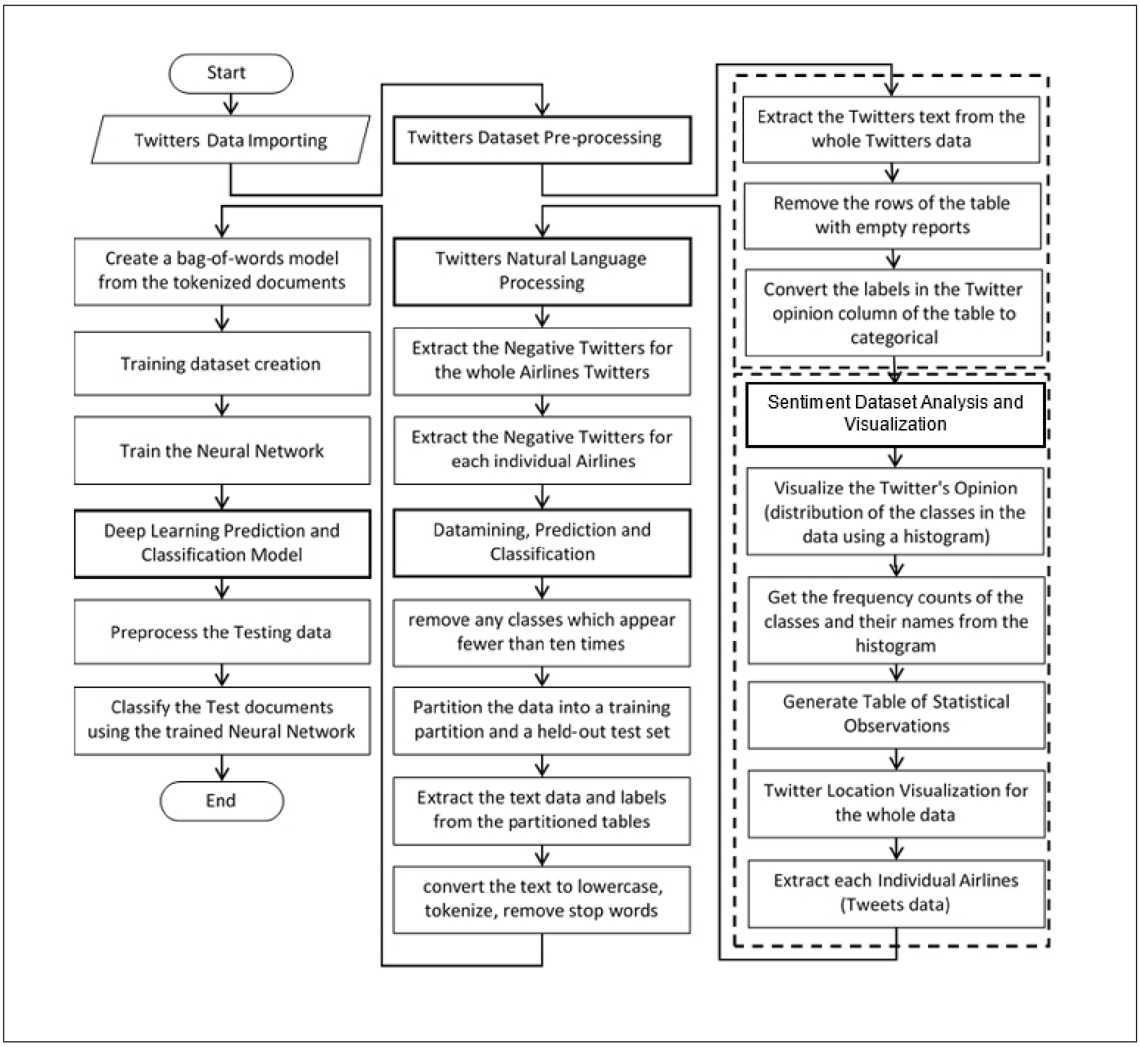


Fig. 4.6 : Opinion-mining based deep learning propose approach.

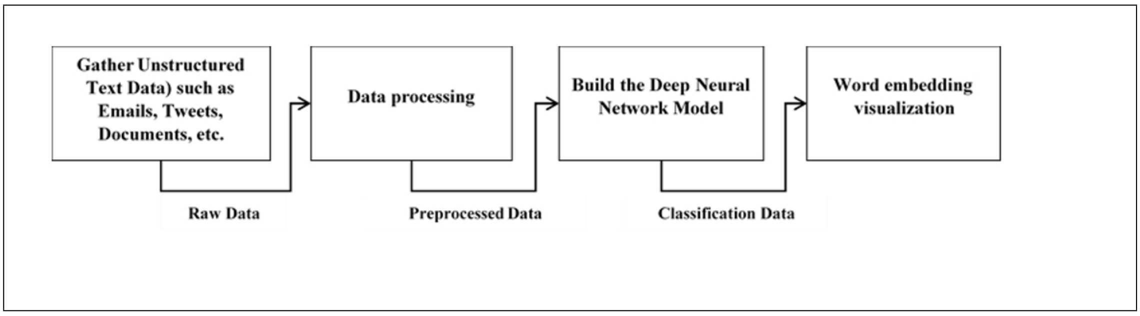


Fig. 4.7 : A high level of text mining general approach using deep learning approach.

4.3.1 First step: Dataset pre-processing based on natural language processing

In this step, some NLP tools are implemented in pre-processing the review dataset in order to extract and predict the opinion-mining approach. At the beginning, the review dataset is processed to extract the row text from the whole dataset and, then to remove the rows of the table with empty reports. Subsequently, we convert the labels in the dataset opinion column of the table to categories. Then, we remove words from the bag-of-words model that do not appear more than twice in total. We then remove any documents containing no words from the bag-of-words model and remove the corresponding entries in the labels.

4.3.2 Second step: Twitter sentiment dataset analysis and visualization

In this step, we visualize Twitter's opinion based on the histogram. The histogram visualization is a common analysis that is used for visualizing a large number of tweets as sentiment analysis. To implant this, some important steps are implemented; as shown in Algorithm 4.4.

Algorithm 4.4 : Twitter dataset pre-processing based NLP

Input: Twitter Data in CSV file
Output: Class Type

1. **Import** Dataset
2. **Import** the Twitters reports data. This data contains labeled textual descriptions of Twitters opinion. That are illustrated in (airline sentiment).
3. **To Import** the text data as string arrays, specify the text type to be 'string'.
4. **Pre-process** the Dataset
5. **Remove** the rows of the table with empty reports.
6. **Convert** the labels in the event type column of the table to categorical.
7. **View** the distribution of the classes in the data using a histogram.
8. **Get** the frequency counts of the classes and their names from the histogram.
9. **Erase** the punctuation, then make everything in the lower case.
10. **Replace** hashtag abbreviations with standard language
11. **Tokenizer** the text data.
12. **Create** a bag-of-words model from the tokenized documents.

4.3.3 Deep-learning prediction and classification

Data-mining approaches are used in this step for mining the user's opinion and its classification. In this task, we use data-mining techniques such as dimensionality reduction and feature selection, as well as a supervised learning classification framework. To implant this, some important steps are implemented, as follows.

Pre-processing of the text data To pre-process the training data, the punctuation is deleted, the text is converted to lowercase, and then it is tokenized. Words are not stemmed or removed, as this can be detrimental to the word-embedding fit.

Conversion of the document to sequences The documents are inputted into the DL network and converted into sequences of word vectors. In this case, the mini batches are created automatically during the training phase. Also, the same length of is adaptive by padding, truncating, or either by splitting the whole input dataset. However, this option is not well suited to the sequences of word vectors that are used in our "tweets dataset." Therefore, instead, the pad and truncate functions of the sequences data are manually applied. The training documents are truncated to a length of 75 words. The anonymous function inputted to the documents function takes the string array input and outputs the first 75 elements. The documents are converted to sequences of word vectors. To convert the training documents into a cell array of sequences, we use the example function doc2sequence, shown at the end of this example. The columns of each sequence are the word vectors. To pad sequences of word vectors for DL network model, you must leave pad the sequences. The sequence-padding option for DL network model, by default, right-pads the sequences, so this must be performed manually.

Sentiment Detection and Opinion Mining Using a Convolutional Neural Network

In general, opinion mining using machine learning consists of many tasks and functions such as opinion clustering, opinion conception, entity extraction, opinion summarization, and opinion classification. The very important techniques of text mining with ML tools are text classification, also known as text categorization. Text classification is defined as the task of automatically distinguishing the unlabeled documents into pre-defined classes [DL18]. The opinion classification approach, as a task of machine learning, should clearly pre-define the mathematical model. In addition to the mathematical model, in this case, we should pre-define the feature-extraction model that is primarily used for extracting the feature vector, which is used later to feed the machine learning model [SMKR13]. A CNN is performed based on the prediction value of the truth labels: The "supervised learning approach. Moreover, the true label y of the training data x plays a very significant role in the CNN that is by measuring the actual loss function based on using either "norm1 L1" or "norm2 L2" that assigns the penalty "prediction difference" to classification prediction errors (the difference between the truth label and prediction value) [DAYD17], as shown in Figure 4.8. The depth of the convolutional layer in the CNN is based on the specific amount of kernels "filters" that are used on the input data to generate the feature map. In addition, the stride number of convolutional kernel "filters" allows the CNN filter to extend during the sliding process and extend the data size dimensions [Kan11]. However, the zero-padding, which is the padding process around the data borders, allows the input to be preserved at the same size. Max-pooling is used to reduce the dimension of the feature map [LSD15]. Figure 3.19 illustrates a max-pooling operation with 2×2 filters and provides an example of a max 2×2 pooling layer that it is used to reduce the filter size. Finally, the fully connected layer is defined to connect all the residue that comes from the last convolutional layer, as shown in Figure 4.23.

The mathematical definition of the opinion-mining and classification model relies on the demonstration of documents to attain the ability to correctly execute a categorization task. Moreover, text classification may define it, as given in Equation (4.10). [MK20].

$$x = (x_1, x_2, \ldots, x_n) \tag{4.10}$$

where n represents the total number of classified documents that have been used to categorize the whole data task [GC17]. As the classification task for the text classification is a main part of the supervised learning discussed above, the machine-learning classification model for text classification is provides a training set example. The training set example has an an input with associated labels or target output. Essentially, the training

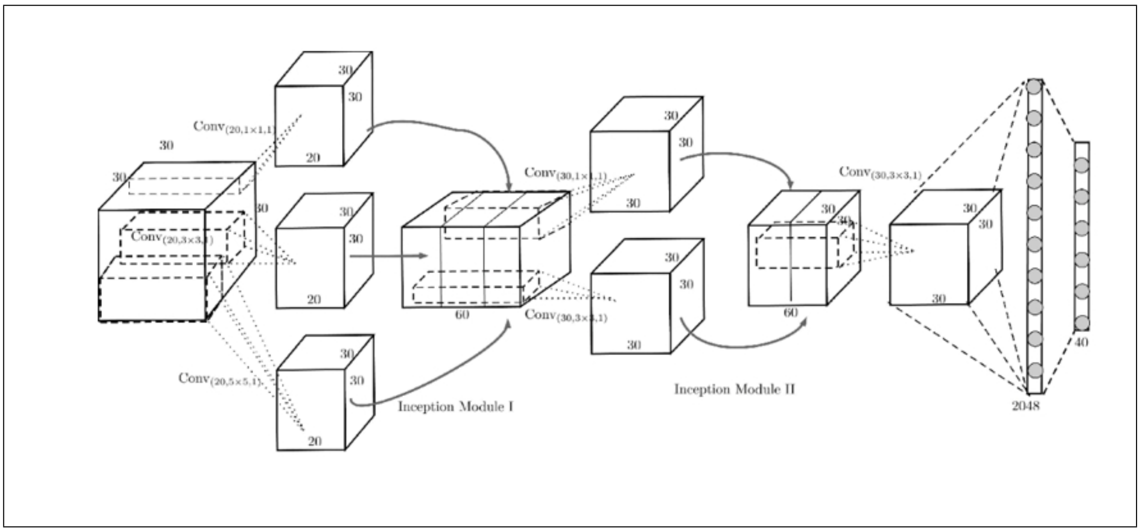


Fig. 4.8 : Convolutional neural network (deep learning) structure [SLX+15].

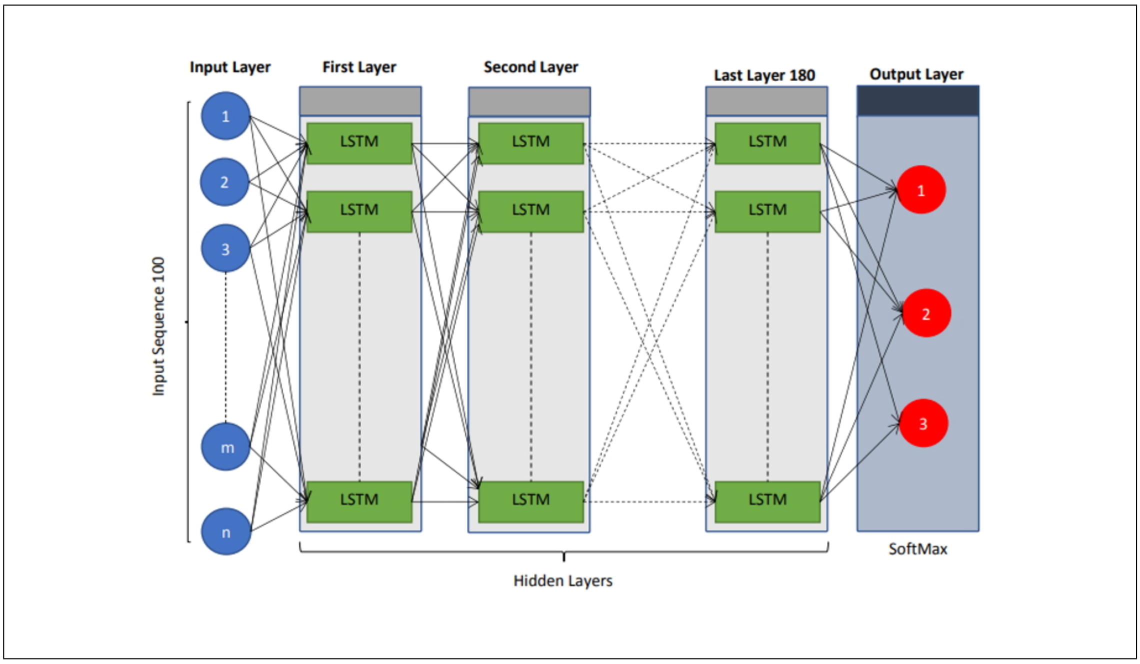


Fig. 4.9 : Deep learning twitter opinion mining classification approach.

examples are in the form of the attribute vectors, so the input is a subset of Rn [GC17]. In this case, it is assigned to the positive class, if the values of the function of X are greater than zero and to the negative class if the value of the function of x are less than zero as given in Equation (4.11) [GC17].

$$Output = \begin{cases} + & iff(x) \geq 0; \\ - & otherwise. \end{cases} \tag{4.11}$$

The function $f(x)$ in this case is represented as a main decision function. That means in each vector that has a target element, the target label is defined in $Y \in \{-1, +1\}$. which

Table 4.2 : Architecture of the convoluted neural network on both training and testing phase for the classification model.

Layer Number	Layer Type	Parameters	
		Dimension	Number of Parameters
Layer 1	Input Sequence (Histogram of Words)	100	100
Layer 2	Word Embedding Layer	100	61495
Layer 3	LSTM	100	190 Hidden Layer
Layer 4	Fully Connected Layer	100	-
Layer 5	Prediction Layer (SoftMax)	3	-
Layer 6	Classification Output	1	Cross Entropy

+1, -1 represent both the positive and negative classes [GC17]. A typical model for machine learning learns the mapping function, which implies $X \Rightarrow Y$, which in this case can be represented by a set of possible learned mapping, such as that given in Equation (3.58). The method that can be used to train a fully connected layer system of CNN is the supervised learning approach

Design and train the deep learning network

The DL network architecture is defined. A DL network layer is included and the output size specified at 180. The DL structure model is described in Table 4.2. The general design of the DL model that has been used in our approach is illustrated in Figure 4.23.

The activation function of the hidden layers is based on Equations (3.57) and (3.58) using the LSTM model described in Chapter 3, Section 3.5.3, using residual neural network training algorithm described in Algorithm 3.1, while the activation function of the output layer is the SoftMax function described in Equation (3.53).

4.4 Sentiment Analysis Dataset

4.4.1 U.S. airlines dataset

The first dataset that is used in this chapter (NLP-based Sentiment Analysis for opinion-mining Prediction) is obtained from the Twitter U.S. Airline Sentiment Dataset in social media [144]. The dataset was primarily collected for a sentiment analysis job. It focuses on the main and major problems that the U.S. airlines face and was collected from Twitter in February 2015. The participants in this analysis based on Twitter data, were asked to first classify their airlines as "positive", "negative," or "neutral," and then to give the reason why they classified it as such as *for example, "late flight" or "rude service"). The dataset has 14,640 rows and 15 columns. It also includes a features set: The tweet ID, the sentiment value, the sentiment confidence score value, the negative reason for the bad tweets, the negative reason confidence value, the airline name, the sentiment gold value, the name, the retweet count value, the tweet text value, the tweet coordinates value, the time of tweet, the date of tweet, the tweet location, and finally the user's time zone.

4.4.2 Amazon review dataset

The second dataset that is used in this chapter is the amazon review dataset, which includes different features, such as the People's opinions about the product after purchasing Which, we get from the metadata in Amazon. The dataset is essentially having over 142.8 million data points (reviews) that are collected between May 1996 and July 2014. The main component of the dataset is distributed as first the product rating which has different data types such as text and votes to rank and describe the products. Second, product metadata has also both text description and classified category to rank the product too. The other part is listing as pricing, brand, an image for each product, and finally a URL link for the product. In these experimental results, we used the small version of this dataset, which contains 5000 reviews and is available free of charge for academic research [145].

4.5 Evaluation Metric

The performance of the opinion-mining detection and classification system using different dimensionality reduction. Is measured using Standard classification criteria such as "accuracy," "precision," and "recall" calculated using the confusion matrix component. The evaluating performance of the proposed system is calculated using three measures: Recognition Rate (RR), Precision (PR), Sensitivity (SE), and Specificity (SP) [AAASCH18, LG15], as are given in Equations (4.12), (4.13), (4.14), and (4.15), respectively.

Recognition rate (accuracy):

The RR is given in Equation (4.12) [AAASCH18, LG15].

$$Accuracy = \frac{TP}{TP+TN} * 100 \tag{4.12}$$

F-measure:

The F-measure is given in Equation (4.13) [AAASCH18, LG15].

$$F - Measure = \frac{TP}{TP+TN} \tag{4.13}$$

Detection rate:

The detection rate is given in Equation (4.14) [AAASCH18, LG15].

$$Detection\ Rate = \frac{TN}{TN+FN} \tag{4.14}$$

False alarm: The False Alarm is given in Equation (4.15) [AAASCH18, LG15].

$$False\ Alarm = \frac{TP}{TP+FN} \tag{4.15}$$

We use 5-fold cross-validation, and the best way to select the folds is by down-sampling. We set our test size to a maximum of 10% of the dataset. This near-equality allows for a more accurate evaluation of the resulting classifiers. The training set is then divided into five folds, each fold including randomly selected images. One-fold is withheld for the validation step. The same previously split folds are used to train and validate our classifiers, and the performance of the trained classifier is decided by the votes collected from the classifier of each fold.

4.6 First Model Experimental Results Using the U.S. Airlines Dataset

The results of our proposed system is obtained and evaluated in this section. This work is implemented using Matlab 2017a supported in the Windows-10 operating system. Different evaluation results have been extracted based on two main categories. The Twitters opinion statistical observation and visualization and the Twitter opinion prediction and classification. In our experiments, we compare each dimensionality reduction algorithm in turn to assess the validation and accuracy of the experimental results compared with our approach. For more precisely indication results, the confusion matrix of size $m \times m$ is designed for detection results based on the total number of classes in our final datasets.

4.6.1 Twitter airlines dataset statistical observation

Whole data observation The first experimental result is the whole Twitter opinion based on the class (mood) distribution for the whole dataset. Figure 4.10 and Table 4.3 show the sentiment analysis (opinion histogram) and the statistical measurement for the whole Twitter airlines dataset based on the observation using the statistical measurement for each mood class.

Table 4.3 : Whole airlines statistical observation.

No	Opinion Class	Opinion Class
1	negative	9178
2	neutral	3099
3	positive	2363
Total		14,640

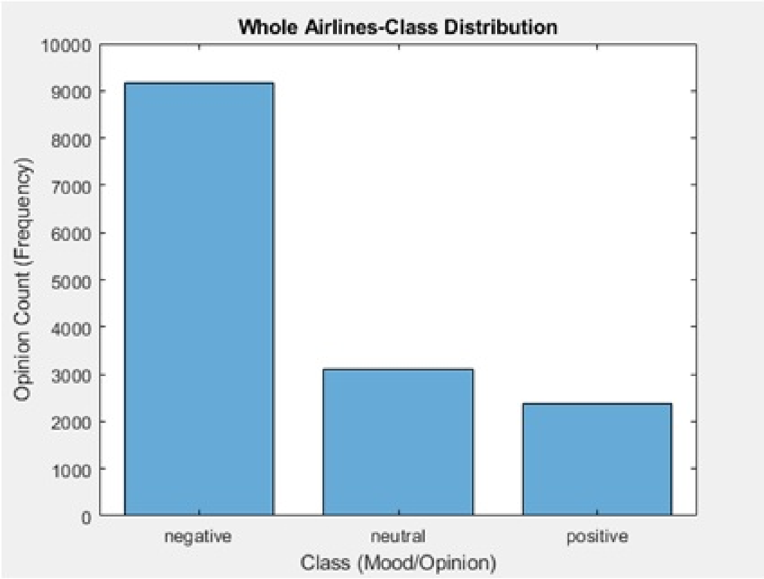


Fig. 4.10 : Whole dataset opinion statistical observation.

Visualization of tweets based on location In this experimental result, we observe the tweet locations for the whole dataset as an indication of the tweet opinion locations based on word cloud visualization. Figure 4.11 shows the cloud visualization based tweets on location.

Fig. 4.11 : Word cloud visualization of each tweet location.

Tweet extraction for each individual airline In this experiment, we extract the statistical observation of each individual airlines. Figure 4.12 and Table 4.4 show the sentiment analysis (opinion histogram) of the Twitter mood based on the Twitter opinion for each individual airline.

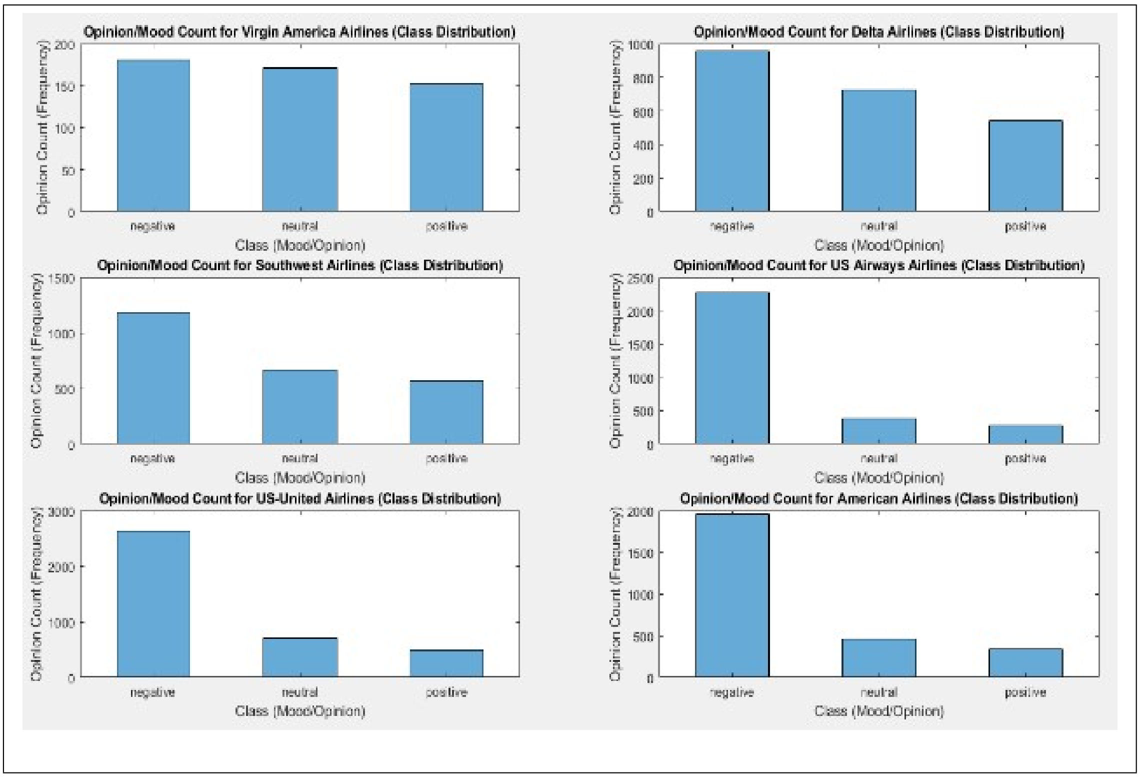
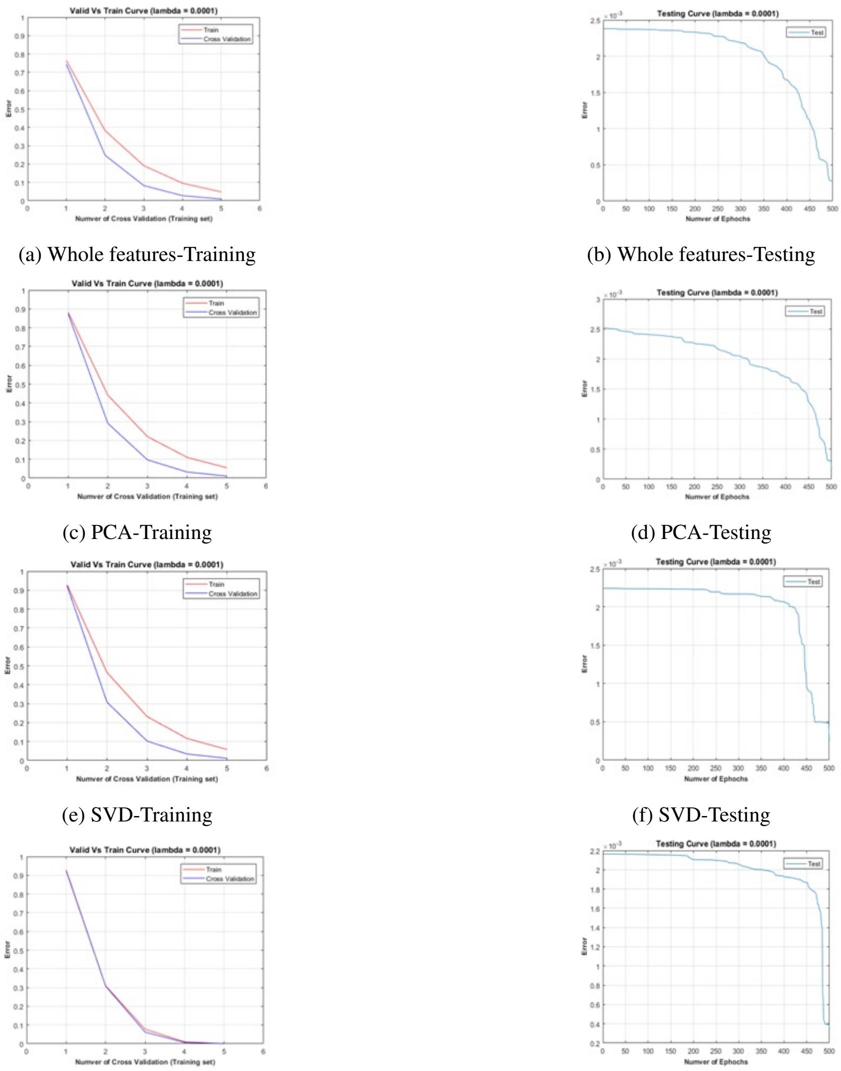


Fig. 4.12 : Sentiment analysis (opinion histogram) of the tweets mood for each individual airlines.

4.6.2 NLP-based tweet opinion prediction and classification performance results

The experimental results for the Twitter opinion-mining approach using BNNs show that our dimensionality reduction approach reduces the original data dimension from 2976 to 532 and achieved a highest accuracy of 93.49% in training and 94.88% in testing. By comparing our results with other dimensionality reduction algorithms (PCA and SVD) using the same dimensions we note that the highest accuracy is lower than our results by 5.45%. The overall accuracy (cross-validation and testing accuracy) for each algorithm is shown in Figure 4.13.

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using the whole features space are showing in Tables 4.5-4.8.



(a) Whole features-Training (b) Whole features-Testing

(c) PCA-Training (d) PCA-Testing

(e) SVD-Training (f) SVD-Testing

(g) Our Approach-Training (h) Our Approach-Testing

Fig. 4.13 : Overall performance results of U.S. airlines opinion mining approach.

Table 4.4 : Statistical observation of each airlines.

Airlines Name	Statistical Observation	
Virgin_America	Negative	181
	Neutral	171
	positive	152
Delta	Negative	955
	Neutral	723
	positive	544
Southwest	Negative	1186
	Neutral	664
	positive	570
US_Airways	Negative	2263
	Neutral	381
	positive	269
United	Negative	2633
	Neutral	697
	positive	492
American	Negative	1960
	Neutral	463
	positive	336

Table 4.5 : Confusion matrix of opinion mining prediction approach during the training phase using the whole features.

	Class1	Class2	Class3
Class1	8177	540	630
Class2	439	1935	501
Class3	562	624	1232

Table 4.6 : The performance results (each class) of the opinion mining approach during the training phase using the whole features.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9187	9347	85.17%	0.87	0.89	0.88
Class 2	3099	2875	85.63%	0.67	0.62	0.65
Class 3	2363	2418	84.17%	0.51	0.52	0.52

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using our dimensionality reduction and features selection method are showing in Tables 4.9-4.12.

Table 4.7 : Confusion matrix of opinion mining prediction approach during the testing phase using the whole features.

	Class 1	Class 2	Class3
Class 1	8528	411	292
Class 2	201	2157	193
Class 3	449	531	1878

Table 4.8 : The performance results (each class) of the opinion mining approach during the testing phase using the whole features.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9187	9231	90.76%	0.92	0.93	0.93
Class 2	3099	2551	90.87%	0.85	0.70	0.76
Class 3	2363	2858	89.99%	0.66	0.79	0.72

Table 4.9 : Confusion matrix of opinion mining prediction approach during the training phase using our dimensionality reduction and features selection method.

	Class 1	Class 2	Class3
Class 1	8849	260	413
Class 2	106	2536	70
Class 3	232	303	1880

Table 4.10 : The performance results (each class) of the opinion mining approach during the training phase using our dimensionality reduction and features selection method.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9187	9522	93.1%	0.93	0.96	0.95
Class 2	3099	2712	94.96%	0.94	0.82	0.87
Class 3	2363	2415	93.05%	0.78	0.80	0.79

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using PCA algorithm are showing in Table 4.13-4.16.

Finally, the experiential results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using SVD algorithm are showing in Table 4.17-4.20.

Table 4.21 shows the overall performance results of the propose system opinion mining detection and classification approach in both training and testing. We can notice that our dimensionality reduction and feature selection approach has satisfied the highest

Table 4.11 : Confusion matrix of opinion mining prediction approach during the testing phase using our dimensionality reduction and features selection method.

	Class 1	Class 2	Class 3
Class 1	8921	269	219
Class 2	88	2675	39
Class 3	169	155	2105

Table 4.12 : The performance results (each class) of the opinion mining approach during the testing phase using our dimensionality reduction and features selection method.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9178	9409	94.91%	0.95	0.97	0.96
Class 2	3099	2802	96.24%	0.95	0.86	0.91
Class 3	2363	2429	96.02%	0.87	0.89	0.88

Table 4.13 : Confusion matrix of opinion mining prediction approach during the training phase using our dimensionality reduction and features selection method.

	Class 1	Class 2	Class3
Class 1	8341	398	323
Class 2	459	2310	389
Class 3	378	391	1651

Table 4.14 : The performance results (each class) of the opinion mining approach during the training phase using our dimensionality reduction and features selection method.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9187	9062	89.36%	0.92	0.91	0.91
Class 2	3099	3158	88.82%	0.73	0.75	0.74
Class 3	2363	2420	89.88%	0.68	0.70	0.69

Table 4.15 : Confusion matrix of opinion mining prediction approach during the testing phase using our dimensionality reduction and features selection method.

	Class 1	Class 2	Class3
Class 1	8548	378	391
Class 2	299	2291	271
Class 3	331	430	1701

accuracy comparing with the other algorithms such as PCA and the SVD, comparing with the original feature space.

Table 4.16 : The performance results (each class) of the opinion mining approach during the testing phase using our dimensionality reduction and features selection method.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9178	9317	90.44%	0.92	0.93	0.92
Class 2	3099	2861	90.59%	0.80	0.74	0.77
Class 3	2363	2462	90.28%	0.69	0.72	0.71

Table 4.17 : Confusion matrix of opinion mining prediction approach during the training phase using SVD algorithm.

	Class 1	Class 2	Class3
Class 1	8202	420	221
Class 2	314	2090	239
Class 3	662	589	1903

Table 4.18 : The performance results (each class) of the opinion mining approach during the training phase using SVD algorithm.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9187	8843	88.95%	0.93	0.89	0.91
Class 2	3099	2643	89.33%	0.79	0.67	0.73
Class 3	2363	3154	88.31%	0.60	0.81	0.69

Table 4.19 : Confusion matrix of opinion mining prediction approach during the testing phase using SVD algorithm.

	Class 1	Class 2	Class3
Class 1	8285	414	110
Class 2	303	2098	128
Class 3	590	587	2125

Table 4.20 : The performance results (each class) of the opinion mining approach during the testing phase using SVD algorithm.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	9178	8809	90.32%	0.94	0.90	0.92
Class 2	3099	2529	90.22%	0.83	0.68	0.75
Class 3	2363	3302	90.33%	0.64	0.90	0.75

Table 4.21 : Overall accuracy.

Algorithm	Data Dimension	Training				Testing			
		Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Non	2976	85.00%	0.68	0.68	0.68	90.54%	0.81	0.8	0.8
Ours	532	93.70%	0.88	0.86	0.87	94.8%	0.92	0.9	0.92
PCA	532	89.40%	0.78	0.79	0.78	90.40%	0.8	0.8	0.8
SVD	532	88.87%	0.77	0.79	0.78	90.29%	0.8	0.83	0.81

4.6.3 Comparison with other prediction approaches

We compare with the performance results of our approach, which is based on a mechanism involving several consecutive steps of sentiment analysis using a proper ANN and dimensionality reduction, with those of data-mining and machine-learning approaches. Table 4.22 shows the performance results for the Twitter U.S. Airlines. Sentiment analysis, using different approaches, such as a decision tree, which achieves 63%, random forest 85.6%, SVM 81.2%, adaBoost 84.5%, and logistic regression 81%. Our approach achieves better results (94.8%) compared with other methods implemented on the same dataset.

Table 4.22 : Overall performance results comparing with different data mining and machine learning approaches.

Methods	**Accuracy**
Decision Tree [RK18]	63%
Random Forest[RK18]	85.6%
SVM [RK18]	81.2%
AdaBoost [RK18]	84.5%
Logistic Regression[RK18]	81%
KNN [RK18]	59%
Our method based DR [AGAAL19]	94.8%

4.7 First Model Experimental Results Using the Amazon Review's Dataset

4.7.1 Amazon review's statistical observation

The second experimental result is for the Amazon review, based on the voting distribution for whole dataset. Figure 4.14 and Table 4.23 show the statistical analysis (voting histogram) and the statistical measurement for the whole Amazon, dataset based on the observation using the statistical measurement for each voting class. We select the small dataset version that contains 5000 review. We pre-process the dataset and assign each

Table 4.23 : Whole Amazon review statistical observation.

No	Opinion Class	Opinion Class
1	negative	63
2	negative	54
3	neutral	196
4	neutral	1207
5	positive	3476
Total		5000

voting count to one of our class, which is "negative", "neutral", "positive" as is shown in Table 4.23.

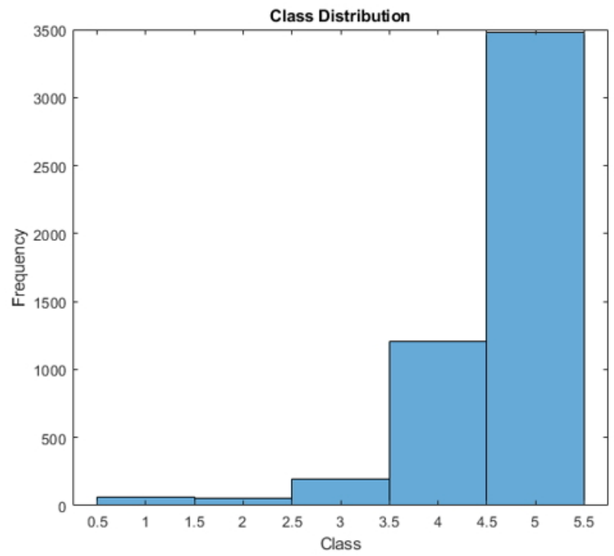
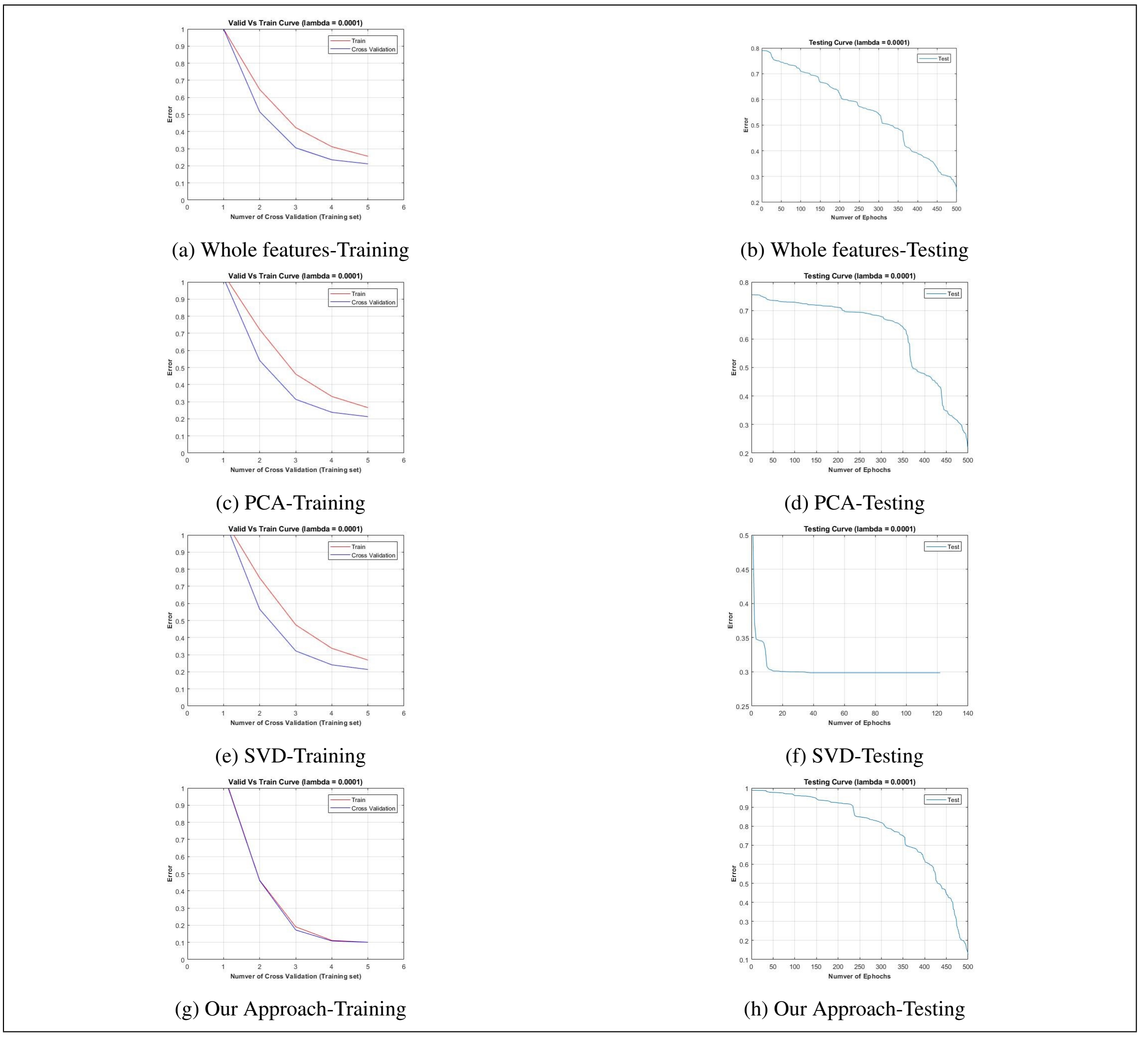


Fig. 4.14 : Word cloud visualization of each tweet location.

4.7.2 NLP-based Amazon review's dataset prediction and classification performance results

The experimental results for the Amazon review's opinion-mining approach using BNNs show that our dimensionality reduction approach achieves a highest accuracy of 93.9% in training and 94.8% in testing. By comparing our results with other dimensionality-reduction algorithms (PCA and SVD) using the same dimensions we notice that the highest accuracy is less than for our results. The overall accuracy (cross-validation and testing accuracy) for each algorithm is shown in Figure 4.15.

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using the whole features space are provided in Tables 4.24-4.27.



(a) Whole features-Training (b) Whole features-Testing

(c) PCA-Training (d) PCA-Testing

(e) SVD-Training (f) SVD-Testing

(g) Our Approach-Training (h) Our Approach-Testing

Fig. 4.15 : Overall performance results of the amazon review opinion mining approach.

Table 4.24 : Confusion matrix of Amazon review's mining prediction approach during the training phase using the whole features.

	Class 1	Class 2	Class 3
Class 1	63	366	450
Class 2	16	522	250
Class 3	14	233	2082

Table 4.25 : The performance results (each class) of the amazon review's mining approach during the testing phase using the whole features.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	93	879	78.83%	0.072	0.68	0.13
Class 2	1121	788	78.35%	0.66	0.47	0.55
Class 3	2782	2329	76.30%	0.89	0.75	0.81

Table 4.26 : Confusion matrix of amazon review's mining prediction approach during the testing phase using the whole features.

	Class 1	Class 2	Class 3
Class 1	8	40	70
Class 2	6	212	136
Class 3	12	20	490

Table 4.27 : The performance results (each class) of the amazon review's mining approach during the testing phase using the whole features.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	26	118	87.25%	0.068	0.31	0.11
Class 2	282	354	78.88%	0.6	0.75	0.67
Class 3	696	532	75.30%	0.92	0.7	0.8

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using our dimensionality-reduction and features-selection algorithm are provided in Tables 4.28-4.31.

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using PCA algorithm are showing in Table 4.32-4.35.

The experimental results of each approach in the training and testing phase in additional to the confusion matrix, precision, recall, F1 score, and accuracy of each class using SVD algorithm are showing in Table 4.36-4.39.

Table 4.40 shows the overall performance results of the proposed system opinion mining detection and classification approach in both training and testing. We can notice that our dimensionality reduction and feature selection approach has satisfied the highest accuracy comparing with the other algorithms such as PCA and the SVD, comparing with the original feature space.

Table 4.28 : Confusion matrix of Amazon review's mining prediction approach during the training phase using our dimensionality reduction and features selection method.

	Class 1	Class 2	Class 3
Class 1	90	90	180
Class 2	1	901	60
Class 3	2	130	2542

Table 4.29 : The performance results (each class) of the Amazon review's mining approach during the training phase using our dimensionality reduction and features selection method.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	93	360	93.17%	0.95	0.97	0.9
Class 2	1121	962	92.97%	0.94	0.8	0.87
Class 3	2782	2674	90.69%	0.95	0.91	0.93

Table 4.30 : Confusion matrix of Amazon review's mining prediction approach during the testing phase using our dimensionality reduction and features selection method.

	Class 1	Class 2	Class 3
Class 1	15	10	25
Class 2	6	267	15
Class 3	3	5	646

Table 4.31 : The performance results (each class) of the Amazon review's mining approach during the testing phase using our dimensionality reduction and features selection algorithm.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	24	50	95.56%	0.3	0.63	0.41
Class 2	282	288	96.37%	0.93	0.95	0.94
Class 3	686	654	95.16%	0.99	0.94	0.96

Table 4.32 : Confusion matrix of Amazon review's mining prediction approach during the training phase using PCA.

	Class 1	Class 2	Class 3
Class 1	10	397	480
Class 2	62	638	560
Class 3	21	86	1742

Table 4.33 : The performance results (each class) of the Amazon review's mining approach during the training phase using PCA.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	93	887	75.98%	0.011	0.11	0.02
Class 2	1121	1260	72.35%	0.51	0.57	0.02
Class 3	2782	1849	71.30%	0.94	0.63	0.75

Table 4.34 : Confusion matrix of Amazon review's mining prediction approach during the testing phase using PCA.

	Class 1	Class 2	Class 3
Class 1	3	138	127
Class 2	4	110	72
Class 3	17	34	497

Table 4.35 : The performance results (each class) of the Amazon review's mining approach during the testing phase using PCA.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	24	268	71.46%	0.011	0.13	0.021
Class 2	282	186	75.25%	0.59	0.39	0.47
Class 3	696	548	75.05%	0.91	0.71	0.8

Table 4.36 : Confusion matrix of Amazon review's mining prediction approach during the training phase using SVD.

	Class 1	Class 2	Class 3
Class 1	30	361	362
Class 2	35	518	421
Class 3	29	242	1999

Table 4.37 : The performance results (each class) of the Amazon review's mining approach during the training phase using SVD.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	93	753	80.33%	0.04	0.32	0.071
Class 2	1121	974	73.50%	0.53	0.46	0.49
Class 3	2782	2269	73.65%	0.88	0.72	0.79

Table 4.38 : Confusion matrix of Amazon review's mining prediction approach during the testing phase using SVD.

	Class 1	Class 2	Class 3
Class 1	2	140	132
Class 2	6	116	68
Class 3	17	26	496

Table 4.39 : The performance results (each class) of the Amazon review's mining approach during the testing phase using SVD.

Classes	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
Class 1	24	274	70.59%	0.0073	0.08	0.013
Class 2	282	190	76.07%	0.61	0.41	0.49
Class 3	696	539	75.77%	0.92	0.71	0.8

Table 4.40 : Overall accuracy.

Algorithm	Training				Testing			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Non	77.80%	54.70%	63.30%	49.70%	80.50%	52.90%	58.70%	52.70%
Ours	93.1%	94.7%	89.30%	90%	95.0%	94.7%	90.70%	93.70%
PCA	73.20%	48.70%	43.60%	26.30%	73.90%	50.40%	41.00%	43%
SVD	75.80%	48.30%	50%	45%	74.10%	51.20%	40%	43.40%

4.7.3 Comparison with other approaches

We compae the performance results of our approach, which based on a mechanism that involves several consecutive steps of sentiment analysis using a proper ANN and dimensionality reduction, with those of several data-mining and machine-learning approaches and a deep- learning approach. Table 4.41 shows the performance results for the Amazon review opinion-mining results using different approaches: Naive base (NB), linear SVM, non-linear SVM, and KNN with different window sizes, such as N=4, 5, and 6. Moreover, we compare our results with the LSTM, linear naive base, SVM, and KNN with and without Glove processing. However, our approach attains a better result, achieving 94.8% than the other methods implemented on the same dataset, for which the highest score, which is achieved by the DL approach, is 81.29%.

Table 4.41 : Overall performance results comparing with different data-mining and machine-learning approaches.

Methods	Accuracy
Deep Learning Sentiment Anlysis [SN19]	81.29%
Multinomial NB [TWX19]	70.60%
Linear SVM [TWX19]	69.60%
RBF SVM [TWX19]	69.20%
KNN-4 [TWX19]	61.70%
KNN-5 [TWX19]	65.40%
KNN-6 [TWX19]	64.60%
LSTM [TWX19]	71.50%
Gaussian NB w/ Glove [TWX19]	52.40%
Linear SVM w/ Glove v	68.60%
KNN-4 w/ Glove [TWX19]	57.60%
KNN-5 w/ Glove [TWX19]	62.20%
KNN-6 w/ Glove [TWX19]	61.60%
LSTM w/ Glove [TWX19]	70.20%
LSTM w/ Glove(Resample) [TWX19]	65.60%
[AAAL19] our method based DR	95.0%

4.8 Second Model Experimental Results Using the U.S. Airlines Dataset

The performance of the opinion mining detection and classification system using a DL approach can be evaluated using various parameters. Technically, more significant time is needed and consumed when the model is training on the CPU rather than using the GPU to train the same mode. To train the DL network first, we input the documents into a DL network structure (LSTM network), by converting the documents into sequences of word vectors. During the DL model (Deep Learning Network) training phase, the same length of different input/output feature maps is generated by using padding or truncating. In another technique, splitting the input feature map is used to quantize them in the same size. Instead, there is a need to pad and truncate the sequences manually. The training setting parameters that are used in our deep learning model are shown in Table 4.42.

Table 4.42 : Training setting parameters in our DL model.

Training Parameter	**Setting Value**	
Momentum	0.9000	
Initial Learning Rate	0.0100	
Learning Rate Schedule Setting	Drop Rate Factor	0.2
	Drop Period	5
L2 Regularization	1.0000e-04	
Max Epochs	20	
Mini Batch Size	32	
Verbose	1	
Verbose Frequency	50	
Validation Data Frequency	50	
Validation Patience	5	
Shuffle	1	
Sequence Padding Value	0	

The first step of training the deep tweets opinion mining prediction is to choose the data (Twitter text) target length as shown in Figure 4.16.

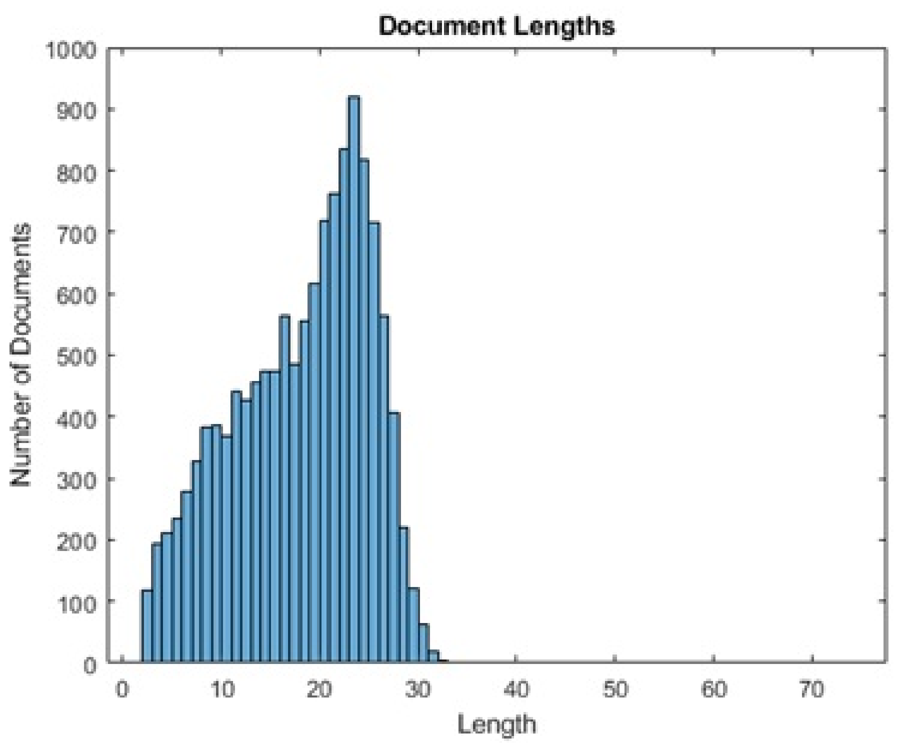


Fig. 4.16 : Sequence training documents data conversation.

The overall training progress of the deep Twitter opinion-mining approach using deep learning approach and the loss function plot are shown in Figure 4.17 and 4.18.

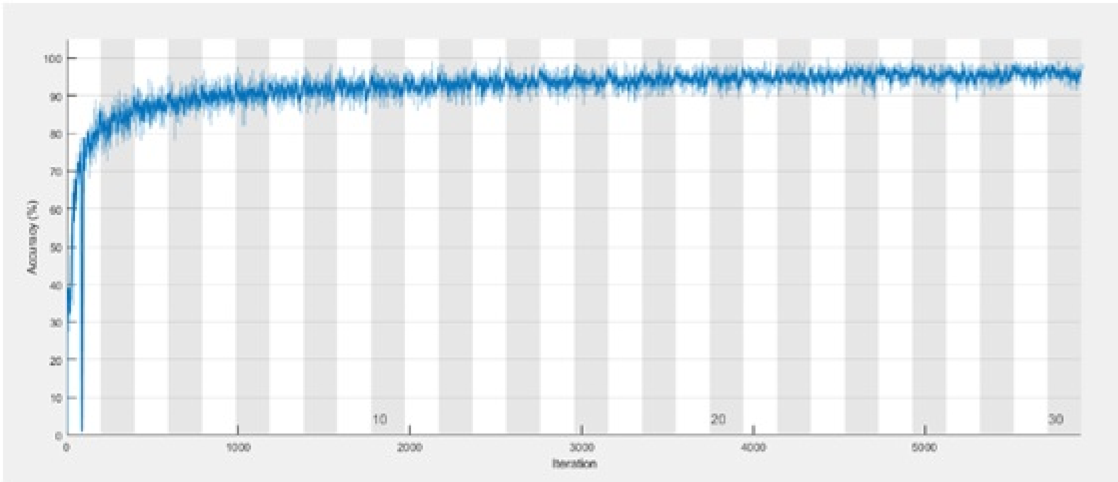


Fig. 4.17 : Overall performance results of the deep opinion mining approach (Training Accuracy).

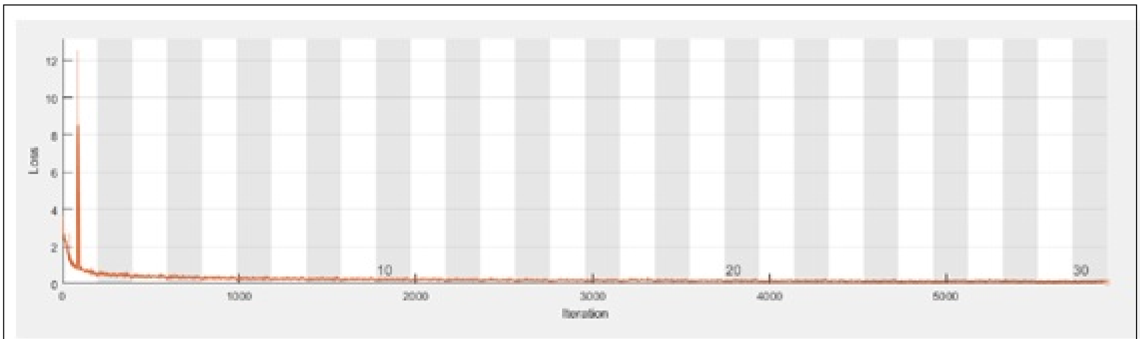
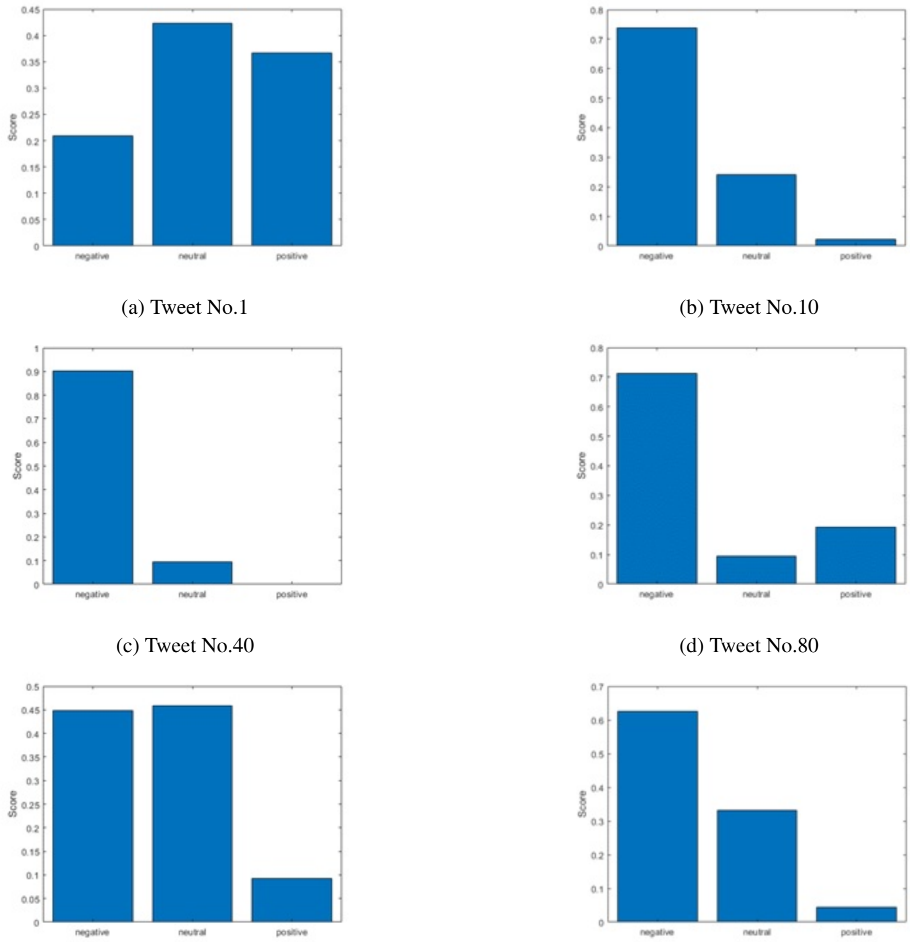


Fig. 4.18 : Overall performance results of the opinion mining approach (Lost function).

4.8.1 Deep Twitter opinion-mining prediction model

To predict and classify the Twitter mood based the trained deep learning model for the testing dataset (Twitter text data), we use the trained parameters to predict each tweet mood using the deep learning test model. To visualize the tweet opinion prediction (mood), we visualize the top three predictions and their scores for each tweet in the testing data, sort the prediction scores, and select the top three values and visualize them. Figure 4.19 shows an example of the first tweets in the testing dataset, as well as the corresponding visualized scores of each predicted mood (opinion).

An example of the overall deep Twitter opinion mining prediction outcome is shown in Table.4.43 and the predicted mood is shown in Figure 4.20.



(a) Tweet No.1 (b) Tweet No.10

(c) Tweet No.40 (d) Tweet No.80

(e) Tweet No.100 (f) Tweet No.1300

Fig. 4.19 : Overall tested samples of the mood predict visualization.

Table 4.43 : Overall tested samples of the mood predict Visualization Results.

Twitter Test No.	**The Requested Twitter**	**The Predicted Mood**	**Actual Mood**	**Prediction Score %**
1	"Virgin America view of downtown los Angeles the Hollywood sign and beyond that rain in the mountains"	neutral	neutral	0.517001
10	"Virgin America is it me or is your website down btw your new website isn't a great user experience time for another redesign"	negative	negative	0.737704
40	"united customer service is atrocious you have disrupted my travel plans you have lost my luggage and it is impossible to talk to a human"	negative	negative	0.902131
80	"Virgin America i spoke with a representative that offered no solution i am a loyal customer who flies on Virgin Atlantic as well"	negative	negative	0.713250
100	"united premier gold desk changes flight waives fees gives me wrong flight now Jana Acosta in salt lake refuses the same service angry"	negative	negative	0.713250
1300	"American air Delaney and Shawn at dfw showed exceptional customer service today will happily choose aa whenever possible now thank you "	positive	positive	0.625260

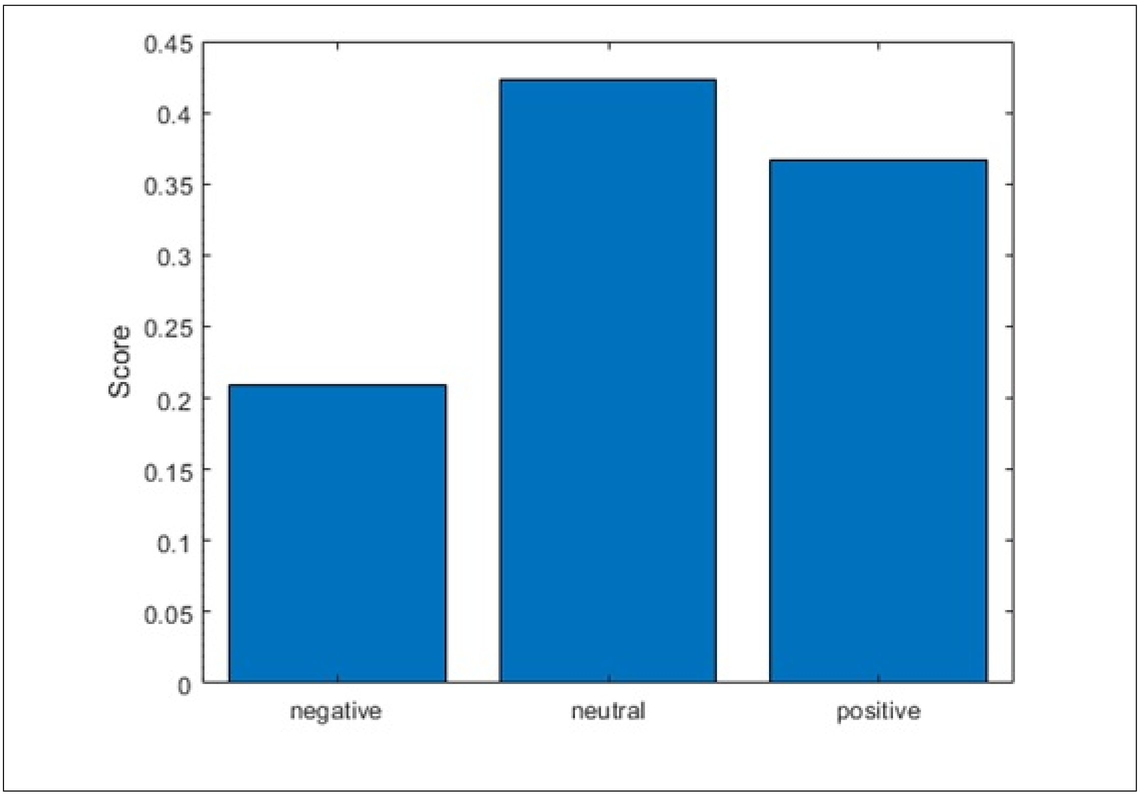


Fig. 4.20 : Predict visualization of the deep twitter opinion mining prediction model.

4.8.2 Comparison with other approaches

We compare the performance results of our approach, which is based on a mechanism that involves several consecutive steps of sentiment analysis using a proper ANN and dimensionality reduction, with those data-mining and machine-learning approaches. Table 4.27 shows the performance results for the Twitter U.S. airlines. Sentiment analysis dataset using different approaches, such as decision tree, which achieves 63%, random forest 85.6%, SVM 81.2%, AdaBoost 84.5%, logistic regression 81%. Our NLP-based sentiment analysis for Twitter opinion mining and visualization using dimensionality reduction and residual neural network achieves 94.8%. In addition, our approach obtains a better result (96.1%) compared with other methods implemented on same dataset.

Table 4.44 : Overall performance results comparing with different data-mining and machine-learning approaches.

Methods	Accuracy
Decision Tree [RK18]	63%
Random Forest[RK18]	85.6%
SVM [RK18]	81.2%
AdaBoost [RK18]	84.5%
Logistic Regression[RK18]	81%
Our method based DR [AGAAL19]	94.8%
Our method based Deep Learning [AAAL19]	96.1%

4.9 Summary

In this chapter, two different models of sentiment analysis and reviewers' opinion-mining prediction are proposed and implemented. Our approaches utilize data mining and machine learning to analyze and visualize users' opinion-mining prediction in different platforms. Two different standard datasets, the U.S. airlines and Amazon reviews datasets, are used to evaluate our models. Our first data-mining model-based on dimensionality reduction and ANN using our dimensionality reduction achieves 94.8% on the testing dataset, while the nearest comparative is the PCA, which achieves 90.34%. By comparing with other approaches on the same dataset (U.S. Airlines Sentiment), the highest score is achieved based on the random forest 85.6%. In addition, our second model (deep opinion mining) achieves even better results than our first model, achieving 96.1% compared with other methods implemented on the same U.S. airlines Sentiment dataset.

Chapter 5

Social Media Image Sentiment Analysis Based Deep Attention Learning Mechanisms

This chapter presents the concept of image sentiment analysis and emotion detection. Deep neural networks have been used based on the attention mechanism to enhance detection accuracy.

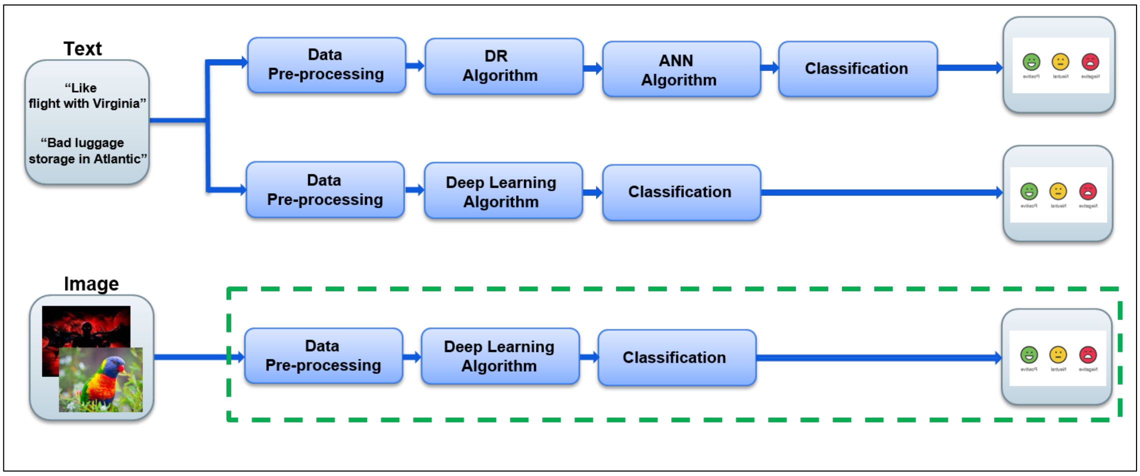


Fig. 5.1 : The highlighted part denotes how chapter 5 contributes to overall workflow of the general proposed system.

5.1 Emotion in Social Media Image

"Sentiment visualization" is a technique that has evolved and spread to deal with complex multidimensional datasets. The issue of visual sentiment analysis in social media involving images is quite new and challenging. In some cases, image emotion analysis may invoke different sentiments in different cultural or geographical circumstances. This chapter introduces a DL system model for analyzing and visualizing emotions in social media images through the classification and recognition of image-embedded sentiment patterns. Through utilizing an adequate technique for DL, namely the attention mechanism, this piece of work introduces a deliverable capable of extracting the implied sentimental status (embedded emotional responses) for each social media image as Happy or Sad. Therefore, the third contribution described in Section 1.6 is fulfilled in this chapter.

5.2 Proposed System

Based on the complexity of the sentiment image analysis in the social network, we propose a deep attention model (DAM) designed around a deep network that has an attention learning-based mechanism, which we call DAM for sentiment image analysis and emotion detection. The whole framework of the proposed network is illustrated in Figure 5.2. The main design of the deep network based on the attention mechanism feedback. Mainly, our design of the feedback (backpropagation attention mechanism) has two main stages. The first is the feed-forward attention stage and feeds the backward stage. In the first stage, the deep model learns and focuses on high-level image features that are extracted based on using very low-level features that also are adopted from the input images. The main features are extracted from the main significant image areas (details) by using the attention approach, and the feed-forward stage abstracts them as discriminative features. However, the second stage (feed backward) tries to acquire the low-level features in which the high-level features are returned in order to learn to extract more learned low-level features.

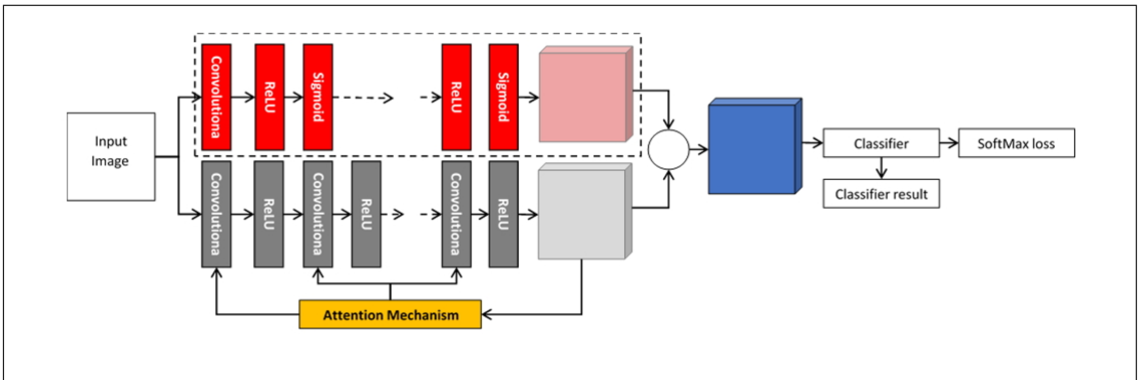


Fig. 5.2 : System flowchart of the deep attention-mechanism for sentiment image classification.

In more detail, the image features (high and low-level) are extracted using a stack of CNN blocks (convolutional layers). The CNN blocks (red and gray) are stacked together and comprise features based on convolution, pooling, and non-linear transformation. In another words, the red blocks in our design illustrate the original CNN blocks that are used primarily for high-level-feature extraction during the first feature-extraction stage (feed-forward pass), while the gray blocks illustrate the second stage (feed backward). The gray blocks (CNN's) use the high-level features that are extracted from the red blocks (CNN's) and stack together based on the attention mechanism, convolution, pooling, and non-linear transformation. The Attention term is defined as a type of action that guides directly to the object. In other words, it is defined as giving need which is the ability mind to allocate the uneven consideration across a field of sensation [SD13b]. Moreover, it helps to focus on and bring certain input to the core of the attention. In the same, diminishing or ignoring the others [Kau92]. Technically, in the neural network, attention action helps in terms of the credit assignment. The main challenge of this action is a long-range dependency, so the prediction becomes more impacting and also more affected by other facts [CS07]. The core probability model of the attention network is based on the Markov Assumption [CLZ17, TSN18], which aim to introduce a model that consists of different probability numbers, as given in Equation (3.65).

The attention network, in this case, can capture information in a human level [SMGS14]. The attention network mechanism is based on the sequence-to-sequence models, and, in this case, the design model can capture the essence of the entire input sequence in a single hidden state, as shown in Figure 5.3 [KDHR17].

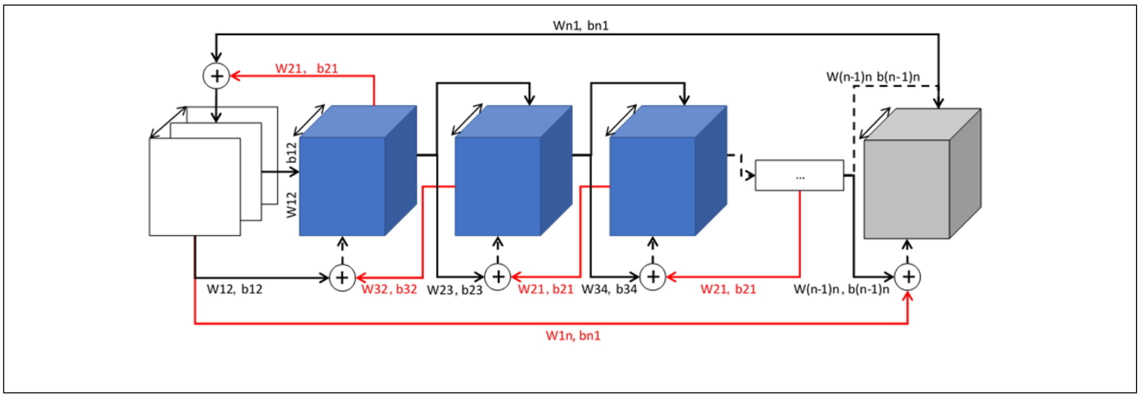


Fig. 5.3 : The attention learning mechanisms. The red line shows the attention layer feed batch for each epoch to combine each weight and bias for each layer; the black lines illustrate the regular feed epoch during each iteration for the whole deep learning attention model [KDHR17].

The structure of our deep attention network has nine double layers in total (18 layers). The first feed-forward structure has nine deep layers, in addition to the backward-feed layers, of which there are also nine. The first five layers in our structure are convolutional layers, followed by three fully connected layers. The SoftMax function is the main learning-based model that is used in last fully connected layer. The main SoftMax function is given in Equation (5.1) [KDHR17].

$$L_s = -\frac{1}{m}\sum_{i=1}^{m} log \frac{e^{W_{Yi}^T f_i + b_{Yi}}}{\sum_i^n e^{W_j^T f_j b_j}} \tag{5.1}$$

The reason for using the SoftMax function is that the SoftMax model provides a significant distribution that can distinguish between the two classes (Highly Positive/Highly Negative) in our binary classification problem. The whole structure of our DAM fundamentally maximizes the essential logistic objective of the multinomial regression, which is equivalent to maximizing the log-probability by maximizing the average of the logistic-function obtained. In this case, the Achievement of the deep attention network is based on the across of the final distribution, which is an arrangement for the final prediction made based on achieving the final labels of each problem class. The original image input size that feeds in to the first layer is $224 \times 224 \times 3$. Each image is 227 by 227 width by height with three channels, as the dataset includes colored images. The first convolutional layer of our DAM is constructed using 256 kernels; the density features map that is constructed from the first layer is $5 \times 5 \times 48$. The ReLU scheme is used in the output of the first layer (wholly-connected in the feed-forward synthesis), as presented in Equation (3.52). The attention mechanism is applied in each back-propagation synthesis, as shown in Equation (3.65). The second convolutional is constructed using 256 kernels; the density features map that is constructed from the first layer is $5 \times 5 \times 48$. The output map from the first layer is $11 \times 11 \times 3$ using stride 4 and padding 0.

Moreover, the second layer is the same first convolutional layer, using the same block based on the attention mechanism (back backward) convolutional layer. The second layer is another convolutional layer and is followed by the normalization layer (pooled and normalization layers). The final layer uses the SoftMax activation function as indicated in Equation (3.52). The full structure of our proposed system (DAM) for image sentiment analysis and classification is described in Table 5.1.

5.3 System Overview and Contribution

Our network design of the stacked DAM is based on using different residual learning approach. In the original one which is called ResNet is based on the residual learning formula as is illustrated as $H_{i,c}(x) = x + F_{i,c}(x)$, where $F_{i,c}(x)$ is denoted as the approximation learning function and x is the generated feature map that is generated using deep CNN [AAASCH18]. The main contribution of our design is based on using M(x) as a another function for feature selector. In this case, the attention mechanism is applied here to extract and select the relevant features during the learning process which we modified the original ResNet to pay more attention on the relevant learnable features only. The overview of our network design is illustrated in the Figure 5.4.

Table 5.1 : Proposed deep learning structure description.

Layer Number	Layer Type	Ker.	Size	Description
I1	Image Input	-	227x227x3	input image
C1	Convolution	96	11x11x3	4×4 stride and $0 \times 0 \times 0$ padding
R1	ReLU	-	-	ReLU activation function
A1	Attention Model	1	-	Attention
N1	Normalization	-	-	normalization
P1	Max Pooling	1	3x3	2×2 stride and $0 \times 0 \times 0$ padding
C2	Convolution	256	5x5x48	1×1 stride and $2 \times 2 \times 2 \times 2 \times 2$ padding
R2	ReLU	-	-	ReLU activation function
A2	Attention Model	1	-	Attention layer
N2	Normalization	-	-	normalization
P2	Max Pooling	1	3x3	2×2 stride and $0 \times 0 \times 0$ padding
C3	Convolution	384	3x3x256	1×1 stride and $2 \times 2 \times 2 \times 2 \times 2$ padding
R3	ReLU	-	-	ReLU activation function
A3	Attention Model	1	-	Attention layer
C4	Convolution	384	3x3x192	1×1 stride and $2 \times 2 \times 2 \times 2 \times 2$ padding
R4	ReLU	-	-	ReLU activation function
A4	Attention Model	1	-	Attention layer
C5	Convolution	256	3x3x192	1×1 stride and $2 \times 2 \times 2 \times 2 \times 2$ padding
R5	ReLU	-	-	ReLU activation function
A5	Attention Model	1	-	Attention layer
P6	Max Pooling	1	3x3	2×2 stride and $0 \times 0 \times 0 \times 0 \times 0$ padding
F7	Fully Connected	1	4096	fully connected layer
R7	ReLU	-	-	ReLU
A7	Attention Model	1	-	Attention
D7	Dropout	-	-	dropout
F8	Fully Connected	1	4096	fully connected layer
R8	ReLU	-	-	ReLU activation function
A8	Attention Model	1	-	Attention layer
D8	Dropout	-	-	dropout
F9	Fully Connected	1	4096	fully connected layer
R9	ReLU	-	-	ReLU activation function
A9	Attention Model	1	-	Attention layer
D9	Dropout	-	-	dropout

Our deep attention network has eight layers. The first five layers are the first convolutional layers. The fully connected layer is found in the last three layers, in addition, to the final output layer that has the SoftMax activation function. However, the other layers have the ReLU activation function instead of the SoftMax. The main reason behind using the SoftMax in the last fully connected layer is our dataset has only two class labels which is either Highly Negative or Highly Positive which means that the binary classification approach is the best model that fits our data. Mainly, over the two class labels our network is designed by the first convolutional layer to maximize the multinomial logistic regression objective. That means it maximizes the log-probability of the average correct label across the training cases under the prediction distribution. Moreover, in another case, the second, fourth, and the fifth convolutional layers are designed as a ResNet attention

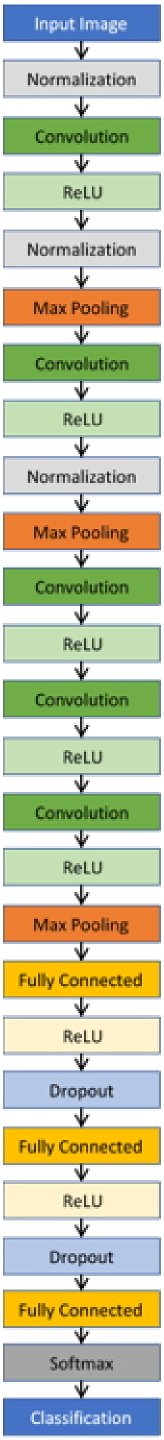


Fig. 5.4 : Deep learning structure for image sentiment analysis and classification. Each block illustrates specific function for the structure and the network configuration.

approach which are connected only to the relevant. However, the third convolutional layer is designed as a regular ResNet which is connected directly to all neurons. The

(a) Highly Negative Sentiment Image Samples.

(b) Highly Positive Sentiment Image Samples.

Fig. 5.5 : Samples of the sentiment images dataset.

reason behind that is to maximize the distribution of the multinomial logistic regression. The network structure like kernel sizes and map demotions are illustrated in Table 5.1.

5.4 Sentiment Analysis Dataset

The data set contains over 15,000 sentiment-scored images with typical positive and/or negative sentiment. In this case, the data set contains multiple URLs of images. Moreover, the sentiment scores may have one case of positive, neutral, or negative [143]. Some samples of the training and testing dataset are shown in Figure 5.5. The whole dataset consists of 4000 images that have been divided into 2000 images as Highly Negative and 2000 images as a Highly Positive. The dataset is split to 70% of images per category intended to train (1399 images for training) and 30% specified as a validation set intended to test (601 images for testing). Our network after it has been trained by specific training options is shown in Table 2. A small value is set as an initial learning rate which is down the training rate. In addition, the validation data and a small validation frequency are specified. Instead, we boost the learning rates of the new layers that we added, so that they change faster than the rest of the network. Thus, earlier layers do not change much, and we quickly learn the weights of the newer layer [OFB19].

5.5 Experimental Results

5.5.1 Training experimental results

Figure 5.6 shows the training accuracy and the lost function score. It is clear that the loss score starts from a higher score and decreases until reaches the lowest loss score by achieving a 10% loss score. However, the accuracy starts from the lower score of 30% and continues to increase until reaches the highest accuracy, which is almost 90% on the training dataset, after undergoing 150 iterations to do so.

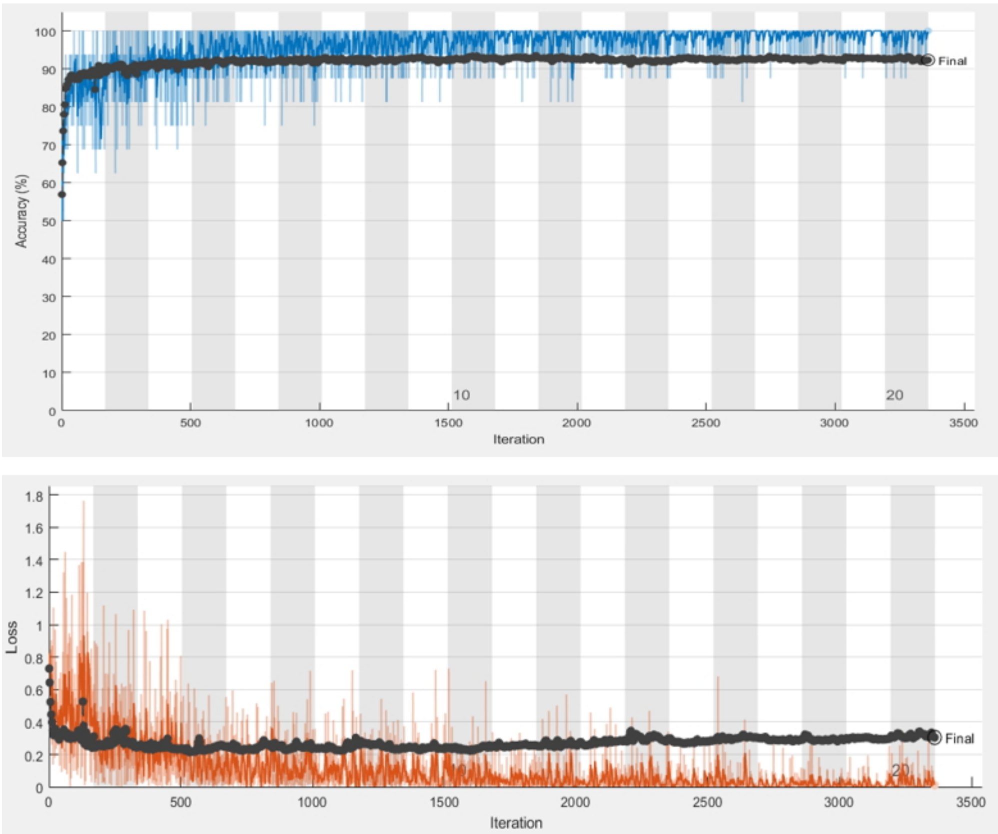


Fig. 5.6 : The overall performance of the training accuracy (blue) and the loss function score (red) during the training phase. in both plots, the dashed lines show the average training a score and the average loss function for each epoch.

Table 5.2 and Figure 5.7 illustrate the mini-batch accuracy based on each iteration during the training step. Table 5.2 indicates that the mini-batch accuracy achieves the highest accuracy of 100% and that the loss score is very low, reaching 0.0093.

The different parameter have been tuned during the training phase; the most stable and appropriate parameters are presented in Table 5.3.

Table 5.2 : Performance results of the training accuracy and the loss function score for each iteration.

Epoch	Iteration	Time Elapsed	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1-150	0:00:17	75.00%	0.583	1.00E-04
2	200-300	0:00:33	75.00%	0.6243	1.00E-04
3	350-500	0:00:54	100.00%	0.0031	1.00E-04
4	550-650	0:01:10	100.00%	0.0033	1.00E-04
5	700-800	0:01:26	93.75%	0.1597	1.00E-04
6	850-1000	0:01:47	93.75%	0.4944	1.00E-04
7	1050-1150	0:02:04	100.00%	0.1246	1.00E-04
8	1200-1300	0:02:20	93.75%	0.0479	1.00E-04
9	1350-1500	0:02:43	100.00%	0.08	1.00E-04
10	4501550-1650	0:02:59	100.00%	0.0154	1.00E-04
11	5001700-1800	0:03:15	93.75%	0.1471	1.00E-04
12	1850-2000	0:03:39	100.00%	0.0073	1.00E-04
13	2050-2150	0:03:56	100.00%	0.0075	1.00E-04
14	2200-2350	0:04:19	100.00%	0.022	1.00E-04
15	2400-2500	0:04:38	100.00%	0.001	1.00E-04
16	2550-2650	0:04:56	100.00%	0.0448	1.00E-04
17	2750-2768	0:05:09	100.00%	8.34E-05	1.00E-04

Table 5.3 : Training function parameters.

Function	Parameter
Training Function	Sigmoid Function
Mini Batch Size	10
Max Epochs	6
Shuffle	every-epoch
Initial Learn Rate	1.00E-04
Validation Data	Used
Validation Frequency	3
Max Epochs Number	20
Max Iterations Number	3360
Iteration per Epoch	168

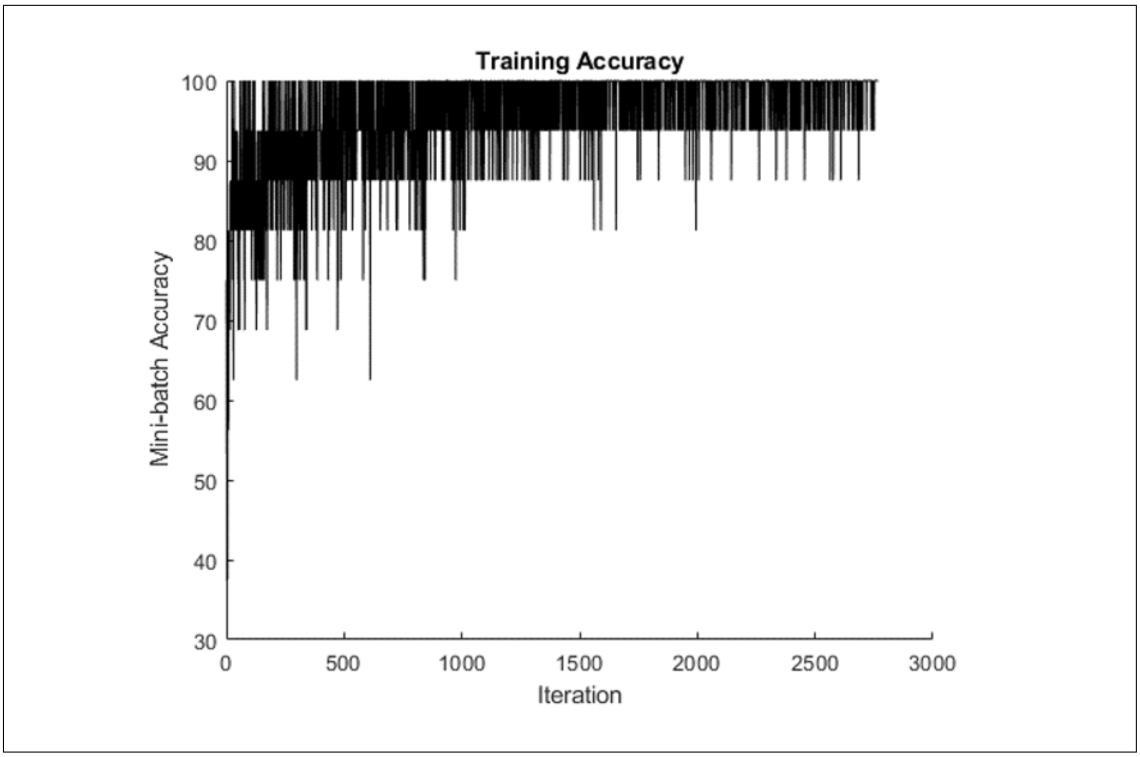


Fig. 5.7 : Mini-batch training accuracy for each iteration during the training phase.

5.5.2 Evaluation criteria

In this section, the performance results of our proposed system are evaluated. For this reason, various criteria are applied and used here for more significant indications. In this case, the most standard measurement approach for the evaluation of the results is measured by different criteria such as classification accuracy, detection rate as well as false-positive rate. To calculate them, different parameters need to be extracted such as TP which is the True Positive that gives the amount of correct detection of the true positive cases in the testing dataset. Also, the TN which is the True Negative that also gives the amount of correct detection of negative cases on the testing dataset. Another parameter is FP which is the False Positive which gives the incorrect classification of the positive cases among the negative cases in the testing dataset. Finally, the last parameter is the FN which is the False Negative which gives the incorrect classification of the negative cases among the positive cases in the testing dataset.

5.5.3 Testing performance and experimental results

The evaluation of the performance of DAM system is calculated using three measures called recognition rate (RR), precision (PR), sensitivity (SE), specificity (SP) [LG15, AAOTC19]. The formulas for calculating these measures are given in Equations (4.12), (4.13), (4.14), and (4.14) respectively. Figure 5.8 illustrates the confusion matrix of

Table 5.4 : Experimental results for the testing dataset.

Measure	Value
Sensitivity	0.9344
Specificity	0.9125
Precision	0.9102
Negative Predictive Value	0.9361
False Positive Rate	0.0875
False Discovery Rate	0.0898
False Negative Rate	0.0656
Accuracy	0.9231
F1 Score	0.9221

the testing dataset. and indicates that the proposed approach (deep attention network) achieves an RR of 92.31% In contrast, the most recent method for sentiment image classification using a regular DL approach achieved 78.1% on the same dataset.

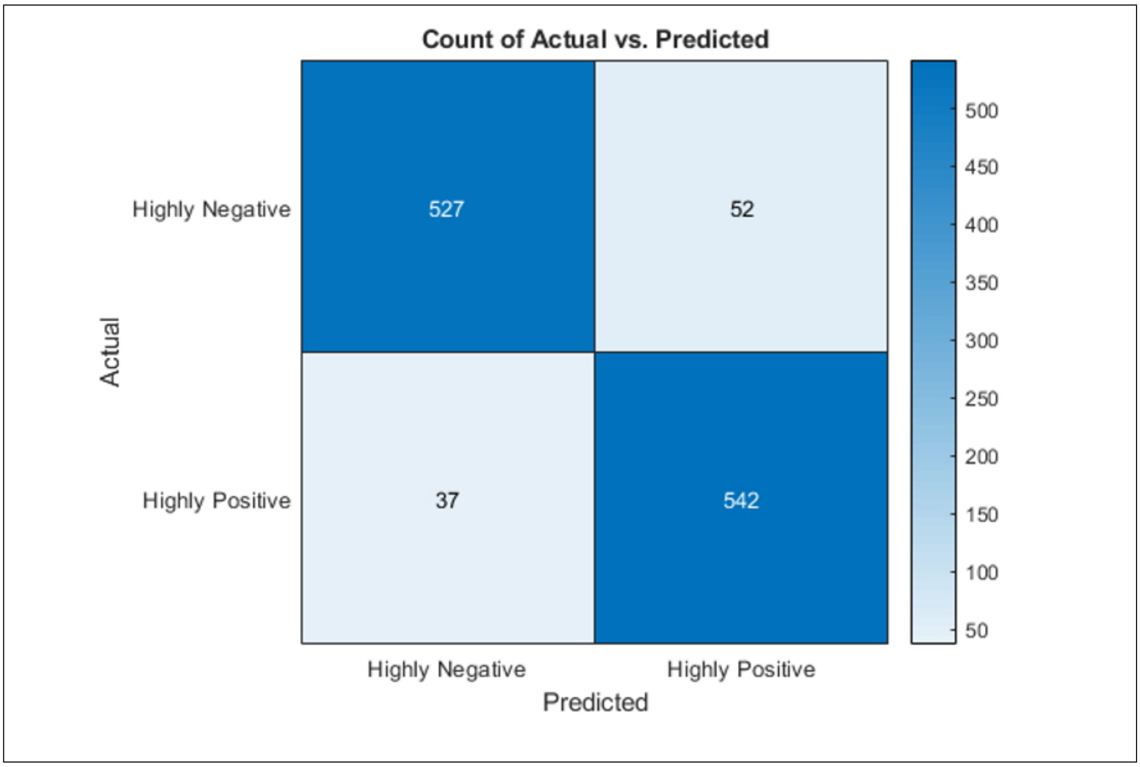


Fig. 5.8 : Confusion matrix of the testing experimental result.

The full performance results of the testing are shown in Table 5.4. the proposed system achieves 93.44% SE, 91.25% SP, 91.02% PR, 93.61% for the negative prediction, 0.875% for the false-positive rate, 0.898% for the false-discovery rate, 0.656% for the false-negative rate, and finally 92.21% as an F1 score.

Figure 5.9 presents some examples that have been randomly selected from the testing dataset. In these cases, the proposed approach is able to attain a confident prediction

Table 5.5 : Overall performance of deep attention model results comparing with different data mining and machine learning Approaches.

Methods	Framework	Approach	Accuracy
[SMDH10]	Low-level Feature-based	global color histograms (GCH)	66.00%
	Low-level Feature-based	local color histogram features (LCH)	66.40%
	Low-level Feature-based	global color histograms (GCH)+bag ofvisual word features (BoW)	66.50%
	Low-level Feature-based	local color histogram features (LCH)+bag of visual word features (BoW)	66.40%
[BJC+13]	Mid-level Feature-based	SentiBank	66.20%
[YMYL13]	Mid-level Feature-based	Sentribute	69.60%
[YLJY15]	CNN	Reqular CNN	66.70%
[YLJY15]	PCNN	Progressive CNN	68.70%
[JWLH18]	CNN	FC7	49%
[JS15b]	CNN	Domain Specific Fine Tuning	53.50%
Our Approach [AAAL20]	CNN	Attention Mechanism	92.31%

score in assigning the images to the correct label. The figure also shows that some cases have less confident scores that the other based on the color variation and the complexity of the tested images.

Comparing the performance results of our DL approach-based attention learning mechanism with other approaches, Table 5.5 indicates the performance results for the sentiment image analysis dataset using different approaches, such as low-level feature-based global color histograms (GCHs) (which archive 66% accuracy), low-level feature-based local color histogram features (LCHs) (66.4% accuracy), bag of visual-word features (BOWs) 66.4% accuracy), mid-level feature-based sentibank (66,2% accuracy), mid-level feature-based sentribute (69.6% accuracy), regular CNN model (66.7% accuracy), progressive CNN (68.7% accuracy) Low-level Feature-based LCH which achieves 66.4% accuracy, Low-level Feature-based GCH and bag of visual word features (BoW) which achieves 66.5%, Low-level Feature-based LCH and BoW which achieves 66.4%, Mid-level Feature-based SentiBank which achieves 66.2 % accuracy, Mid-level Feature-based Sentribute which achieves 69.6 % accuracy, regular CNN model which achieves 66.7% accuracy, Progressive CNN which achieves 68.7% accuracy, CNN model based FC7 achieves 49% accuracy, and CNN Model-based Domain-Specific Fine-Tuning achieves 53% accuracy. It notices that our deep attention model is more powerful than the other approaches by achieving 92.31% accuracy.

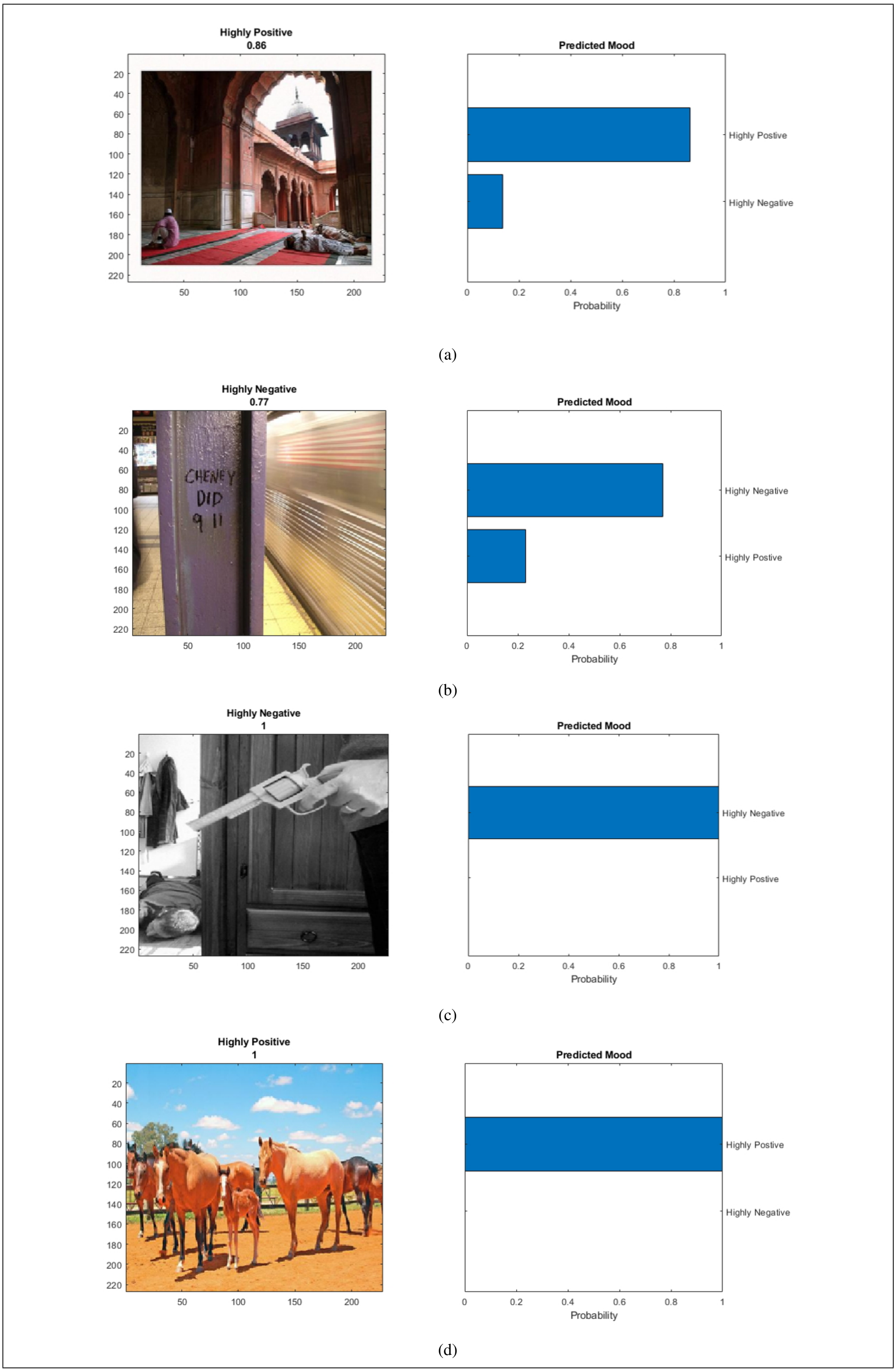


Fig. 5.9 : The performance results in the testing experimental result.

5.6 Summary

Sentiment analysis tackles the challenging problem of interpreting the high-level content of large-scale visual data based on algorithms devised from computer vision. In this work, we design a DANM to achieve a high level of social media sentiment image analysis and to classify the images as Highly Positive mood or Highly Negative mood. The DANM produces features maps through utilizing the adequate focusing technique of ML based on a proper CNN. The proposed network presents significantly higher accuracy and efficiency in the performance results by achieving 92.31% compared with the highest accuracy of 68.7% that was achieved by using the progressive CNN with the same standard image sentiment dataset.

Chapter 6

Conclusion and Future Works

The main aim of this thesis was to evaluate the impact of data mining and classification on sentiment analysis and emotion detection. The experimental presented in Chapter 4 in the first system proved that, based on DR methods, sentiment analysis and opinion mining of unstructured social data can generally be improved. In the same chapter, the evaluation results in the second system revealed that the use of a deep neural network to detect the opinion increases the accuracy of the classification. Moreover, the evaluation results described in Chapter 5 in the third system proved that using a CNN-based attention mechanism can generally improve emotion detection in social images. This thesis also introduced several classification approaches that are based on how to classify the emotion in unstructured data on social media. The main objective of these approaches was to increase the accuracy of detecting and extracting useful information from huge unstructured datasets available online to support decision-making processes. This chapter emphasizes how the research questions have been addressed, highlights the contributions Section 6.1, summarizes the methods and the findings of this research Section 6.2, and sets out directions for future work Section 6.3. The final section summarizes the contribution and the research questions and illustrates the questions that have been addressed.

6.1 Addressing the Research Questions

This research adopted a methodology that included building a system capable of extracting important information from large text and image datasets, in order to address the research questions.
This methodology emphasized a three-stage process: (1) an analysis stage, which involved the investigation of the literature to formulate the hypothesis and determine the main influential entities that needed to be modeled; (2) system development; and (3) a design stage which involved describing the experiment's set-up to answer the research questions. The methodology served to answer the research questions, including the following.

1. The evaluation of dimension reduction as a technique to enhance the accuracy and speed of detection and classification processes in the textual social data. The evaluation involved proposing a new DR method to reduce the dimensions of text before the classification stage (Chapter 4).

2. The evaluation of classification, which involved using a deep neural network to increase the accuracy of results in the detection process (Chapter 4).

3. The provision of systems that can extract emotion from images using CNN-based attention mechanisms to increase the focus on important features in the classified image (Chapter 5).

The research questions and how they were answered are outlined below

1. How can information about sentiment be extracted from texts that contain huge numbers of words to help in complex decision making?

   Unstructured data in the form of text can be found everywhere: Emails, reviews, comments, social media, support tickets, surveys, and so on. Generally, text is a type of data that contains a lot of information; but, of course, extracting useful information from this type of data is a difficult process due to the high volume of data in terms of high dimensions. Therefore, it needs a lot of time to process. Businesses and companies are restructuring big data using data-mining algorithms and classification methods to improve decision making. Therefore, to answer this research question, this thesis introduced a novel dimension-reduction method in order to solve the problem of high dimensions in text data and to then classify the text into a category based on ANN algorithms. The experimental methodology and results are presented in Chapter 4.

2. How can the process of sentiment detection described in answering the first research question be improved by increasing performance and accuracy while maintaining the premise of huge amounts of data?

   To answer this question, a deep sentiment system based on CNNs and LSTM is proposed. The model uses one layer of CNNs to capture the partial features of the text after which the features are fed to the LSTM, which captures the contextual information. The proposed sentiment-detection model and the performance evaluation is presented in Chapter 4.

3. How can emotion be inferred from a given image?

   In a search for tag "Victory" on Flickr, an enormous variety of images is found: For example, the sun, a sign of victory, the Statue of Liberty, soldiers dancing and pictures of a runner having crossed the finish line. All the images obtained bear the connotation of victory, but each one differs from the others due to the different perspectives and also to the different nature of the individual expressing victory. In this thesis, we used a DL-based attention mechanism for increased focus on features in order to classify and detect the emotion that images carry. The results that we obtained were very good, and they proved that deep neural networks are able to learn and extract the emotions inherent in images. These prediction systems can be used in many areas; for example, to determine out people's opinions on an important topic, such as an election.

4. Is it possible to provide automated systems that can understand the sentiment and emotions of data available on those social networking platforms?

   The evaluation of the detection of emotion from unstructured data and the proposed classification algorithms using different data types was carried out in this thesis using the designed systems. Therefore, this research question was answered through building a sentiment analysis model. As the building of any model of the real world should demonstrate that the model reflects the reality, we parameterized the presented models with data for real individuals' opinions and sentiments from social media. To ensure the model's validity before any experiments were designed, we validate the presented models by comparing them against those in other sentiment-analysis and emotion-detection studies using same datasets. The models and the gathered measurements are illustrated in Chapters 4 and 5.

6.2 Summary

Nowadays, social networking sites occupy a large part of individuals' lives and are also the focus and attention of researchers and scientists. Sentiment analysis is a challenge and attempts to interpret the high-level content of large-scale visual data based on algorithms devised from computer vision. In this thesis, different data-mining and ML approaches are proposed for sentiment analysis (opinion-mining prediction and classification) using various social media data types, such as tweets, review text, and images. We proposed an intelligent opinion-mining approach that can visualize the Twitter opinion and also predict and classify the Twitter mood based on the data-mining framework. As this paper's focus was on the tweet's text processing for Twitter's opinion-mining prediction and classification, dimension reduction-based data-mining approaches were proposed to reduce the data's dimensionality in addition to the supervised learning classification approach, such as a BPNN. The main contribution of this work is to propose a different dimensionality-reduction and feature-selection approach that helps the BPNN to predict and classify with higher accuracy than the other approaches, such as the PCA and the

SVD, even with the whole feature space. Our approach achieved 94.8% accuracy in the testing dataset, while the nearest comparator was the 90.34% achieved by the PCA. In addition, a DL framework for a text opinion-mining prediction and classification approach was proposed in this thesis. This is a new visualization model for the prediction of Twitter mood based on DL approach. The proposed system visualizes the top three predictions and their scores for each tweet in the testing data after the DL model is trained using the training dataset that includes randomly selected items from the original dataset. Then the proposed system sorts the prediction scores and selects the top three values, visualizing them as a main predicted Twitter mood (opinion). Our approach attained 96.1% in the testing dataset, while the nearest comparator was the 94.8% achieved by the our dimension rediction method [AGAAL19].

Finally, a DANM was proposed as a third ML approach to achieve an increased level of social media sentiment image analysis and to classify the data as Highly Positive mood or Highly Negative mood. The DANM produces features maps utilizing the focusing adequate technique of ML-based a proper CNN. The proposed network presented high accuracy and efficiency in the performance results by achieving 92.31% higher than the most recent work that has been tested using the same dataset, which achieved 78.1%.

6.3 Future Work

For future research, we will continue working on mixtures of images, video, and text data on social media. We will determine user opinion, the sentiment of the text, and the emotion from images by proposing a topic aspects-based generative mixture model for a movie recommendation system using a deep convolutional network. We aim to build a system for recommending a movie based on analyzing previous users' emotions and by extracting sentiment and topics from movie descriptions. Movie-recommendation systems have become ubiquitous in most aspects of our lives. Currently, they are far from optimal. In the future, we will present a movie-recommendation system based on ML by utilizing a deep convolutional network and generative modeling of public previous aspects mixtures. The objective of the future work is to introduce such a recommendation system to help users in selecting datasets of movies according to certain pre-specified measurements and data. The applied methodology on implementing the system using different sentimental analysis algorithms. These algorithms provide a solution for the full-stack developers by employing a trained model using their datasets. This will help users in visualizing their interest or informing a better scope of visualization. We expect this system to be convenient for business process-modeling applications, various data-mining applications, and e-commerce websites, in addition to most online platforms that people use including social media.

References

[AAAL19] Maha Alghalibi, Adil Al-Azzawi, and Kai Lawonn. Deep Tweets Analyzer Model for Twitter Mood Visualization and Prediction Based Deep Learning Approach. *International Journal of Computer and Communication Engineering*, 8:1–17, 2019.

[AAAL20] Maha Alghalibi, Adil Al-Azzawi, and Kai Lawonn. Deep attention learning mechanisms for social media sentiment image revelation. *International Journal of Computer and Communication Engineering*, pages 1–17, 2020.

[AAASCH18] Adil Al-Azzawi, Hasanain Al-Sadr, Jianlin Cheng, and Tony X Han. Localized Deep Norm-CNN Structure for Face Verification. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 8–15. IEEE, 2018.

[AAOTC19] Adil Al-Azzawi, Anes Ouadou, John J Tanner, and Jianlin Cheng. AutoCryoPicker: an unsupervised learning approach for fully automated single particle picking in Cryo-EM images. *BMC bioinformatics*, 20(1):326, 2019.

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[Abn07] Steven Abney. *Semisupervised learning for computational linguistics*. Chapman and Hall/CRC, 2007.

[AGAA18] Maha Al-Ghalibi and Adil Al-Azzawi. Time series social network visualization based on dimension reduction. 2018.

[AGAAL19] Maha Al-Ghalibi, Adil Al-Azzawi, and Kai Lawonn. NLP based sentiment analysis for Twitter's opinion mining and visualization. In *Eleventh International Conference on Machine Vision (ICMV 2018)*, volume 11041, page 110412A. International Society for Optics and Photonics, 2019.

[Agg15] Charu C Aggarwal. *Data mining: the textbook*. Springer, 2015.

[Agg16] Charu C Aggarwal. An introduction to recommender systems. In *Recommender systems*, pages 1–28. Springer, 2016.

[AHMJP12] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*, pages 4277–4280. IEEE, 2012.

[Alp09] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2009.

[AWO10] Charu C Aggarwal, Haixun Wang, and Others. *Managing and mining graph data*, volume 40. Springer, 2010.

[AZ12a] Charu C Aggarwal and ChengXiang Zhai. A survey of text clustering algorithms. In *Mining text data*, pages 77–128. Springer, 2012.

[AZ12b] Charu C Aggarwal and ChengXiang Zhai. *Mining text data*. Springer Science & Business Media, 2012.

[BH11] J C A Barata and M S Hussein. The Moore-Penrose Pseudoinverse. *A tutorial review of the theory*, 2011.

[BIG03] Adi Ben-Israel and Thomas N E Greville. *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media, 2003.

[Bis06] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[BJC$^+$13] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM, 2013.

[BK18] Ashray Bhandare and Devinder Kaur. Designing Convolutional Neural Network Architecture Using Genetic Algorithms. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, pages 150–156. The Steering Committee of The World Congress in Computer Science, Computer˜. . . , 2018.

[BL12] M A Beyer and D Laney. The importance of'big data': A definition. Stamford, CT: Gartner Inc. Report ID number: G00235055, 2012.

[BM01] Ella Bingham and Heikki Mannila. *Random projection in dimensionality reduction: applications to image and text data*. 2001.

[BMZ11] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

[BWZ16] Christos Boutsidis, David P Woodruff, and Peilin Zhong. Optimal principal component analysis in distributed and streaming models. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 236–249. ACM, 2016.

[CC19] Ching-Hsue Cheng and Hsien-Hsiu Chen. Sentimental text mining based on an additional features method for text classification. *PloS one*, 14(6), 2019.

[CCH$^+$14] Yan-Ying Chen, Tao Chen, Winston H Hsu, Hong-Yuan Mark Liao, and Shih-Fu Chang. Predicting viewer affective comments based on image content in social media. In *proceedings of international conference on multimedia retrieval*, page 233. ACM, 2014.

[CEHM14] Toon Calders, Floriana Esposito, Eyke Hüllermeier, and Rosa Meo. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings*, volume 8724. Springer, 2014.

[CGGS12] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851, 2012.

[Cho14] Anna Ewa Choromanska. *Selected machine learning reductions*. PhD thesis, Columbia University, 2014.

[CKNP17] Lingjiao Chen, Arun Kumar, Jeffrey Naughton, and Jignesh M Patel. Towards linear algebra over normalized data. *Proceedings of the VLDB Endowment*, 10(11):1214–1225, 2017.

[CLZ17] Jinyoung Choi, Beom-Jin Lee, and Byoung-Tak Zhang. Multi-focus attention network for efficient deep reinforcement learning. In *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[CMH+18] Bo Chang, Lili Meng, Eldad Haber, Lars Ruthotto, David Begert, and Elliot Holtham. Reversible architectures for arbitrarily deep residual neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[CPS98] Krzysztof J Cios, Witold Pedrycz, and Roman W Swiniarski. Data mining and knowledge discovery. In *Data mining methods for knowledge discovery*, pages 1–26. Springer, 1998.

[CS07] Daniel S Coming and Oliver G Staadt. Velocity-aligned discrete oriented polytopes for dynamic collision detection. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):1–12, 2007.

[DAYD17] Jose Dolz, Ismail Ben Ayed, Jing Yuan, and Christian Desrosiers. HyperDense-Net: A hyper-densely connected CNN for multi-modal image semantic segmentation. *arXiv preprint arXiv:1710.05956*, 2017.

[DD08] Werner Dubitzky and Werner Dubitzky. *Data mining techniques in grid computing environments*. Wiley Online Library, 2008.

[DL18] Guozhu Dong and Huan Liu. *Feature engineering for machine learning and data analytics*. CRC Press, 2018.

[DTR10] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters*, pages 241–249. Association for Computational Linguistics, 2010.

[EL13] L Einav and J D Levin. The data revolution and economic analysis (Working Paper No. 19035). *Retrieved from National Bureau of Economic Research website: http://www. nber. org/papers/w19035*, 2013.

[FHL14] Jianqing Fan, Fang Han, and Han Liu. Challenges of big data analysis. *National science review*, 1(2):293–314, 2014.

[Fre03] Ivar Fredholm. Sur une classe d'équations fonctionnelles. *Acta mathematica*, 27(1):365–390, 1903.

[FZ09] Francesca Fallucchi and Fabio Massimo Zanzotto. SVD feature selection for probabilistic taxonomy learning. In *Proceedings of the workshop on geometrical models of natural language semantics*, pages 66–73, 2009.

[GBH09] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009, 2009.

[GC17] Lucía Martín Gómez and María Navarro Cáceres. Applying data mining for sentiment analysis in music. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 198–205. Springer, 2017.

[Gei19] Stefan Geißler. Maschinelles Lernen und NLP: Reif für die industrielle Anwendung! *Information-Wissenschaft & Praxis*, 70(2-3):134–140, 2019.

[GG15] Vasavi Gajarla and Aditi Gupta. Emotion detection and sentiment analysis of images. *Georgia Institute of Technology*, 2015.

[GH15] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2):137–144, 2015.

[Gia08] Eugenia Giannopoulou. *Data mining in medical and biological research*. BoD–Books on Demand, 2008.

[GK65] Gene Golub and William Kahan. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, 2(2):205–224, 1965.

[GPW10] Mrudula Gudadhe, Prakash Prasad, and Lecturer Kapil Wankhade. A new data mining based network intrusion detection model. In *2010 International Conference on Computer and Communication Technology (ICCCT)*, pages 731–735. IEEE, 2010.

[Gra76] Franklin A Graybill. *Theory and application of the linear model*. Number QA279. G72 1976. Duxbury press North Scituate, MA, 1976.

[GRUG17] Aidan N Gomez, Mengye Ren, Raquel Urtasun, and Roger B Grosse. The reversible residual network: Backpropagation without storing activations. In *Advances in neural information processing systems*, pages 2214–2224, 2017.

[HB14] Roger A Horn and Nelson H F Beebe. A Bibliography of Matrix Analysis, Second. 2014.

[HDE15] Tanri Hedresta, Eko Darwiyanto, and Veronikha Effendy. Analisis dan Implementasi Web Usage Mining Menggunakan Metode Self Organizing Map dan K-Means (Studi Kasus: Aktifitas Internet Telkom University). In *Indonesia Symposium on Computing 2015*, 2015.

[HK06] Jiawei Hang and Micheline Kamber. Data mining: concepts and techniques, 2006.

[Hou58] Alston S Householder. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM (JACM)*, 5(4):339–342, 1958.

[Hui04] Liu Huiqing. *Effective use of data mining technologies on biological and clinical data*. PhD thesis, Citeseer, 2004.

[IK19] Runar Isaksen and Aline Iyagizeneza Knudsen Eirik Bøand Walde. A deep learning segmentation approach to calories and weight estimation of food images. Master's thesis, Universitetet i Agder; University of Agder, 2019.

[ILW16] M Iftene, Q Liu, and Y Wang. Very high resolution images classification by fine tuning deep convolutional neural networks. In *Eighth International Conference on Digital Image Processing (ICDIP 2016)*, volume 10033, page 100332D. International Society for Optics and Photonics, 2016.

[IZ16] Jyoti Islam and Yanqing Zhang. Visual sentiment analysis for social images using transfer learning approach. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 124–130. IEEE, 2016.

[Jef13] Adrianne Jeffries. The man behind flickr on making the serviceawesome again. *The Verge*, 20, 2013.

[JS15a] Stuti Jindal and Sanjay Singh. Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning. In *2015 International Conference on Information Processing (ICIP)*, pages 447–451. IEEE, 2015.

[JS15b] Stuti Jindal and Sanjay Singh. Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning. pages 447–451, 2015.

[JWLH18] Hai Jin, Xun Wang, Yuanfeng Lian, and Jing Hua. Emotion information visualization through learning of 3D morphable face model. *The Visual Computer*, feb 2018.

[JWLH19] Hai Jin, Xun Wang, Yuanfeng Lian, and Jing Hua. Emotion information visualization through learning of 3D morphable face model. *The Visual Computer*, 35(4):535–548, 2019.

[JYZ+11] Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 151–160. Association for Computational Linguistics, 2011.

[JZSC09] Bernard J Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188, 2009.

[KA15] Olga Kanishcheva and Galia Angelova. About Emotion Identification in Visual Sentiment Analysis. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 258–265, 2015.

[Kan11] Mehmed Kantardzic. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.

[Kau92] C J Kaufman. Rocky Mountain Research Laboratories. *Boulder, CO, personal communication*, 1992.

[KDHR17] Yoon Kim, Carl Denton, Luong Hoang, and Alexander M Rush. Structured attention networks. *arXiv preprint arXiv:1702.00887*, 2017.

[KFD+07] Daphne Koller, Nir Friedman, Sašo Džeroski, Charles Sutton, Andrew McCallum, Avi Pfeffer, Pieter Abbeel, Ming-Fai Wong, David Heckerman, Chris Meek, and Others. *Introduction to statistical relational learning*. MIT press, 2007.

[KGR+20] Sudhanshu Kumar, Monika Gahalawat, Partha Pratim Roy, Debi Prosad Dogra, and Byung-Gyu Kim. Exploring impact of age and gender on sentiment analysis using machine learning. *Electronics*, 9(2):374, 2020.

[Koh82] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.

[Kou16] Jayanth Koushik. Understanding convolutional neural networks. *arXiv preprint arXiv:1605.09081*, 2016.

[KQM13] Daniel Keim, Huamin Qu, and Kwan-Liu Ma. Big-data visualization. *IEEE Computer Graphics and Applications*, 33(4):20–21, 2013.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[LG15] Jianwei Liu and Lei Guo. A new brain MRI image segmentation strategy based on wavelet transform and K-means clustering. In *2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–4. IEEE, 2015.

[LJL16] Wootaek Lim, Daeyoung Jang, and Taejin Lee. Speech emotion recognition using convolutional and recurrent neural networks. In *2016 Asia-Pacific signal and information processing association annual summit and conference (APSIPA)*, pages 1–4. IEEE, 2016.

[Lju17] Benjamin Fayyazuddin Ljungberg. Dimensionality reduction for bag-of-words models: Pca vs lsa. *Semanticscholar. org*, 2017.

[LL14] Daniel T Larose and Chantal D Larose. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.

[Loh12] Steve Lohr. The age of big data. *New York Times*, 11(2012), 2012.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[LZ12] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis. In *Mining text data*, pages 415–463. Springer, 2012.

[LZN20] Haoye Lu, Haolong Zhang, and Amit Nayak. A deep neural network for audio classification with a classifier attention mechanism. *arXiv preprint arXiv:2006.09815*, 2020.

[MA05] Sushmita Mitra and Tinku Acharya. *Data mining: multimedia, soft computing, and bioinformatics*. John Wiley & Sons, 2005.

[MD15] P Mohata and Sheetal Dhande. Web Data Mining Techniques and Implementation for Handling Big Data. *Computer Science and Mobile Computing*, 4(4):330–334, 2015.

[ME20] Avinash Madasu and Sivasankar Elango. Efficient feature selection techniques for sentiment analysis. *Multimedia Tools and Applications*, 79(9):6313–6335, 2020.

[MH13] Prakash Mahindrakar and M Hanumanthappa. Data mining in healthcare: A survey of techniques and algorithms with its limitations and challenges. *International Journal of Engineering Research and Applications*, 3(6):937–941, 2013.

[MK20] D T Mane and Uday V Kulkarni. A survey on supervised convolutional neural network and its major applications. In *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications*, pages 1058–1071. IGI Global, 2020.

[MKA11] Kornel Mucsi, Ata M Khan, and Mojtaba Ahmadi. An Adaptive Neuro-Fuzzy Inference System for estimating the number of vehicles for queue management at signalized intersections. *Transportation Research Part C: Emerging Technologies*, 19(6):1033–1047, 2011.

[Moh20] AA Mohamed. An effective dimension reduction algorithm for clustering arabic text. *Egyptian Informatics Journal*, 21(1):1–5, 2020.

[Moo20] Eliakim H Moore. On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, 26:394–395, 1920.

[Mor60] David D Morrison. Remarks on the unitary triangularization of a nonsymmetric matrix. *Journal of the ACM (JACM)*, 7(2):185–186, 1960.

[MSC13] Viktor Mayer-Schönberger and Kenneth Cukier. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.

[MSJ18] Dhwani Mehta, Mohammad Faridul Haque Siddiqui, and Ahmad Y Javaid. Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors*, 18(2):416, 2018.

[MSK17] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):1–23, 2017.

[OBRS10] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

[OFB19] Alessandro Ortis, Giovanni Maria Farinella, and Sebastiano Battiato. An overview on image sentiment analysis: Methods datasets and current challenges. In *Proc. 16th Int. Joint Conf. E-Bus. Telecommun.*, volume 1, pages 290–300, 2019.

[PLO08] Bo Pang, Lillian Lee, and Others. Opinion mining and sentiment analysis. *Foundations and Trends®in Information Retrieval*, 2(1–2):1–135, 2008.

[POG99] J Paul Peter, Jerry C Olson, and Klaus G Grunert. *Consumer behaviour and marketing strategy*. McGraw-hill London, 1999.

[PPB15] Vishakha Patel, Gayatri Prabhu, and Kiran Bhowmick. A survey of opinion mining and sentiment analysis. *International Journal of Computer Applications*, 131(1):24–27, 2015.

[Rak97] Vladimir Rakočević. On continuity of the Moore-Penrose and Drazin inverses. *Matematichki Vesnik*, 49(3-4):163–172, 1997.

[Ren19] Thomas Renault. Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. *Digital Finance*, pages 1–13, 2019.

[RGC+97] Wilburn E Reddick, John O Glass, Edwin N Cook, T David Elkin, and Russell J Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. *IEEE Transactions on medical imaging*, 16(6):911–918, 1997.

[RK18] Ankita Rane and Anand Kumar. Sentiment classification system of Twitter data for US airline service analysis. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, volume 1, pages 769–773. IEEE, 2018.

[RPG06] Sanjay Rawat, Arun K Pujari, and Ved Prakash Gulati. On the use of singular value decomposition for a fast intrusion detection system. *Electronic Notes in Theoretical Computer Science*, 142:215–228, 2006.

[SD12] Anuj Sharma and Shubhamoy Dey. A comparative study of feature selection and machine learning techniques for sentiment analysis. In *Proceedings of the 2012 ACM research in applied computation symposium*, pages 1–7, 2012.

[SD13a] Anuj Sharma and Shubhamoy Dey. A boosted svm based sentiment analysis approach for online opinionated text. In *Proceedings of the 2013 research in adaptive and convergent systems*, pages 28–34. 2013.

[SD13b] Ms Snehal Suryawanshi and V Dixit. Improved technique for detection of malaria parasites within the blood cell images. *Int J Sci Eng Res*, 4:373–375, 2013.

[SEZS00] Matthew G Schultz, Eleazar Eskin, F Zadok, and Salvatore J Stolfo. Data mining methods for detection of new malicious executables. In *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001*, pages 38–49. IEEE, 2000.

[SGM+19] Martin Suero, Carsten Paul Gassen, Dimitrije Mitic, Ning Xiong, and Miguel Leon. A deep neural network model for music genre recognition. In *The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pages 377–384. Springer, 2019.

[SHK+14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[Sim15] George Simos. How much data is generated every minute on social media. *We Are Social Media*, 19, 2015.

[SLX+15] Hai Su, Fujun Liu, Yuanpu Xie, Fuyong Xing, Sreenivasan Meyyappan, and Lin Yang. Region segmentation in histopathological breast cancer images using deep convolutional neural network. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 55–58. IEEE, 2015.

[SMDH10] Stefan Siersdorfer, Enrico Minack, Fan Deng, and Jonathon Hare. Analyzing and Predicting Sentiment of Images on the Social Web. *Proceedings of the 18th ACM International Conference on Multimedia*, pages 715–718, 2010.

[SMGS14] Marijn F Stollenga, Jonathan Masci, Faustino Gomez, and Jürgen Schmidhuber. Deep networks with internal selective attention through feedback connections. In *Advances in neural information processing systems*, pages 3545–3553, 2014.

[SMKR13] Tara N Sainath, Abdel-rahman Mohamed, Brian Kingsbury, and Bhuvana Ramabhadran. Deep convolutional neural networks for LVCSR. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 8614–8618. IEEE, 2013.

[SN19] Nishit Shrestha and Fatma Nasoz. Deep learning sentiment analysis for amazon reviews and ratings. *arXiv preprint arXiv:1904.04096*, 2019.

[SO03] Jiri Sima and Pekka Orponen. General-purpose computation with neural networks: A survey of complexity theoretic results. *Neural Computation*, 15(12):2727–2778, 2003.

[SS15a] Yustinus Eko Soelistio and Martinus Raditia Sigit Surendra. Simple text mining for sentiment analysis of political figure using naive bayes classifier method. *arXiv preprint arXiv:1508.05163*, 2015.

[SS15b] Mariusz Szwoch and Wioleta Szwoch. Emotion recognition for affect aware video games. In *Image Processing & Communications Challenges 6*, pages 227–236. Springer, 2015.

[SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[Ste12] Michael A Stelzner. *2012 social media marketing industry report: how marketers are using social media to grow their businesses*. social media examiner, 2012.

[SVM14] Carlos Oscar Sánchez Sorzano, Javier Vargas, and A Pascual Montano. A survey of dimensionality reduction techniques. *arXiv preprint arXiv:1403.2877*, 2014.

[Swa16] Japa Swadia. *A Study of Text Mining Framework for Automated Classification of Software Requirements in Enterprise Systems*. Arizona State University, 2016.

[TGL15] Nima Tajbakhsh, Michael B Gotway, and Jianming Liang. Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 62–69. Springer, 2015.

[TS12] Deepika Tewari and Sanjay Kumar Srivastava. A visual recognition of static hand gestures in indian sign language based on kohonen self-organizing map algorithm. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(2):165–170, 2012.

[TSN18] Gongbo Tang, Rico Sennrich, and Joakim Nivre. An analysis of attention mechanisms: The case of word sense disambiguation in neural machine translation. *arXiv preprint arXiv:1810.07595*, 2018.

[TWX] Wanliang Tan, Xinyu Wang, and Xu Xinyu. Sentiment analysis for amazon reviews.

[TWX19] Wanliang Tan, Xinyu Wang, and Xinyu Xu. Sentiment analysis for amazon reviews. *Reseaechgate*, (March):0–9, 2019.

[VC14] G Vinodhini and RM Chandrasekaran. Opinion mining using principal component analysis based ensemble model for e-commerce application. *CSI transactions on ICT*, 2(3):169–179, 2014.

[VL15] Andrea Vedaldi and Karel Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692. ACM, 2015.

[VPZ13] Herna L Viktor, Eric Paquet, and Jing Zhao. Artificial neural networks for predicting 3D protein shapes from amino acid sequences. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2013.

[VVW$^+$15] Mitko Veta, Paul J Van Diest, Stefan M Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders B L Larsen, Jacob S Vestergaard, Anders B Dahl, and Others. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical image analysis*, 20(1):237–248, 2015.

[Wat04] David S Watkins. *Fundamentals of matrix computations*, volume 64. John Wiley & Sons, 2004.

[YIAH14] Alireza Yousefpour, Roliana Ibrahim, and Haza Nuzly Abdull Hamed. A novel feature reduction method in sentiment analysis. *International Journal of Innovative Computing*, 4(1), 2014.

[YJ12] Quanzeng You and San Jose. Visual Sentiment Analysis by Attending on Local Image Regions. pages 231–237, 2012.

[YLJY15] Quanzeng You, Jiebo Luo, Hailin Jin, and Jianchao Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[YMYL13] Jianbo Yuan, Sean Mcdonough, Quanzeng You, and Jiebo Luo. Sentribute: image sentiment analysis from a mid-level perspective. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 10. ACM, 2013.

[ZB16a] Charles Y. Zheng and Yuval Benjamini. Estimating mutual information in high dimensions via classification error. (Nips), 2016.

[ZB16b] Charles Y Zheng and Yuval Benjamini. Estimating mutual information in high dimensions via classification error. *arXiv preprint arXiv:1606.05229*, 2016.

Internet-resources

[143] DATA.WORLD (). https://data.world/crowdflower/image-sentiment-polarity, Last accessed on May 2019.

[144] KAGGLE (). https://www.kaggle.com/crowdflower/twitter-airline-sentiment, Last accessed on May 2019.

[145] MCAuley, J. (). http://jmcauley.ucsd.edu/data/amazon/, Last accessed on July 2020.

Appendix A

Singular Value Decomposition (SVD) [KFD+07]

Algorithm A.1: Singular Value Decomposition (SVD)

Input: Generate Data Matrix X
Output: New Dimensions C

1. **Repeat**
2. Applying SVD to the matrix X as $X = USV^T$
   X $\rightarrow$ is an m$\times nmatrix(m\rightarrow$ no. of sessions (vectors)
   n $\rightarrow$ is no. of attributes)
   $U \leftarrow XX^T$ matrix of the eigenvectors
   S is matrix which is diagonal
   $V \leftarrow$ is matrix the eigenvectors.
3. **Construct** the covariance matrix from this decomposition by
   $XX^TXX^T \leftarrow (USV^T)(USV^T)^T = (USV^T)(VSU^T)$
4. V $\rightarrow$ an orthogonal matrix
   $(V^TV = I), XX^T = US2U^T$
5. The square roots of the eigenvalues of XX^T are the singular values of X
6. **until** Represent every transaction Ii over the time interval t as a vector $x(t)_i$
7. **Return** U^TX

Principle Component Analysis (PCA) [GPW10]

Algorithm A.2: Principle Component Analysis (PCA)

Input: Generate Data matrix X (features of KDD 99) Number of principle component d
Output: New Dimensions N

1. **Repeat**
2. Compute the mean of transactions$\mu \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$
3. Subtract the mean from each transactionX $(t) \leftarrow x_i - \mu$
4. Compute the covariance matrix $co(t) \leftarrow \frac{1}{m} X_n X_n^T$
5. From $Co(t)$ Compute eigenvectors u(t)of AA^T
6. Consider matrix AA^T as a matrix $M \times M$ matrix
7. Compute the eigenvectors v(t) of AA^T such that:
8. $AA^T V_i \rightarrow \mu_i V_i$
9. $\mu_i V_i \rightarrow AA^T AV_i$
10. Compute the best eigenvectors of AA^T : $\mu_i \leftarrow AV_i$
11. Keep only K eigenvectors, (K features with their values).
12. U $\leftarrow$ Top(eigenvector(C,d)
13. **until** Represent every transaction Ii over the time interval t as a vector $x(t)_i$
14. Return $N \leftarrow U^T$ X

Mutual Information (MI) [ZB16b]

Algorithm A.3: Mutual Information (MI)

Input: Feature Space X
Output: Mutual Scoring I

1. **Repeat**
2. **Compute** the MI for each two variables X and Y
3. **I(X,Y)** $\leftarrow_{x \in X \ y \in Y} P(x,y) log \frac{P(x,y)}{P(x)p(y)}$

a **X:** is the first feature space

b **Y:** is the second feature space

c **P:** Probability function

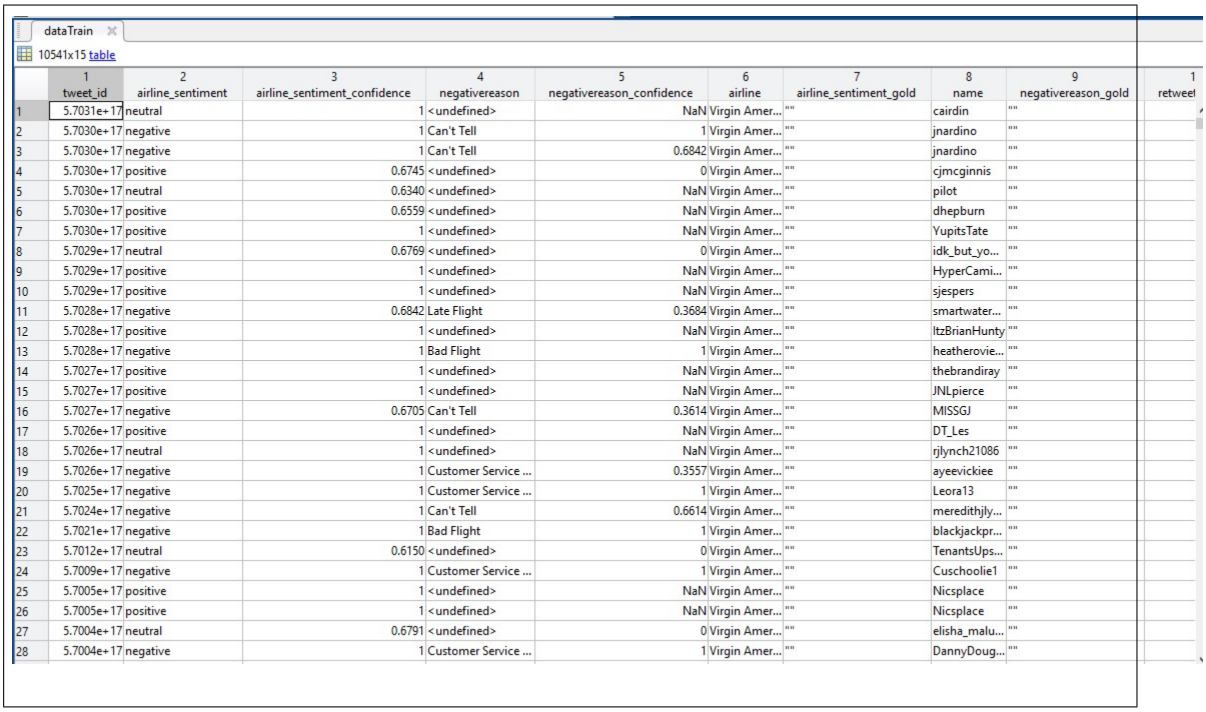
4. **Import** the next two variables in the feature space
5. **Until** consumed all variable in the whole variables
6. **Return** the mutual scoring vector. **End**

Appendix B

Step by Step of NLP based Twitter's Opinion Mining Prediction Using Dimensionality Reduction

- First step, visualize the Twitter's Opinion based on the distribution of the classes in the data using a histogram. Which is very common analyses way that can be performed on a large number of tweets. Sentiment analysis is scored based on the words contained in a tweet. Sentiment analysis provides a convenient way to take the pulse of the tweeting.

- Second, in terms of applying the NLP and data mining approach (training and testing) on an actual data, we extract the text data and labels from the partitioned tables. as the airlines-sentiments tables shown in Figure 1:


dataTrain

10541x15 table

	1 tweet_id	2 airline_sentiment	3 airline_sentiment_confidence	4 negativereason	5 negativereason_confidence	6 airline	7 airline_sentiment_gold	8 name	9 negativereason_gold	1 retweet
1	5.7031e+17	neutral	1	<undefined>	NaN	Virgin Amer...	""	cairdin	""	
2	5.7030e+17	negative	1	Can't Tell	1	Virgin Amer...	""	jnardino	""	
3	5.7030e+17	negative	1	Can't Tell	0.6842	Virgin Amer...	""	jnardino	""	
4	5.7030e+17	positive	0.6745	<undefined>	0	Virgin Amer...	""	cjmcginnis	""	
5	5.7030e+17	neutral	0.6340	<undefined>	NaN	Virgin Amer...	""	pilot	""	
6	5.7030e+17	positive	0.6559	<undefined>	NaN	Virgin Amer...	""	dhepburn	""	
7	5.7030e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	YupitsTate	""	
8	5.7029e+17	neutral	0.6769	<undefined>	0	Virgin Amer...	""	idk_but_yo...	""	
9	5.7029e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	HyperCami...	""	
10	5.7029e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	sjespers	""	
11	5.7028e+17	negative	0.6842	Late Flight	0.3684	Virgin Amer...	""	smartwater...	""	
12	5.7028e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	ItzBrianHunty	""	
13	5.7028e+17	negative	1	Bad Flight	1	Virgin Amer...	""	heatherovie...	""	
14	5.7027e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	thebrandiray	""	
15	5.7027e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	JNLpierce	""	
16	5.7027e+17	negative	0.6705	Can't Tell	0.3614	Virgin Amer...	""	MISSGJ	""	
17	5.7026e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	DT_Les	""	
18	5.7026e+17	neutral	1	<undefined>	NaN	Virgin Amer...	""	rjlynch21086	""	
19	5.7026e+17	negative	1	Customer Service ...	0.3557	Virgin Amer...	""	ayeevickiee	""	
20	5.7025e+17	negative	1	Customer Service ...	1	Virgin Amer...	""	Leora13	""	
21	5.7024e+17	negative	1	Can't Tell	0.6614	Virgin Amer...	""	meredithjly...	""	
22	5.7021e+17	negative	1	Bad Flight	1	Virgin Amer...	""	blackjackpr...	""	
23	5.7012e+17	neutral	0.6150	<undefined>	0	Virgin Amer...	""	TenantsUps...	""	
24	5.7009e+17	negative	1	Customer Service ...	1	Virgin Amer...	""	Cuschoolie1	""	
25	5.7005e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	Nicsplace	""	
26	5.7005e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	Nicsplace	""	
27	5.7004e+17	neutral	0.6791	<undefined>	0	Virgin Amer...	""	elisha_malu...	""	
28	5.7004e+17	negative	1	Customer Service ...	1	Virgin Amer...	""	DannyDoug...	""	

Fig. 1 Text data and class labels extraction from the training dataset.

- The same step is applied on the testing dataset. as is shown in Figure 2 below:

- convert document to sequence step for the training and testing datasets as are shown in Figure 3 and 4:

- in this step labels extraction procedure implemented. and shown in Figures 5:

dataTest

1171x15 table

	1 tweet_id	2 airline_sentiment	3 airline_sentiment_confidence	4 negativereason	5 negativereason_confidence	6 airline	7 airline_sentiment_gold	8 name
1	5.7030e+17	positive	0.3486	<undefined>	0	Virgin Amer...	""	jnardino
2	5.7030e+17	neutral	0.6837	<undefined>	NaN	Virgin Amer...	""	yvonnaly
3	5.7029e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	HyperCa
4	5.7026e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	ElvinaBe
5	5.7022e+17	neutral	0.6854	<undefined>	NaN	Virgin Amer...	""	AdamSin
6	5.7001e+17	positive	0.6570	<undefined>	NaN	Virgin Amer...	""	joyabsalc
7	5.6997e+17	positive	0.6922	<undefined>	NaN	Virgin Amer...	""	Travelzoo
8	5.6994e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	mrmicha
9	5.6993e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	artisticwr
10	5.6993e+17	negative	1	Can't Tell	1	Virgin Amer...	""	GunsNDi
11	5.6992e+17	negative	1	Can't Tell	0.6513	Virgin Amer...	""	MaryAnr
12	5.6989e+17	negative	1	Late Flight	0.3486	Virgin Amer...	""	GunsNDi
13	5.6977e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	SamBritt
14	5.6962e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	SkateMa
15	5.6962e+17	neutral	0.3550	<undefined>	0	Virgin Amer...	""	FiDiFami
16	5.6961e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	Hollywoo
17	5.6930e+17	negative	1	Can't Tell	1	Virgin Amer...	""	MarwaYo
18	5.6928e+17	negative	1	Cancelled Flight	1	Virgin Amer...	""	FiDiFami
19	5.6926e+17	positive	1	<undefined>	NaN	Virgin Amer...	""	carabube
20	5.6924e+17	neutral	0.6340	<undefined>	0	Virgin Amer...	""	floatinga
21	5.6922e+17	negative	1	Flight Booking Pr...	0.6954	Virgin Amer...	""	propsonl
22	5.6921e+17	positive	0.7081	<undefined>	NaN	Virgin Amer...	""	stefanpir
23	5.6899e+17	neutral	1	<undefined>	NaN	Virgin Amer...	""	AirlineFu
24	5.6893e+17	neutral	1	<undefined>	NaN	Virgin Amer...	""	AirlineFu
25	5.6890e+17	negative	1	Bad Flight	0.6761	Virgin Amer...	""	MsReese
26	5.6882e+17	negative	1	Customer Service ...	1	Virgin Amer...	""	miekd
27	5.6878e+17	negative	1	Flight Booking Pr...	0.6667	Virgin Amer...	""	perMicah
28	5.6866e+17	negative	1	Flight Booking Pr...	0.6841	Virgin Amer...	""	davidhfe

Fig. 2 Text data and class labels extraction from the testing dataset.

- The results of the NLP pre-processing steps for the training dataset are shown in Figure 6:
- make everything lower case step re shown in Figure 7:
- replace hashtag abbreviations with standard language step shown in Figure 8:
- Vocabulary list construction step shown in Figure 9:
- Bag of Words (BoW) Construction for the training dataset as is shown below in Figure 10:
- Training dataset features vector construction as is shown in Figure 11:
- Training dataset labels extraction and conversion step as Figure 12 shows below:
- in this step rank the feature space using the Mutual Information as is shown in figure 13:

  which the total number of the selected features is (536) as is shown in Figure 14:
- Finally, the new Feature Domain that is selected based dimensionality reduction and feature selection approach is (11712x536) features as is shown in Figure 15:

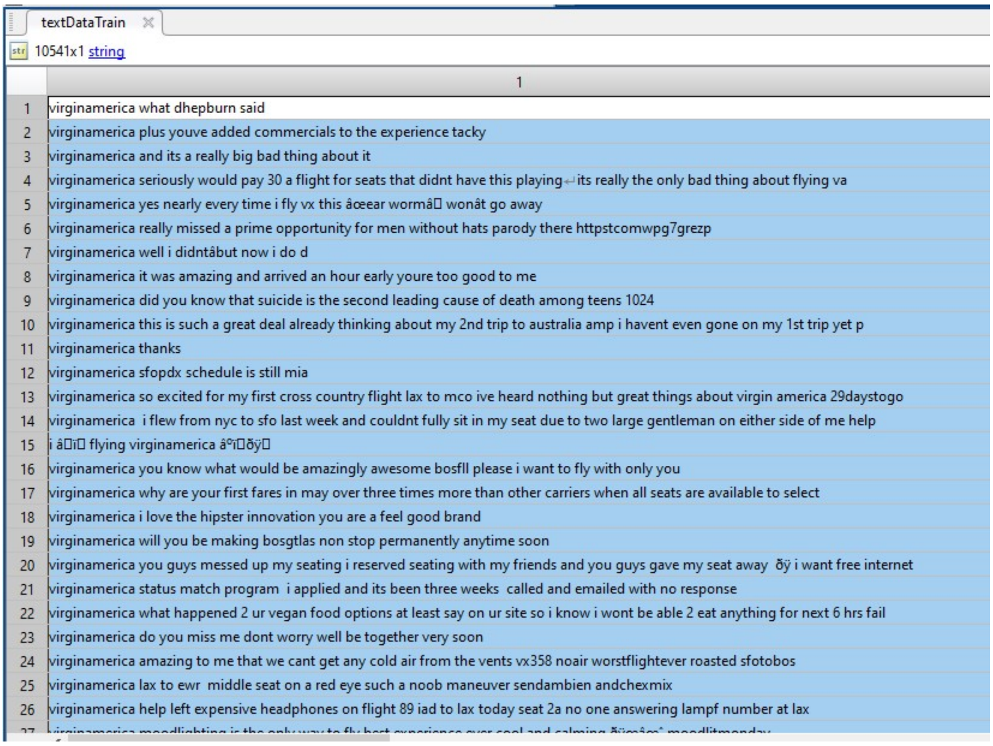


Fig. 3 Text data extraction and document to sequence converting for the text training dataset.

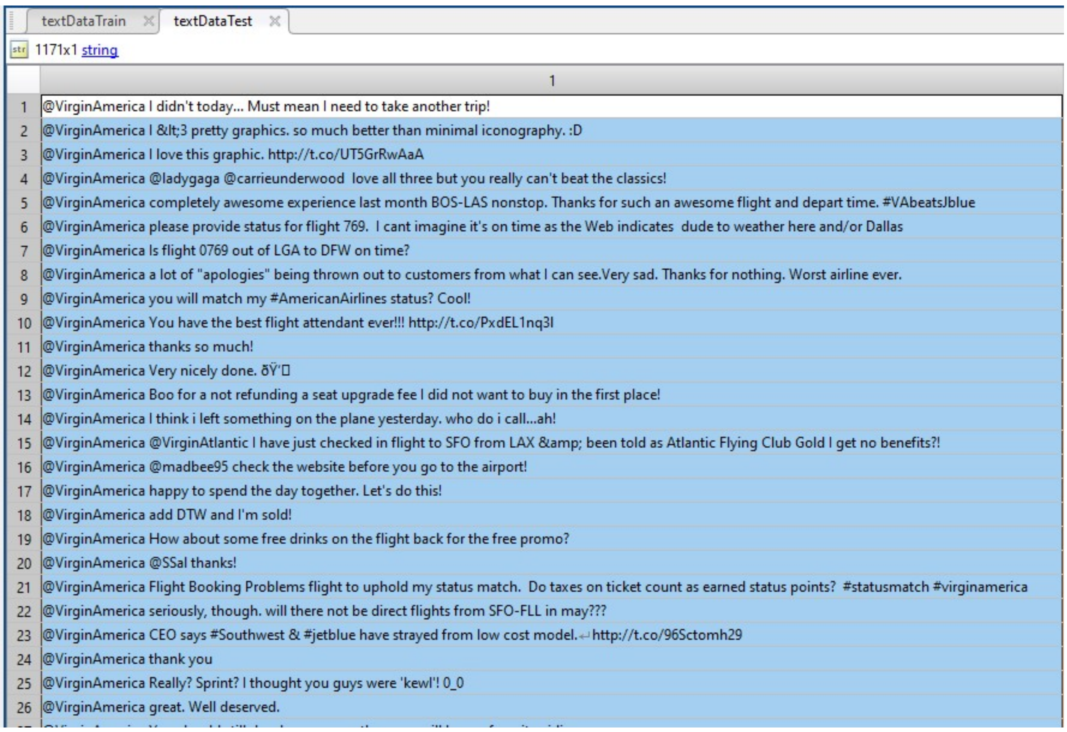


Fig. 4 Text data extraction and document to sequence converting for the text testing dataset.

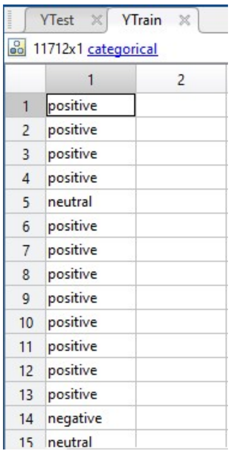


Fig. 5 Training and testing text labels extraction.

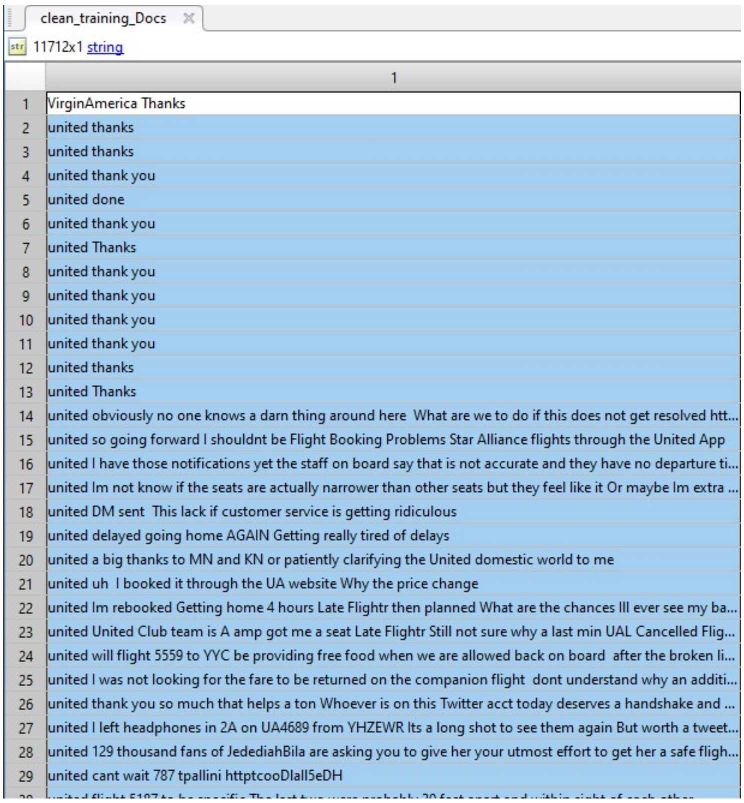


Fig. 6 NLP based text processing for the training dataset (erase punctuation step).

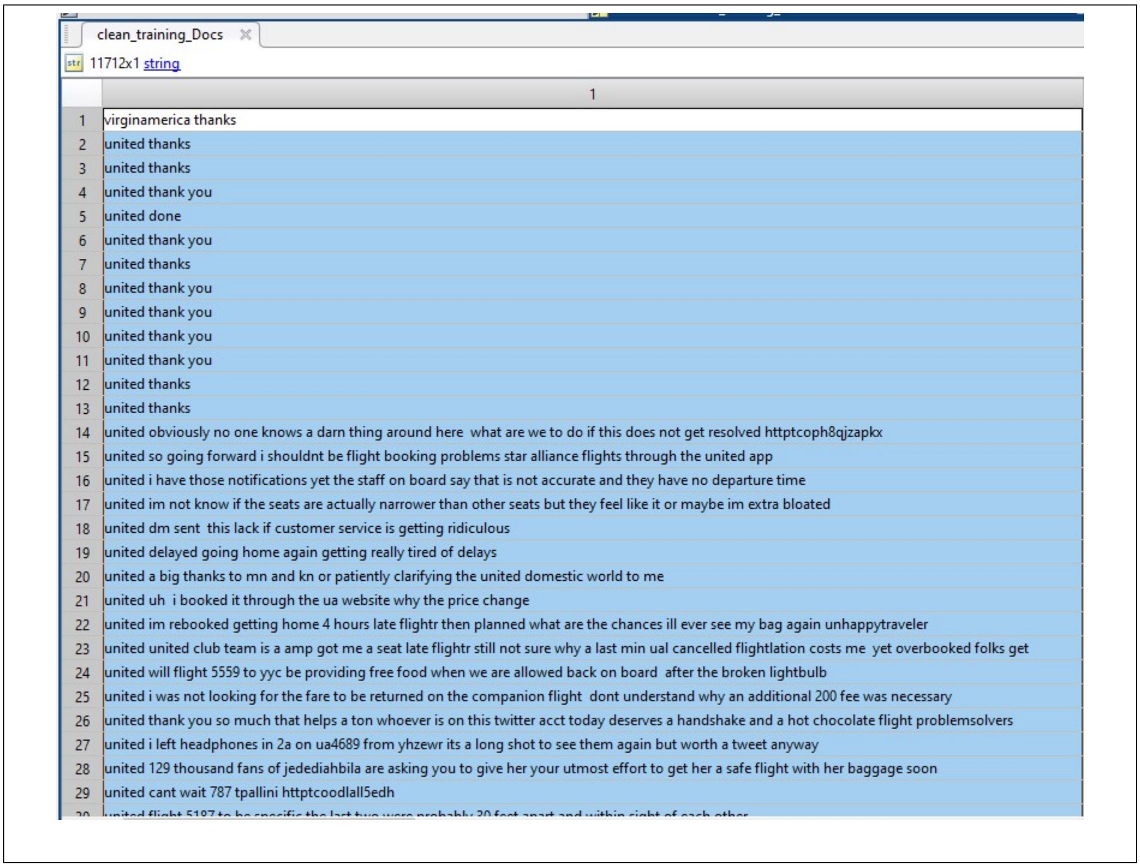


Fig. 7 NLP based text processing for the training dataset (make everything lower case step).

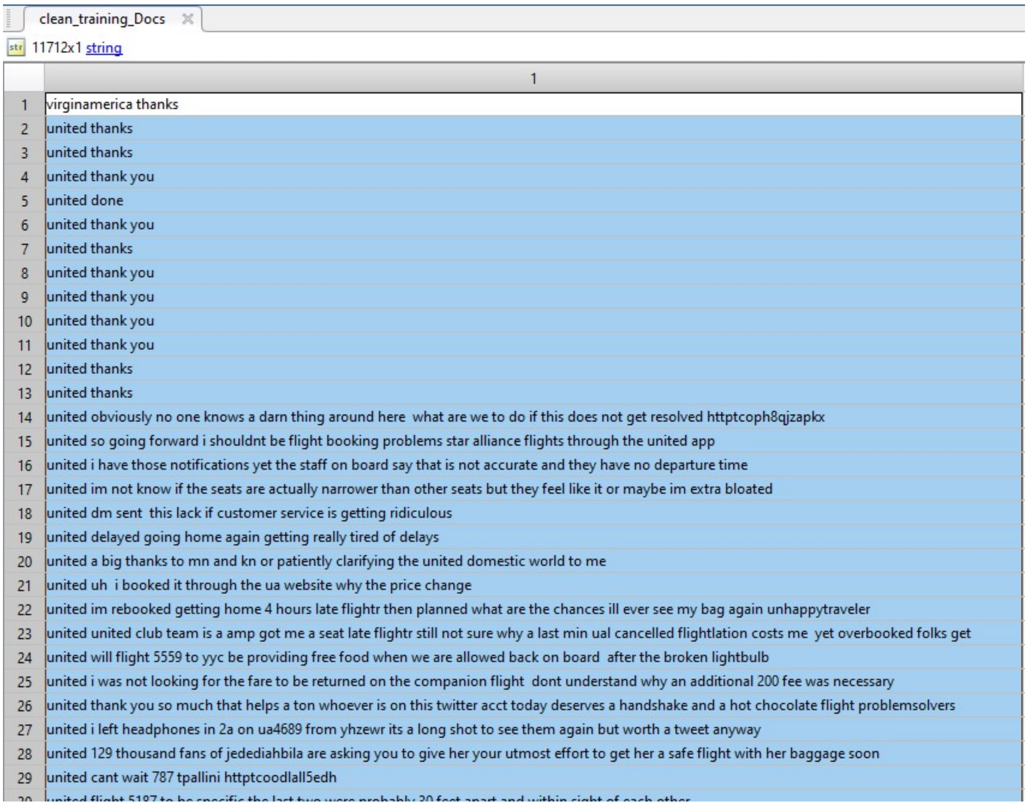


Fig. 8 NLP based text processing for the training dataset (replace hashtag abbreviations with standard language.

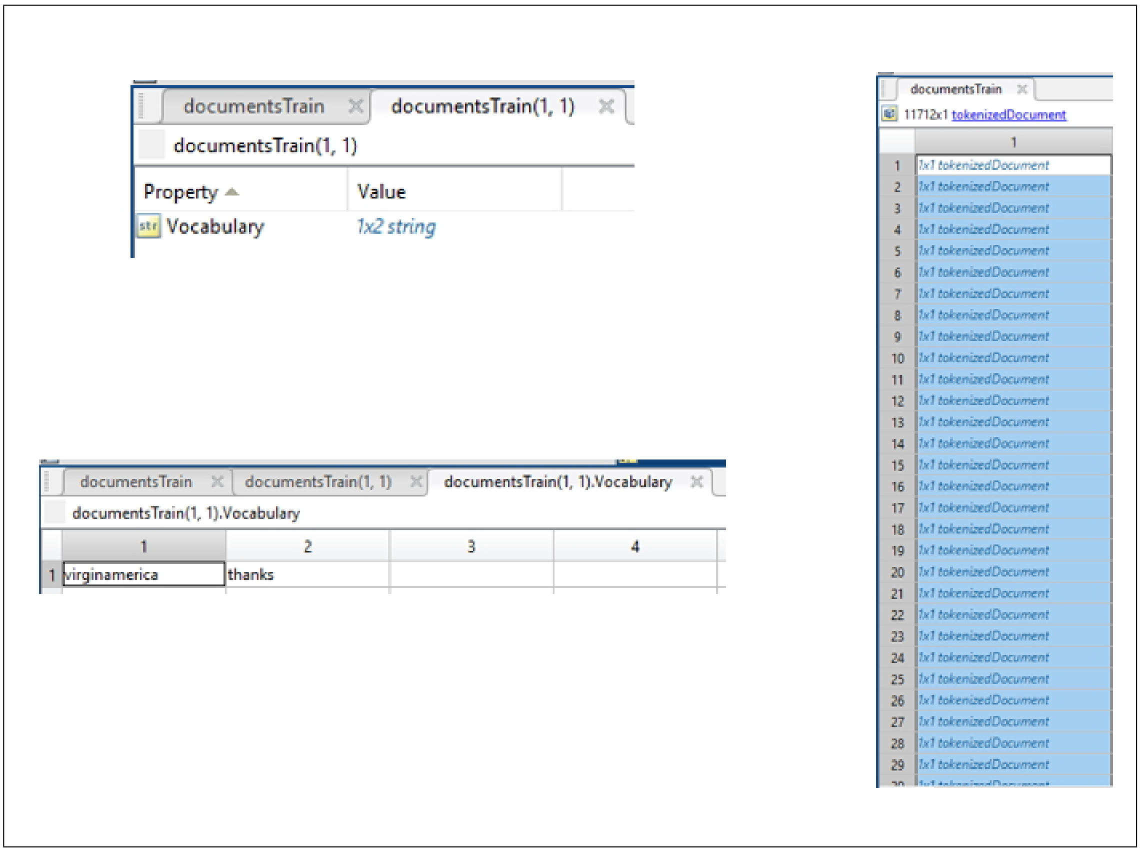


Fig. 9 NLP based text processing for the training dataset (Vocabulary list construction).

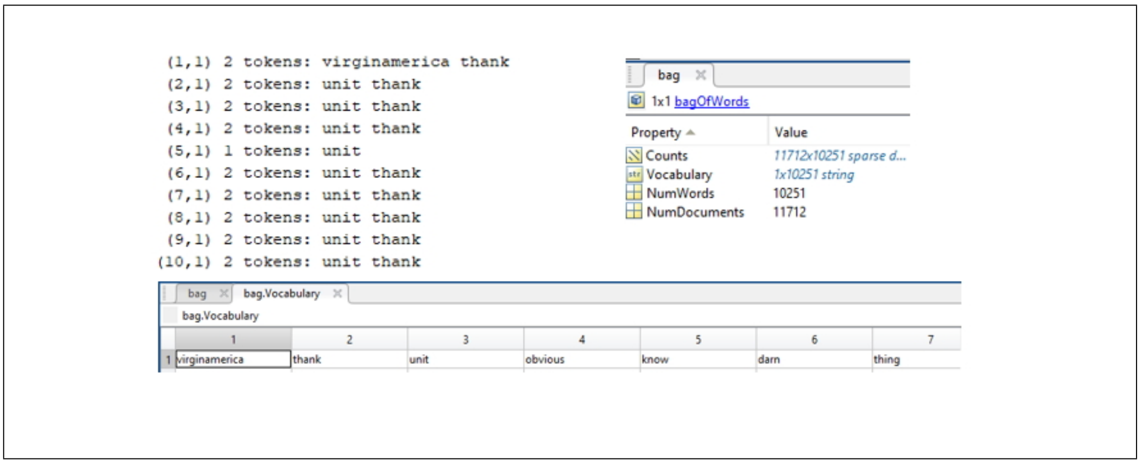


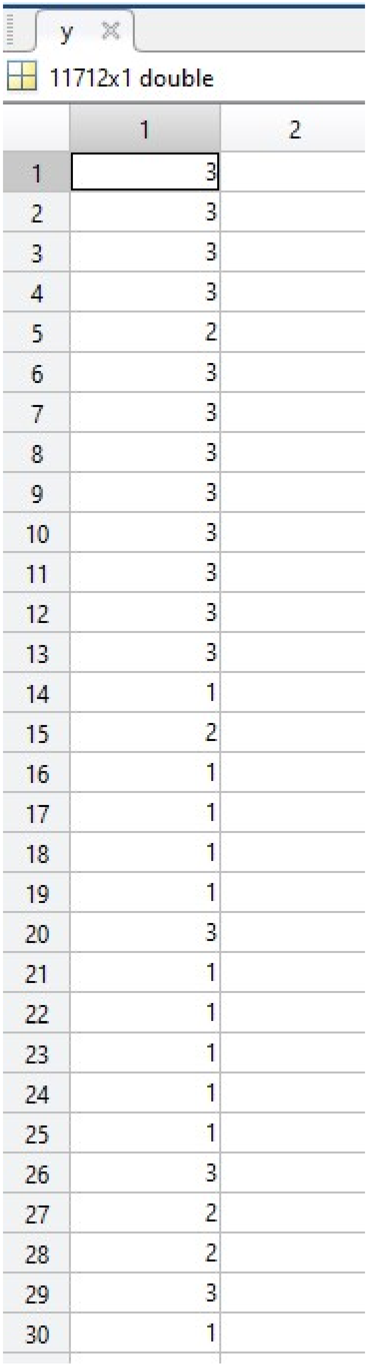
Fig. 10 Histogram of Words (HoW) or Bag of Words (BoW) Construction for the training dataset.

XTrain1

11712x2976 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
15	0	0	2	0	0	0	0	0	0	0	1	1	1	2	1	1
16	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
18	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
19	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
20	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
22	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
23	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0
24	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
25	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
26	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
27	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
29	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
31	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
32	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0

Fig. 11 Training dataset features vector construction.



y

11712x1 double

	1	2
1	3	
2	3	
3	3	
4	3	
5	2	
6	3	
7	3	
8	3	
9	3	
10	3	
11	3	
12	3	
13	3	
14	1	
15	2	
16	1	
17	1	
18	1	
19	1	
20	3	
21	1	
22	1	
23	1	
24	1	
25	1	
26	3	
27	2	
28	2	
29	3	
30	1	

Fig. 12 Training dataset labels extraction and conversion.

features

2976x1 double

	1	2
1	2	
2	54	
3	2123	
4	463	
5	754	
6	560	
7	39	
8	71	
9	14	
10	538	
11	115	
12	245	
13	931	
14	36	
15	123	
16	2513	
17	134	
18	309	
19	334	
20	142	
21	37	
22	313	
23	61	
24	2516	
25	69	
26	50	
27	670	
28	433	
29	635	
30	514	
31	66	
32	200	
33	173	
34	238	

(a) features number

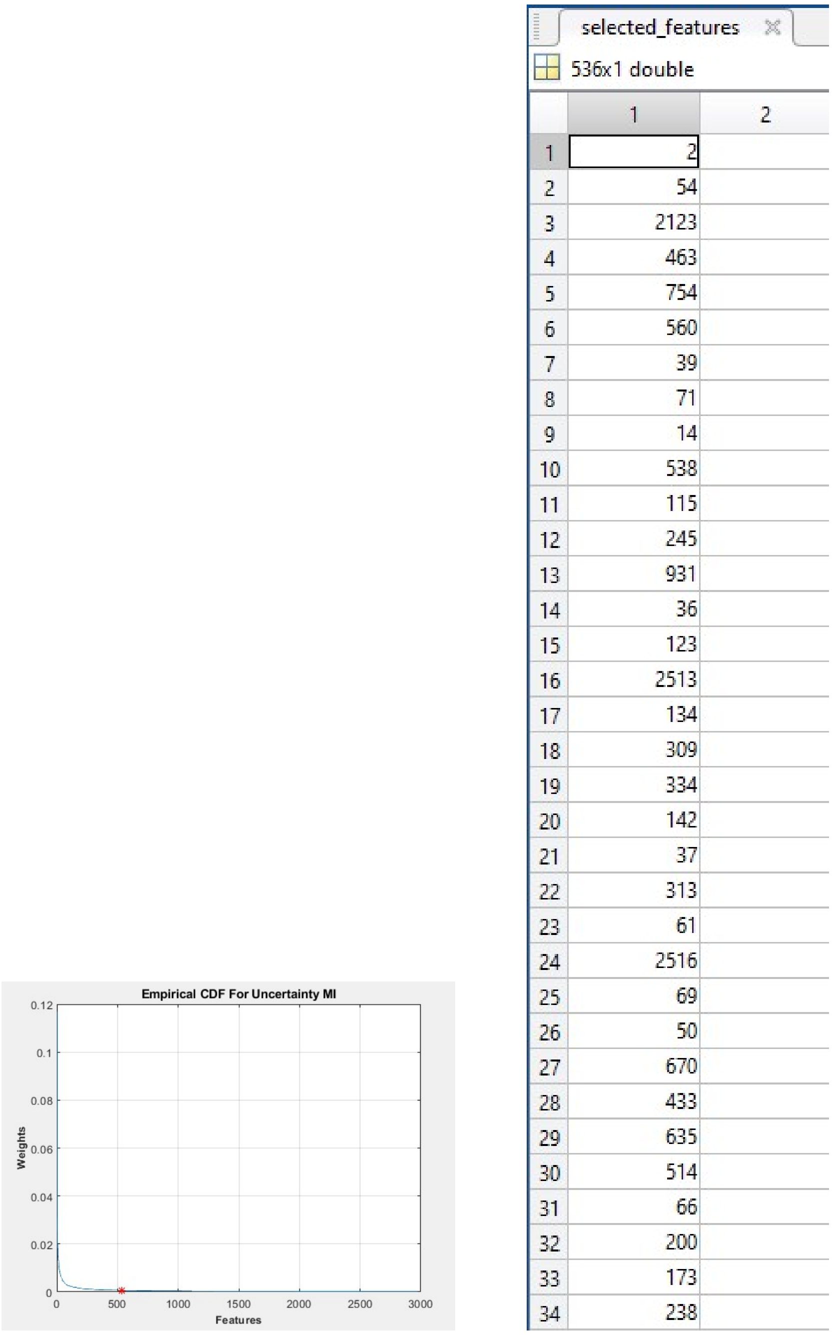
weights

2976x1 double

	1	2
1	0.1167	
2	0.0364	
3	0.0277	
4	0.0252	
5	0.0250	
6	0.0200	
7	0.0181	
8	0.0180	
9	0.0156	
10	0.0143	
11	0.0141	
12	0.0140	
13	0.0132	
14	0.0129	
15	0.0123	
16	0.0108	
17	0.0107	
18	0.0103	
19	0.0098	
20	0.0093	
21	0.0090	
22	0.0088	
23	0.0088	
24	0.0087	
25	0.0082	
26	0.0081	
27	0.0077	
28	0.0073	
29	0.0071	
30	0.0070	
31	0.0068	
32	0.0065	
33	0.0065	
34	0.0063	

(b) MI scores

Fig. 13 Mutual Information (MI) features ranking, (a) features number, (b) MI scores



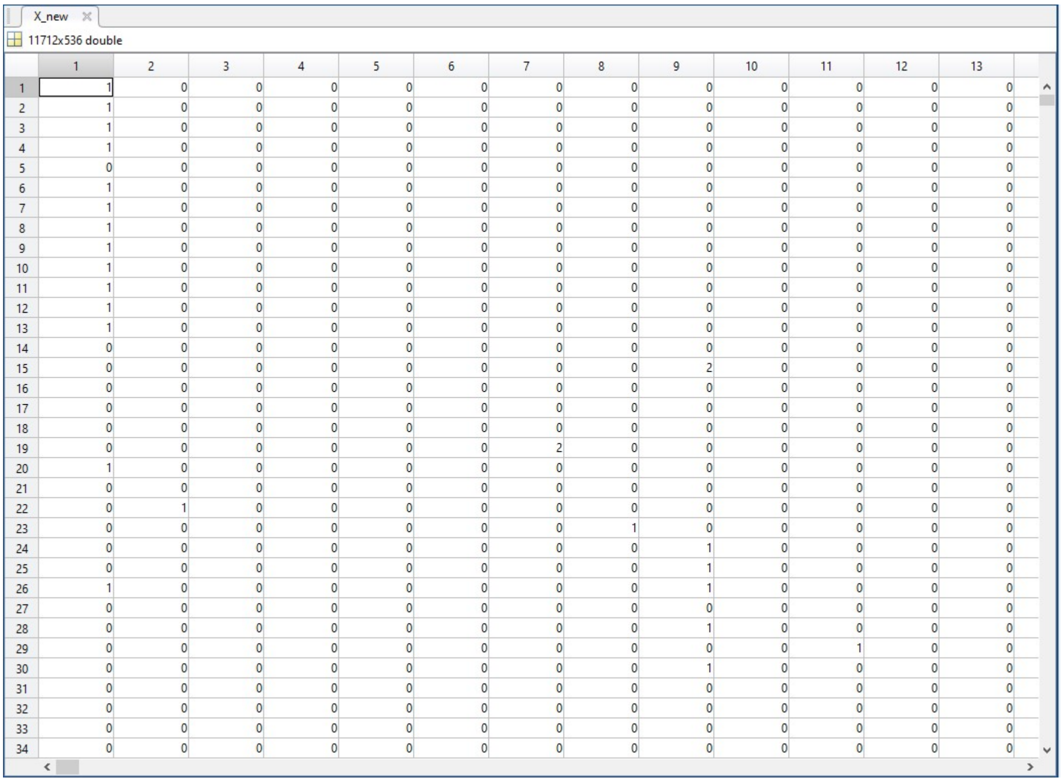
selected_features

536x1 double

	1	2
1	2	
2	54	
3	2123	
4	463	
5	754	
6	560	
7	39	
8	71	
9	14	
10	538	
11	115	
12	245	
13	931	
14	36	
15	123	
16	2513	
17	134	
18	309	
19	334	
20	142	
21	37	
22	313	
23	61	
24	2516	
25	69	
26	50	
27	670	
28	433	
29	635	
30	514	
31	66	
32	200	
33	173	
34	238	

(a) UDP point (threshold) (b) Sleeted features

Fig. 14 CDF function for the Uncertainty Point Detection (UPD) (a) UDP point (threshold), (b) Sleeted features.



X_new

11712x536 double

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	2	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	2	0	0	0	0	0	0
20	1	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	1	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	1	0	0	0	0	0
24	0	0	0	0	0	0	0	0	1	0	0	0	0
25	0	0	0	0	0	0	0	0	1	0	0	0	0
26	1	0	0	0	0	0	0	0	1	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	1	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	1	0	0
30	0	0	0	0	0	0	0	0	1	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 15 Dimensionality reduction and feature selection results after applied our CDF and UPD algorithms.

List of Abbreviations

SA Sentiment Analysis

NLP Natural Language Processing

IR Information Retrieval

ML Machine Learning

SVD Singular Value Decomposition

MI Mutual Information

PCA Principal Component Analysis

NN Neural Network

BNN Backpropagation Neural Network

DL Deep Learning

ANN Artificial Neural Network

CNN Convolutional Neural Network

DANM Deep Attention Network Mechanism

SVM Support Vector Machine

NFM Non-Negative Matrix Factorization

FCM Fuzzy C-means

LDA Latent Dirichlet Allocation

VAC Viewer Affect Concept

SVR Support Vector Regression

GCH Global Color Histograms

LCH Locl Color Histogram

BOW Bag of Visual Word

RR Recognition Rate

PR Precision

SE Sensitivity

SP Specificity

FN False Negative

FP False Positive

TN True Negative

TP True Positive

RNN Recurrent Neural Network

FS Feature Selection

FE Feature Extraction

KNN K-Nearest Neighbors

ERCOF Entropy Based Rank Sum Test and Correlation Filtering

NMF Non Negative Matrix Factorization

CRFs Conditional Random Fields

CM C-Means

FCM Fuzzy C-Means

SSP Secondary Structure Predictor

PAA Plan Acquisition Architecture

CMA-Es Covariance Matrix Adaptation Evolution Strategy

Doc2Vec Document to Vector

VAC Viewer Affect Concepts

LSA Latent Semantic Analysis

TF-IDF Term Frequency Inverse Document Frequency

DT Decision Tree

NB Naive Bayes

VA Valence Arousal

SLFFNs Single Layer Feed-Forward Networks

FFNN Feed-Forward Neural Network

BP Back Propagation

DANM Deep Attention Mechanism

SV Sentiment Visualization

ReLU

VA 2D Emotion Space

ReLU Rectified Linear Unit

Maha AlGhalibi

Address: Germany, Koblenz, 56075

EXPERIENCE

Iraq- Basrah
Lecturer
July 2013 - August 2014
Iraq University college

Iraq- Basrah
Computer teacher
October 2005 - June 2010
Al auroba school

Iraq- Basrah
Computer teacher
January 2004 - September 2005
Diginets company

EDUCATION

Germany 2021
PHD. Computer Science
University of Koblenz-landau

India 2012
MSc. Computer science
Dr. Babasaheb Ambedkar Marathwada University

Iraq 2004
BSc. Computer Science
University of Basrah

SKILLS

Programming languages skills: Java, Python, Matlab, C++.

LANGUAGES

Arabic Native language English good writing and communication skills.

PUBLICATIONS

Maha Al-Ghalibi and Kai Lawonn "Topic aspects-based generative mixture model for movie recommendation system using deep convolutional network", Proc. SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019), 114333K (31 January 2020); https://doi.org/10.1117/12.2556294

Maha Al-Ghalibi, Adil Al-Azzawi, and Kai Lawonn "NLP based sentiment analysis for Twitter's opinion mining and visualization", Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018), 110412A (15 March 2019).

Maha Alghalibi, Adil Al-Azzawi, Kai Lawonn, "Deep Tweets Analyzer Model for Twitter Mood Visualization and Prediction Based Deep Learning Approach," International Journal of Computer and Communication Engineering vol. 8, no. 1, pp. 1-17, 2019.

Maha Alghalibi, Adil Al-Azzawi, Kai Lawonn, "Deep Attention Learning Mechanisms for Social Media Sentiment Image Revelation," International Journal of Computer and Communication Engineering vol. 9, no. 1, pp. 1-17, 2020.

Maha Al-Ghalibi, Adil Al-Azzawi, and Kai Lawonn " Time series social network visualization based on dimension reduction", 26th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision co-operation with EUROGRAPHICS Association, p. 33-41