# A Novel Framework for Sentiment Analysis using Deep Learning

Andleeb Aslam
Fall 2017-MS-17(CSE)00000203554
MS-17(CSE)

Supervisor
Dr. Usman Qamar

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD

JANUARY, 2020

# A Novel Framework for Sentiment Analysis using Deep Learning


Author

Andleeb Aslam

Fall 2017-MS-17(CSE)00000203554


A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Software Engineering

Thesis Supervisor

Dr. Usman Qamar

Thesis Supervisor's Signature: ________________________________________


DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD


JANUARY, 2020

# Declaration

I certify that this research work titled "*A Novel Framework for Sentiment Analysis using Deep Learning"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

ANDLEEB ASLAM

2017-NUST-Ms-CSE-00000203554

## Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

Andleeb Aslam

Registration Number

00000203554

Signature of Supervisor

Dr. Usman Qamar

## Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. The thesis is also according to the format given by the university.

Signature of Student

Andleeb Aslam

Registration Number

00000203554

Signature of Supervisor

Dr. Usman Qamar

# Copyright Statement

## Acknowledgements

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought which You setup in my mind to improve it. Indeed, I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life.

I would also like to express special thanks to my supervisor Dr. Usman Qamar for his help throughout my thesis and also for data engineering and web engineering courses which he has taught me. I can safely say that I haven't learned any other engineering subject in such depth than the ones which he has taught..

I would also like to pay special thanks to Dr. Saad for his tremendous support and cooperation. Each time I got stuck in something, he came up with the solution. Without his help I wouldn’t have been able to complete my thesis. I appreciate his patience and guidance throughout the whole thesis.

I would also like to thank Dr. Wasi Haider and Dr. Muhammad Abbas my GEC committee for their support and cooperation. Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

*Dedicated to my exceptional parents and adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment.*

Abstract

With the advancement of Social media, data present online is also growing progressively. Internet is full of people opinions but mostly in unstructured form. This type of data is useful for different people who are interested to know about specific product accurately. To make this unstructured data useful for different purposes like improving business, customer satisfaction; different techniques of Natural language processing are used. Extracting useful data by converting the unstructured data to structured form is an important step in NLP. Sentiment analysis is used to analyze people sentiments available online in the form of blogs, reviews, comments etc. One of the main tasks of NLP is sentiment analysis. The main problem arises because of text-only reviews where people don't provide the rating but write down a whole statement for a product, organization or a restaurant. And it became difficult for business analysts to know about their customers reviews and thus it become difficult for them to improve their products according to customers demand.

Neural networks language models fail to deal with long sequence of words and also cannot deal with contextual information. Different machine learning models are used for prediction in literature. Recently new methodologies are introduced using deep neural networks like Recurrent neural networks (RNN). But RNN does not save the previous information, to overcome this issue Long-Short-Term-Memory (LSTM) are introduced. Other improved models include BiDirectional LSTM (BiLSTM), Gated Recurrent neural network (1) and Convoluted Neural Network (CNN). To improve the prediction accuracy and to generate more reliable results a novel hybrid model is introduced having five different models including RNN, LSTM, BiLSTM, GRU and CNN and Glove word Embedding. The use of word embedding techniques is also an efficient step in sentiment analysis task. The Provided rating is converted to positive and negative Sentiment firstly in order to predict the Sentiment related to a given review. The proposed model (RNN-LSTM-BiLSTM-GRU-CNN) is tested on different state of the art datasets including SST-1, SST-2, Movie Review and IMDB movie review dataset. The Results shows an improved accuracy using proposed model.

**Key Words:** *Deep neural network, Natural language processing (NLP), LSTM, BiLSTM, RNN, GRU, CNN, word Embeddings*

# Table of Contents

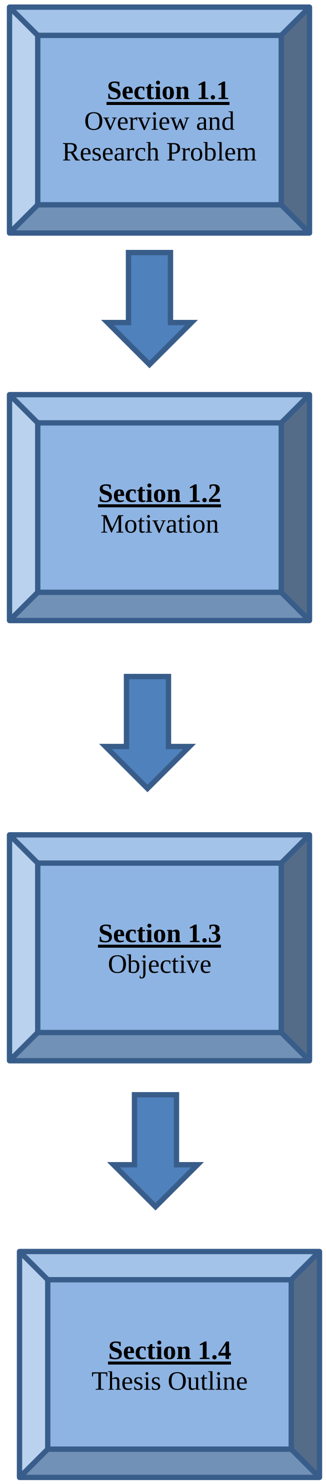
# List of Figures

# List of Tables

# Chapter 1

# Introduction

# Chapter 1: Introduction



# CHAPTER 1: INTRODUCTION

The thesis is about Sentiment analysis and deals with online reviews. The different approaches and models are also discussed in it for correctly predicting the polarity and labels against reviews available online. Special focus is on text-only reviews and improving their accuracy for the task of predicting final label.

## 1.1 Overview and Research Problem

Different researches have been conducted to analyze the sentiment polarity against reviews available online. Mostly reviews available online are not well structured and need to be structured before further steps to analyze sentiments associated with them is performed. Natural language processing (NLP) is used for this task i-e to convert unstructured data to structured form [1]. Prediction of polarity against Text-only reviews is also performed to easily determine polarity associated with customers reviews.

Different novel and efficient techniques and models are proposed for the task of SA in recent years. Researches shows remarkable work in the field of sentiment analysis and NLP. Machine learning and lexicon-based approaches and models are also present for this task. Useful information from raw and un-structured data is extracted along with useful features. But there exist some limitations as well. Deep learning models are also a type of Machine learning algorithms [2]. They result in automatic extraction of features and these features are than use as input to the model. Figure 1.1 1 explains SA and its approaches [3]:

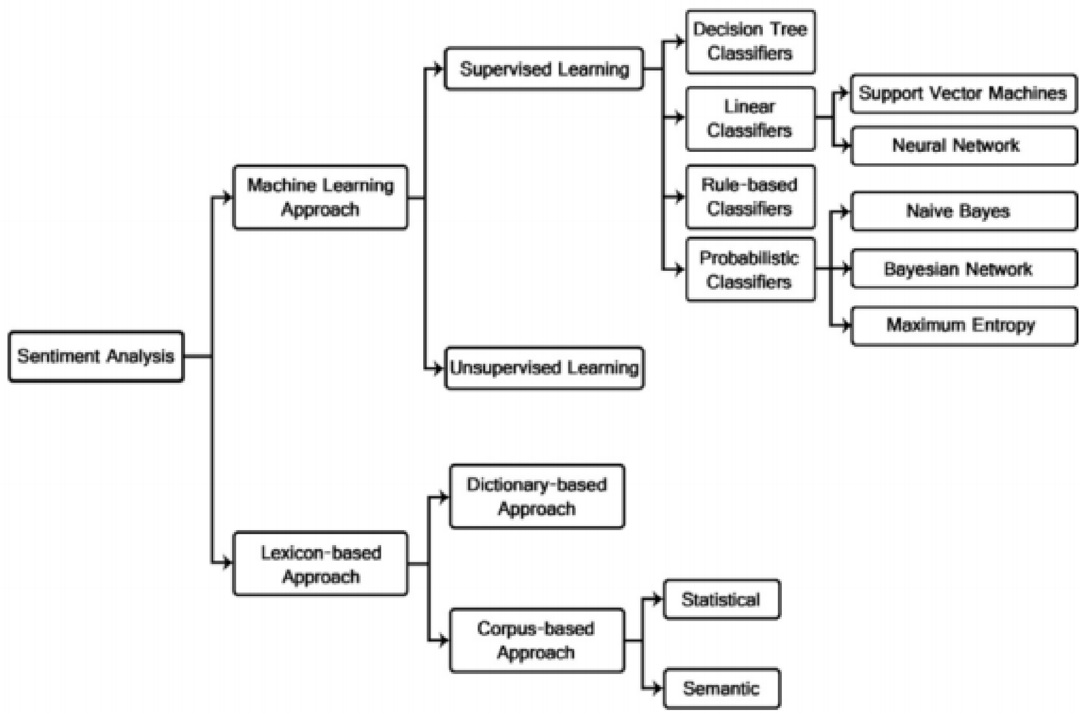



Figure 1.1-1 Sentiment Analysis Approaches

Advancement in online markets and businesses also lead to increase in data available online in the form of opinions and reviews. These reviews are important for customers as well as business analyst for better performance. Internet is full of unstructured data and this data needs to be in structured form for further utilization in tasks like sentiment analysis [4]. It has been observed that the SA task is more complex and difficult to handle than topic-based classification of text. The limited amount of text available for Sentiment analysis is a real problem as it becomes difficult to analyze the context of the words in terms of semantics. Due to vanishing gradient problem it is difficult to capture long term dependencies present in text with gradient descent using neural network language model [5]. NLP (Natural language processing) deals each word as atomic symbol and is not able to capture much detail about semantics between them. Advance and recent work of NLP involves deep neural network for text classification and prediction of ratings for text-only reviews. By increasing the depth of neural network models, complexity also increase as the number of parameters increase and it also lead to vanishing gradient problem [4],[5]. The model thus forms is also difficult to optimize and is more prone to overfitting. Different models have their own pros and cons. Our proposed model using five different type of models is able to give more reliable results. Majority voting helps in producing better and reliable results. The final result is basically dependent on all five models results and the result that gets most vote is considered as final result.

Sentence level sentiment analysis usually has less amount of data and thus less performance in predicting actual sentiment associated with the sentence. Un-supervised pre-trained word embeddings helps in analyzing the contextual meaning of the word and thus increase performance [5].

## 1.2 Motivation

Increase in people knowledge and with increasing use of social media sites, online shopping sites and movies review sites results in data explosion online especially in the form of text. It can be seen that users are now a days active on social media platforms to comment about any topic, event or a person. Different sites are available online where users can watch movies like IMDB and can leave their comments about the movie [6],[7]. Similarly, different sites like Amazon are also present for shopping online and users can leave their review about any product in the form of emoticon or in the form of text. Another important review site is of different hotels and these reviews can help the other customers and people to have an idea about the hotel before booking for it. So, these reviews are very helpful for both the customers as well as for the hotels to know about the requirements and the needs of their customers and thus helps in improving the services they provide. Hotel review sites include TripAdvisor, Airbnb etc.[8],[9]. The product owner can improve their product according to the customers need in the next version thus, improving the profit amount. But with the advancement in technologies and online shopping the users review available online also increase in number and it became difficult to handle such a large amount of data manually.

This is not practically possible to manually analyze and deal with such a massive amount of data. This is a big challenge in the field of Sentiment analysis for researchers to automatically extract sentiments from such a huge amount of data.SA also known as opinion mining is used for analyzing correct sentiment present in text using machine learning. Natural language processing techniques are also used for the purpose of SA along with machine learning algorithms [10]. Different models and algorithms are developed by researchers for this task. Classification is an important task for predicting labels against text-only review [11]. Text-only review are reviews having no rating given by the user but only the text form review is given. Deep neural networks gave good results for the prediction of labels against reviews and for sentiment classification [12].

## 1.3 Objective

The objective of this thesis is to define all steps needed for the task of sentiment analysis including pre-processing of text of reviews available online. Movie reviews are used as input to the model. The raw data is firstly cleaned using NLP toolkit. After that the neural network language models are constructed for prediction of labels against reviews. Eight different hybrid models are constructed and accuracies are compared for all these models. Following research questions are answered.

The first research question is about accuracy value. All model accuracies are compared and the final model is selected based on the highest accuracy. All hybrid models along with their accuracies are also mentioned in thesis. The models are run on different datasets available online of movie reviews like SST-1, SST-2, MR and IMDB datasets.

The next research question is of misclassified review labels. The percentage accuracy shows the number of accurate predictions. The reviews actual labels are compared with the predicted labels and thus the accuracy is calculated.

## 1.4 Thesis Outline

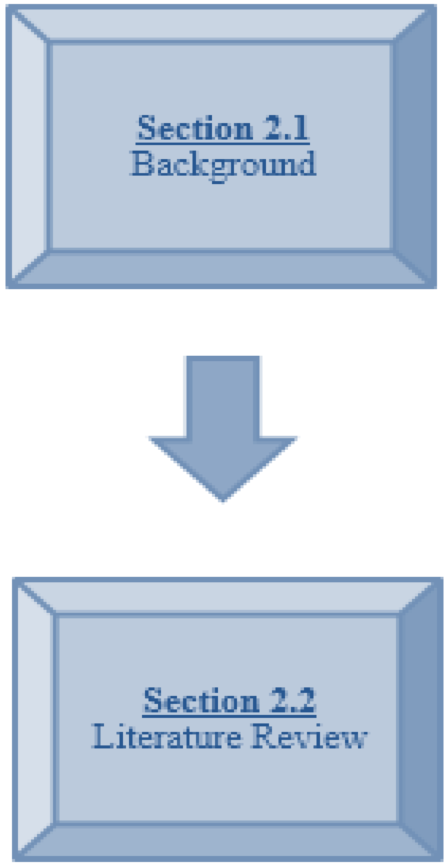
The thesis discusses in detail the task of sentiment analysis and presents an improved framework that provides results having better accuracy. Chapter 1 is all about problem statement, motivation and objectives of the thesis. The next chapter is all about background of sentiment analysis and discusses different models presented in literature for predicting polarity against text-only reviews.

Chapter 3 mentions the basics of neural networks and about different optimization techniques. Chapter 4 explains about proposed methodology and different datasets used. Datasets attributes and specifications are also mentioned in it along with how classification is performed using proposed framework. Chapter 5 discusses the architecture of models used and also in experimental setup different libraries used are mentioned. Also proposed solution is also discussed in this chapter. The next chapter i-e chapter 6 is all about results and results evaluation. The comparison of our results is also performed in this chapter.

# Chapter 2

# Background & Literature Review

# Chapter 2: Background & Literature Review



# CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

In this chapter a background about different approaches of sentiment analysis is provided. And the second part covers the existing work done in this field using different models and techniques.

Most of the organizations rely on the task of sentiment analysis of customers review to have a knowledge about their customers satisfaction level about their product and about their organisations [13]. Factors important and expected by the customers that are missing in the product or the improvements needed in the product that is already available in market can be easily be known using task of sentiment analysis and according to the customers reviews product can be improved, thus generating profit to the organisation. A vast research is being performed using conventional text (Reviews, discussion forums and blogs) and using social networks data such as tweets and microblogs [14]. This text provides information related to opinions of customers about product, personality or any topic etc.

In short, Sentiment analysis approaches can be divided into two main categories namely [15]:

- Traditional approaches

It depends on words present in text along with syntactic features that reflects sentiment present in text.

- Semantic approaches

In this type of approach for sentiment analysis task not only the latent meaning of the word is considered but also the context in which the word is being used is also considered and accordingly sentiment orientation is updated.

In this chapter overview of sentiment analysis and the existing work done in this field is mentioned targeting mainly the areas involving reviews. In particular different dimensions and main elements of the problems related to sentiment analysis are introduced in section 2.1. Literature review on different approaches for the task of sentiment analysis is mentioned in section 2.2.

## 2.1 Background

This section provides basic understanding and knowledge about sentiment analysis, which is the basis of the research conducted.

### 2.1.1 Fundamentals

Sentiment analysis is the task of identifying different polarity of reviews related to any product, individual or event. Polarity can be either in binary or in labels form as positive, negative or neutral. The task of sentiment analysis is multidisciplinary involving different techniques such as Natural language and machine learning for identifying the polarity of text under consideration [16].

Sentiment analysis is a vast problem comprising of different tasks, aspects and dimensions. A vast amount of research has conducted in this area. The problem of analysing sentiment can broadly be divided into four major categories or dimensions summarised below [15],[16]:

1. The granularity of text determines sentiment analysis level like Sentence level, document level and phrase level etc.
2. The sentiment analysis tasks more than just finding the polarity of the text like mood prediction, detection of sentiment strength and emotion analysis.
3. The approach of sentiment analysis which defines the type of sentiment analysis approaches like hybrid, lexicon and supervised approaches.
4. Conventional as well as microblogging data are used to analyze and extract sentiments present in text.

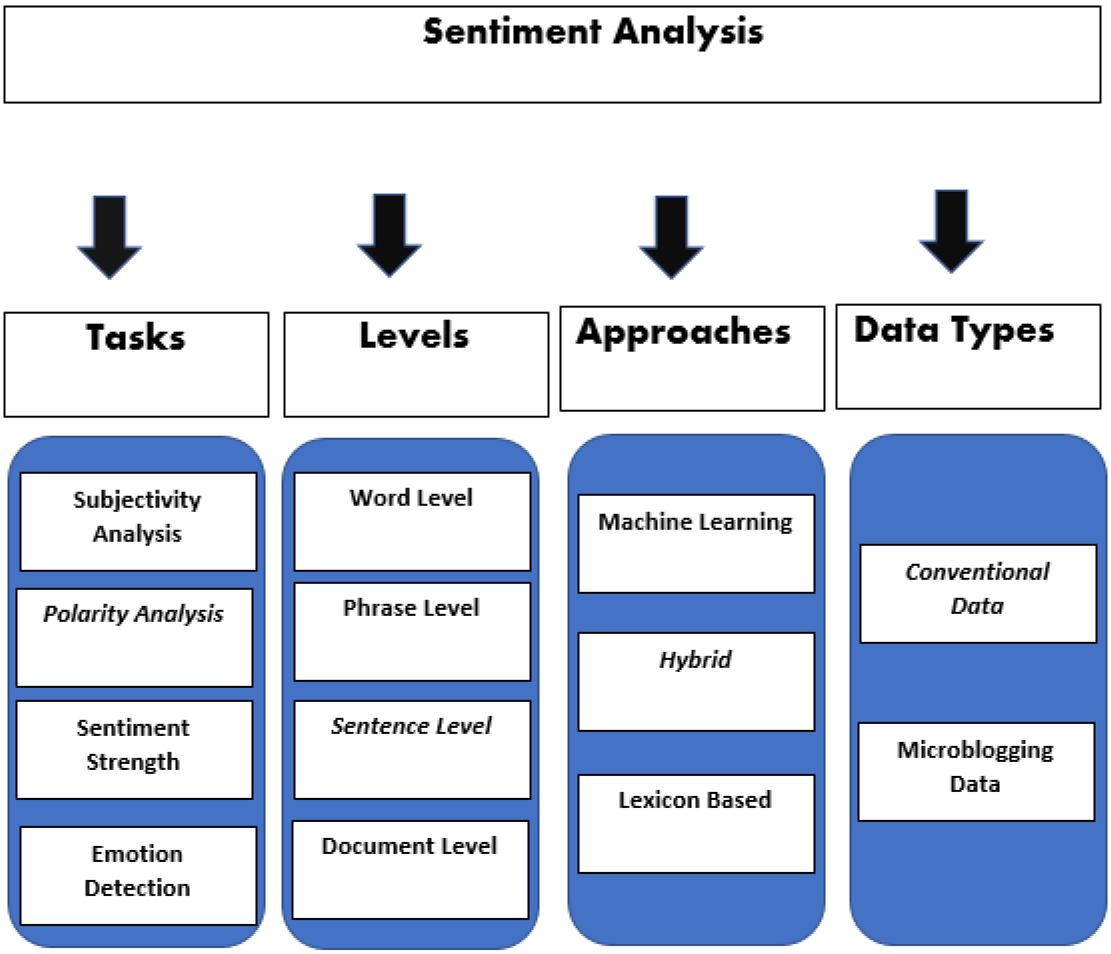



Figure 2.1-1 Dimensions of Sentiment Analysis

To explain these dimensions of sentiment analysis an example of customers review is given [16]:

*"(1)I have recently upgraded to iPhone 5, (2) I am not happy with the screen size, (3) It is just too small ☹. (4) My best friend got Galaxy Note3.*

*(5) He got a much Larger screen than mine!!! (6) even his hardware outperforms mine, easily outrun Him in with all apps I can get ☺"*

#### 2.1.1.1 Sentiment Analysis Tasks:

Polarity detection is an important subtask of sentiment analysis. Our proposed model is focusing on finding correct polarity of the given text. It detects whether the given text is positive or negative. In the above review example 2 and 3 are of negative polarity while 6 shows a positive polarity.

The next important concern is whether the text is subjective or objective in nature. The sentences and text showing any factual information are objective like 1 and 4 are objective in nature. While all other sentences are subjective.

The task of analysing sentiment is not only limited to polarity detection and determining subjective and objective nature of the text but is also concerned with emotion detection such as "happiness", "Sadness", "angry "and strength of the emotion in sentence like, sentence 5 is showing intensity in emotion. Sentence 3 and 6 shows sad and happy emotions respectively.

#### 2.1.1.2 Sentiment analysis (Levels):

There are four different levels are granularity of text. Extensive research in literature has also been done in this area. All these levels are briefly explained below [18]:

- **Word-level:**

  In word level sentiment analysis, each word is considered to express sentiment. Each word W in a sentence S represents an opinion. If this opinion word is predefined like to a product, personality or a place, this is called as entity-level sentiment detection. In the above example "iPhone" and "Note 3" is an entity from the domain of "product" where iPhone receives an overall negative Sentient while Galaxy Note 3 has positive polarity.

- **Phrase / Expression Level:**

  Expressions in a sentence determines the overall polarity in this type. Different types of expressions can exist like in the example mentioned, the phrase "I am not happy" shows negative sentiment.

- **Sentence Level:**

A Sentence has multiple expressions and words present in it. The overall sentence determines the polarity and thus sentiment of a sentence. For example, sentence 2 and 6 has negative and positive polarity respectively.

- **Document Level:**

  In this type of sentiment analysis, the polarity of each sentence is considered and then the overall polarity of a document is calculated based on it. This is done by averaging sentences polarity.

#### 2.1.1.3 Sentiment Analysis Approaches:

**S**entiment analysis applications are in many areas like marketing, business, politics. Vast amount of research has been conducted in this area resulting in new models for this task. Basically, there are two different type of approaches of Sentiment analysis [19]:

##### 2.1.1.3.1 Lexicon Based Approach

I

n this type of approach, the overall sentiment of a given document is calculated and determined by averaging the polarity of words and phrases present in it [19]. Pre-built dictionaries are used in the lexicon type of approach like SentiWordNet , LIWC lexicon and MPQA lexicon[22].

##### 2.1.1.3.2 Machine Learning

This algorithm and approaches deal with text classification problem. Deep Learning is also included in Machine learning. Annotated data is used as input to machine learning algorithms like Naïve Bayes and Maximum entropy. Annotated data is already labelled data. Training is done using this labelled corpora and unseen data is than used to analyse the sentiments of it. In literature authors has used movie review dataset using mainly three classifiers including Naïve Bayes, Maximum entropy and SVM. This work is done by Pang and Lee in year 2002[20]. Many other approaches are also used using variations in the already proposed model [21].

#### 2.1.1.4 Data Types:

Vast amount of textual data is present online on social networking sites like Facebook, Myspace and also present on other microblogging sites as Twitter, Tumbler along with on different discussion forums and news portal etc.

Broadly analysing there are two types of data [23]:

- Conventional Data
- Microblogging Data

Conventional data is a type of data which is well structured and manually annotated for the purpose of sentiment analysis. Reviews available online are an important source of knowing customer expectations regarding the product along with it this type of data also include data and opinions of customers available on discussion forums, microblogging sites etc. A large number of corpuses are available online like MR dataset, Customer review dataset and Amazon food review dataset etc. conventional sentiment analysis is another name when sentiment analysis is done using this type of data. In our research work the dataset used is also of conventional datatype.

The other type of data is microblogging data from twitter and Facebook. A lot of online data is being uploaded every minute on different social networking sites like Facebook. Exponentially data is being produced. This area is new in the field of sentiment analysis. The data of this type is not structured like conventional data type but is unstructured and not well formatted. The size of the data is also smaller in size than other data type.

## 2.2 Literature Review

In the paper [24] author is using Amazon review dataset scrapped from web to perform task of sentiment analysis by using different algorithms of machine learning including SVM, NB and ME. The results are classified into three classes namely positive, negative and neutral. The results are useful for customers as well as companies to improve their business.

In the paper [25] the author proposed a novel model (CNN-LSTM) for sentiment analysis. In the proposed method not, all data is provided to CNN as input but different regions are made of data. Each sentence acts as a region, that divides the input to multiple regions. It helps in weighing and to extract useful information from text, that ultimately provide better prediction of rating for text of Valence arousal (VA).

In the paper [26] an ensemble approach using majority voting is applied to analyse sentiments of Airline data of customer reviews collected from twitter along with other six individual classifiers including C4.5 Decision Tree, SVM, Bayesian Network, Naive Bayes and Random Forest algorithms. Ensemble approach outperform other classifiers with an accuracy of 91.7 %. To validate the classifiers 10 cross validation is also performed.

In this paper [27] author proposed model integrating CNN and LSTM that provides advantages of both models in a single model. Local dependencies are captured by the CNN employed filters. While LSTM captures the long distance relationship. Freezing technique is proposed to avoid overfitting while training.

Author of paper [28] proposed a novel method using unified feature set and word embedding to analyse sentiments of user reviews. Deep learning model RNSA is proposed having RNN which is composed of LSTM. Performance is highly improved using this model.

In paper [29] NgramCNN is proposed using pre-trained word embeddings and a simple classifier having only a single layer. Author works on feature selection to improve the performance of sentiment analysis. IMDB dataset is used and achieved the performance of 91.2 %.

Paper [30] used a google's word2vec algorithm for the task of text classification. It maps each word to its vector form so that basic algebra operations can be performed to calculate the similarity in text available for the task of sentiment analysis. The dataset used is movie review dataset having 50,000 reviews and the models used are RNN and decision tree.

In this paper [31] author used huge amount of data form Social Network Service (SNS) for the purpose of sentiment analysis. Such a huge amount of data can't be handled easily so, deep learning models are implemented in this study. Using RNNs and CNNs the author proposed a deep learning architecture to predict the polarity of text data available.

In this paper [32] an unsupervised method for sentiment analysis is proposed by the use of word embeddings of large data from twitter corpora. This method calculates similarity and syntactic relation between the words of available text. The word semantic polarity score and the n-gram features are used along with word embedding to form the feature set for the task of analysing sentiments of text. Convolution neural network receives this feature set as input for the task of predicting polarity of tweets. Accuracy is improved using the novel model when performed on five different datasets obtained from twitter.

In the paper [33] author proposed a novel approach to obtain semantic information by the use of word embedding using a lexicon-enhanced LSTM model. Word sentiment classifier is first trained using sentiment lexicon and then embedding of words representing sentiments along with words not present in lexicon is used. Word representation becomes more accurate when sentiment lexicon and the embedding of word is used. The proposed LSTM improves the accuracy for predicting polarity of sentiments than existing models.

In this paper [34] for the task of sentiment analysis different ensemble models are proposed and by extracting features the accuracy and performance of different models of deep learning is improved. Manual extraction of features is done for the proposed ensemble models. The classification is performed by using different feature extraction methods and it results in improvement in accuracy and performance. Different datasets are used for evaluation of results of movie reviews.

The author proposed a novel neural network model in the paper [35] for the sentiment analysis. Traditional approaches and different models in literature fails to capture behavioral meaning along with textual meaning. The model proposed by the author can capture behavioral as well as textual meanings thus resulting in improved accuracy. CNN is used for the task of sentiment analysis. Tweets are used for the SA task and it improves the accuracy. The main architecture has different layers including convolution, pooling and the softmax function is also used. Total of 40 different features are used in the proposed framework along with word embeddings. SentiStrength is used for labelling the tweets for the proposed model. The results show that the proposed model outperforms existing models and is more advanced.

An ensemble is proposed in the paper [36] using aspect base sentiment analysis along with feature extraction. The features are extracted using PSO (particle swarm optimization). In the first step different terms are extracted for the purpose of feature extraction and in the second step the task of classification is performed based on the sentiments. The ensemble of different models is proposed having basically three classifiers as base classifiers including ME, SVM an CRF. The proposed ensemble model is given a reduced feature set rather than a complete list of features. The reduction in feature set results in better classification results for the aspect base SA task. The proposed approach is tested on different domains and using voting as well as weighted voting methods along with PSO for feature extraction.

In the paper [37] an ensemble model is proposed for polarity shift problem in sentiment analysis. The basic level of problem of sentiment analysis is dealt at document level. Negations are specifically handled using the proposed approach of polarity shift. The training is also done using the proposed model with weighted combination on different datasets. The model shows better accuracy than existing approached dealing at document level.

The author of the paper [38] deals with the task of sentiment analysis using both textual and multimodal information. The dataset used is of twitter and it is an example of visual data that is on daily basis provided by the online users of the twitter. Tree structure CNN (T- CNN) model is developed for this task along with another framework of joint SA. For the proposed framework pre-trained embeddings are used. The results show an improvement in the accuracy for twitter dataset.

The author of paper [39] contributes in the SA field by proposing a novel framework comprising CNN for visual content only. The dataset used is of twitter having only visual data. According to the author images provide better and more accurate results for analyzing sentiments. The proposed framework helps in predicting users sentiments associated with the visual contents they posted online on different social media platforms especially twitter. The model proposed using visual content as input results in better accuracy than other techniques proposed in literature. In the paper [40] a deep neural network based models proposed for sentiment analysis. CNN model is used for analysing sentiments of visual content. Along with CNN for visual contents Paragraph vector model is also used for dealing with textual content. Twitter data is used for final prediction. In this contribution [41] the author mainly deals with cyber crime and the main persons that are involve in this activity. The proposed framework involves deep learning to solve the problem along with sampling method snow ball. By the sentiment analysis using deep learning the facilitators of this crime activity are identified.

The author of the paper [42] shows the impact of deep CNN for capturing the semantics of short text as tweets and reviews of movies. Movies datasets includes the famous SSTb and STS datasets. Semantics and context of short sentences is difficult to capture than long text and paragraphs. The proposed architecture also used word, sentence and character level representations for better performance of sentiment analysis. The results outperform other state-of-the-art results. Pre-trained word-embeddings are also used to improve results by capturing contextual information.

For the purpose of feature reduction and brand specific lexicon development, a data driven supervised approach is proposed in this paper [43]. Preprocessing is performed and lexicon development is also made for Justin Bieber brand. Artificial neural networks are used for the purpose of classification. Classification accuracy is also improved and from 181 expressions only expressions are left. It helps in analysing views about the brand. Twitter API v1.0 is used to collect data from twitter about the brand. For improving performance, the data is divided into testing and training data.

A novel deep neural network structure is proposed in this contribution [44] namely WSDNNs. Weakly shared layers of neurons are used to extract features specific to language. Four languages namely German, English, Japanese and French are used for cross lingual sentiment classification task. The information is transformed form source to target using Deep neural networks. The proposed approach is also tested on data from Amazon and results demonstrates improve in performance of classification task of cross lingual than studies available in literature.

Table 2.2-1 Comparison Between State-of-art Techniques

| | **Approach and Idea** | **Datasets** | **Model** | **Main Contribution** | **Accuracy** | **Ref.** |
|---|---|---|---|---|---|---|
| 1 | Different types of Ensemble using basic classifiers and deep learning models are proposed. | Seven public datasets are used including movie review (**IMDB**), STS-Gold, SemEval(2013) and (2014), PL04, Sentiment140 and PL04 | Ensemble Models using Bigrams | Improvement in Performance | 87.87 | [34] |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | A neural network model that can capture behavioural information is used (CNN) | Two Datasets from twitter are used<br>- SemEval-1(2016)<br>-SemEval-2(2016) | CNN | Provides much better results than baseline method as Naïve Bayes and outperformed in capturing behavioural information than other models | 88 | [35] |
| 3 | Aspect based sentiment analysis using cascaded framework of classifiers and feature selection is proposed. PSO is used for this purpose. | Benchmark dataset of SemEval(2014) is used for two domains of:<br>-Restaurant<br>-Laptops | PSO based + SVM | Proposal of model independent of domain proving efficient results than existing methods and models. | 84.52 | [36] |
| 4 | Cascaded model having 3 stages is | Four datasets from Amazon are used:<br>-Books | Hybrid Model (PSDEE) | A hybrid model and a polarity shift elimination | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | proposed to deal with polarity shift problem for the task of sentiment analysis. | -DVD<br>-electronic<br>-Kitchen | | method are proposed. | 87.1 | [37] |
| 5 | Sentiment analysis is performed using<br>• visual<br>• textual content. | Datasets using Flickr and Getty Images, visual sentiment analysis dataset. | Progressive CNN(PCNN) | Performance improvement in task of sentiment analysis for datasets of joint visual and textual information. | 78% | [38] |
| 6 | A novel approach for Sentiment analysis for visual analysis is proposed. | Twitter Dataset of images (1269) | CNN | This model outperforms other state-of-art models available in literature. | 86% | [39] |
| 7 | Images as well as text is analysed | Twitter Dataset is used having | combination of CNN and paragraph | Results and experimentation show | 77% | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | using a deep learning model. | images and text. | vector model are used. | improvement in accuracy of results when both image as well as text is used for sentiment analysis | | [40] |
| 8 | Framework to find the malware sellers using deep learning to find sentiment analysis of customer along with snowball sampling and classification. | Data is collected from Russian forums containing 69,385 threads and 485,019 posts (2004-2013) | Word2Vec + RNTN(Recursive Neural Tensor Network) | Results outperform-ed using this framework of deep learning based sentiment analysis. | 88.9% | [41] |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | Short text based sentiment analysis is perfrmed using deep CNN. | “SSTb (Stanford sentiment tree bank)and STS(stanford twitter sentiment corpora)” | Deep CNN model | Compared to traditional ML and Information Retrieval (IR), training time is less and classification performance is enhanced. | 86% | [42] |
| 10 | Different sentiments of people related to a brand are analysed using Artificial neural network | Twitter corpus (10,345,184 ) | DANN (Dynamic Artificial Neural Network. | Customers reviews about a specific brand are identified to improve business. | 80% | [43] |
| 11 | For the purpose of cross lingual sentiment classificatio | “Reviews form four languages having 2000 reviews(1000 | WSDNNs(weakly shared deep neural network) RNNLMs | “Improvement in task of cross lingual sentiment analysis than existing | 58.9 | [44] |

| | | | | | | |
|---|---|---|---|---|---|---|
| | -n deep belief networks is used. | positive and 1000 negative reviews." | | models in literature." | | |

The different approaches used in literature along with their accuracies, the model used by them are described below. Datasets used in the studies present in literature are also mentioned.

Table 2.2-2 Comparison

| Sr # | Model | Datasets | Accuracy | Reference |
|---|---|---|---|---|
| 1 | 3-layer NN (CNN-f+LSTM-f) | Pang & Lee movie reviews(MR), Stanford Twitter Sentiment , SST, IMDB large movie reviews, SenTube | 81.59 | [27] |
| | 3-layer NN (CNN-f+DVngram) | | 81.11 | |
| 2 | GRNN-SR | MR, SST, Amazon Product Reviews | 80.6 | [45] |
| 3 | CNN-GRU | YNM, VLSP | 85.8 | . [46] |
| 4 | Ensemble model | SemEval2013/14, Vader, STS-Gold, Sentiment140, IMDB, PL04 | 86.49% | [34] |

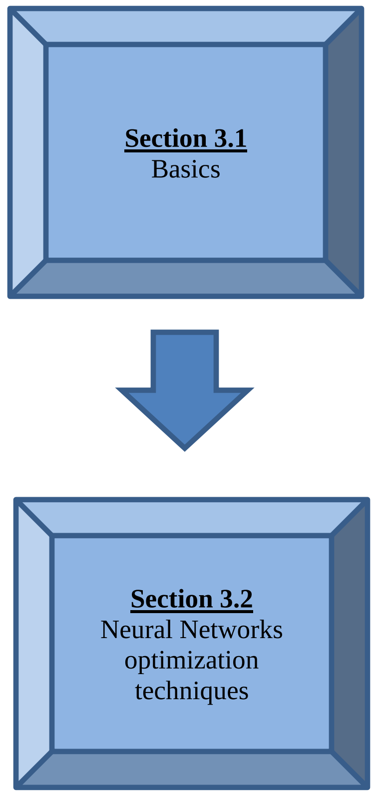
| | | | | |
|---|---|---|---|---|
| 5 | CNN-LSTM | Arabic Tweets | 65.05% | [47] |
| 6 | BiLSTM-CRF | Movie Review, Customer Review, SST-1, SST-2 | 88.3 | [48] |
| 7 | CNN | Sentiment 140 corpus | 74.9 | [49] |
| 8 | NBM, SVM, MEM, DT | Turkish and English Movie and Product review | 84.8 | [50] |
| 9 | SDRNN | Movie Review, MPQA, Customer Review | 86.7 | [51] |
| 10 | DCNN | Twitter data | 80.69 | [52] |
| 11 | NCNM | SST dataset and Cornell MRD dataset | 83.9 | [53] |
| 12 | Sent_Comp (SSC) | Chinese Blog Review Dataset | 88.95 | [54] |
| 13 | Deep belief nets (DBN) | Twitter database, Sina microblog (in Chinese), COAE2014 (in Chinese) | 76% | [55] |

| | | | | |
|---|---|---|---|---|
| 14 | RNN | Movie review, MPQA Opinion, Customer Review | 86.7% | [56] |
| 15 | CNN | Google Embeddings, Amazon Embeddings, SemEval 2014 dataset | 87.17% | [57] |
| 16 | Naïve Bayes, SVM | Textblob, SentiWordNet and Word Sense Disambiguation (WSD) | 79% | [58] |
| 17 | RNTN | Sentiment Treebank | 80.7 | [59] |
| 18 | LSTM-CNN | FiQA 2018 | 69% | [60] |
| 19 | CNN-LSTM | Twitter data | 88% | [61] |

# Chapter 3

# Neural Networks

# Chapter 3: Neural Network



# CHAPTER 3: NEURAL NETWORK

This chapter explains different concepts and algorithms used for sentiment analysis. Neural networks details and concepts are explained along with basic module of it. Deep neural networks details are also mentioned along with different regularization and training techniques. Natural language has different concepts including sentiment classification and word embeddings. These concepts are also explained. For evaluation of different model different criteria are used like accuracy.

## 3.1 Basics

This section explains about a neural network, its architecture and different models of neural networks used in this research.

Artificial neural network has different layers and the layer is also comprised of single neuron. The single neuron is denoted by the following equation:

$$y_j = f\left(b + \sum_{i=1}^{n} x_i W_j\right)$$

In the above equation X= [X1, X2, X3……… Xn] and W= [W1, W2, W3……… Wn] are vectors representing input and weights respectively. Different weights are assigned to input and the result of multiplication of input and weight is temporary. This temporary result having input and some weight value is added to the bias value which is denoted by b. n shows number of inputs and j specifies the jth neuron in the layer. The equation is to calculate output value of input having specific neuron (jth neuron) [62].

Activation function is also a part of deep neural network and take input in numerical form. After performing some mathematical calculations, it gave the final output. Different activation functions exist for the neuron like sigmoid, Tanh and ReLu.

Sigmoid function has the range of 0 and 1 for input. This activation function outputs 1 and 0. For the negative value it output 0 while for a positive value it outputs 1. In our proposed framework we have to predict the labels for sentiments present in text. And this function is also used for better performance. Tanh is also an activation function of nonlinear type having its range for output as -1 and 1 [62].

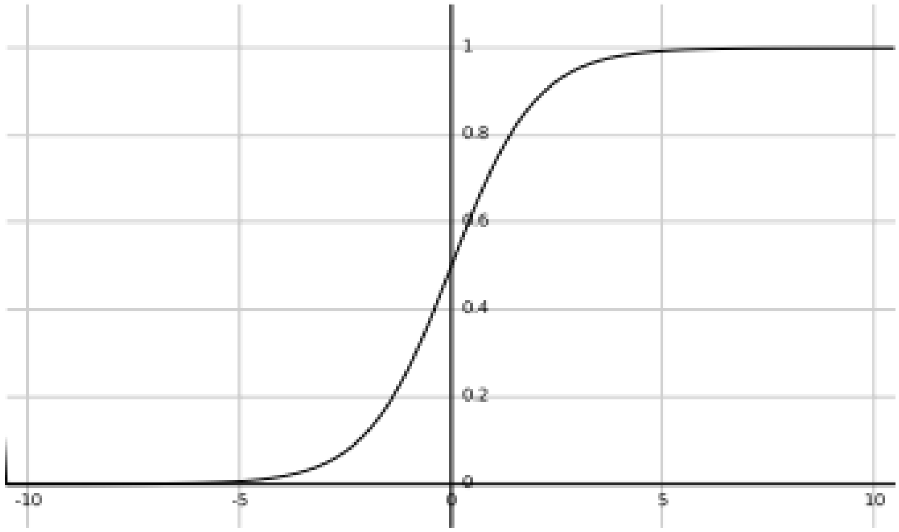



Figure 3.1-1 Activation function of sigmoid

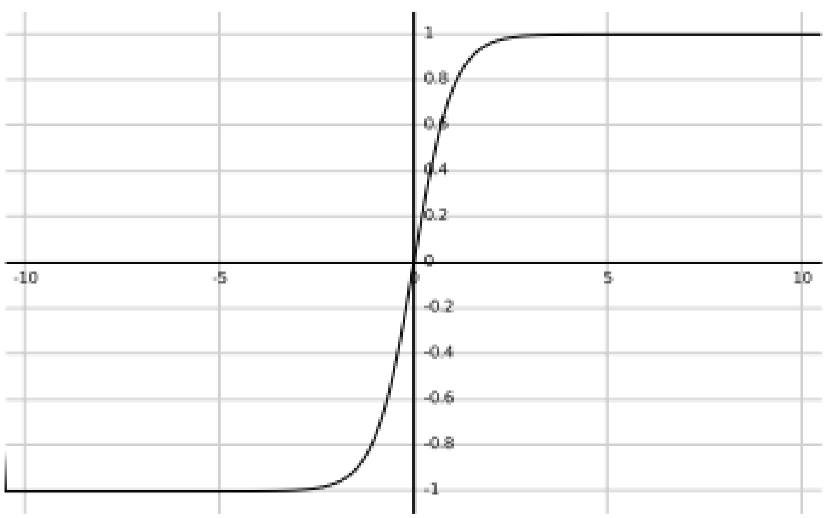



Figure 3.1-2 Activation function of Tanh

There also exist some linear activation functions like ReLu. They differ from other two activation functions mentioned before. The constant gradient of ReLu is the reason that it does not saturates like sigmoid and tanh functions. ReLu has an advance version that results in more fast training of the input data than simple ReLu. This advance and more improved ReLu is called as leaky ReLu function [63].

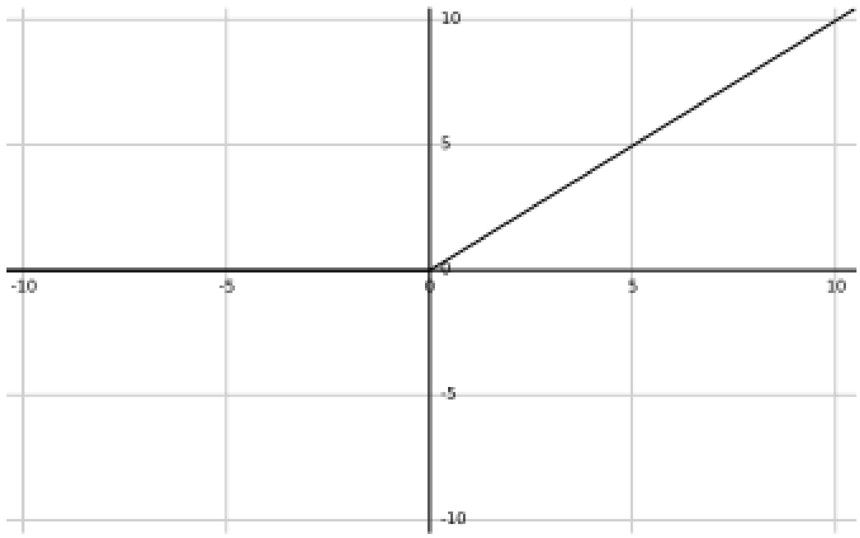



Figure 3.1-3 Activation function of ReLU

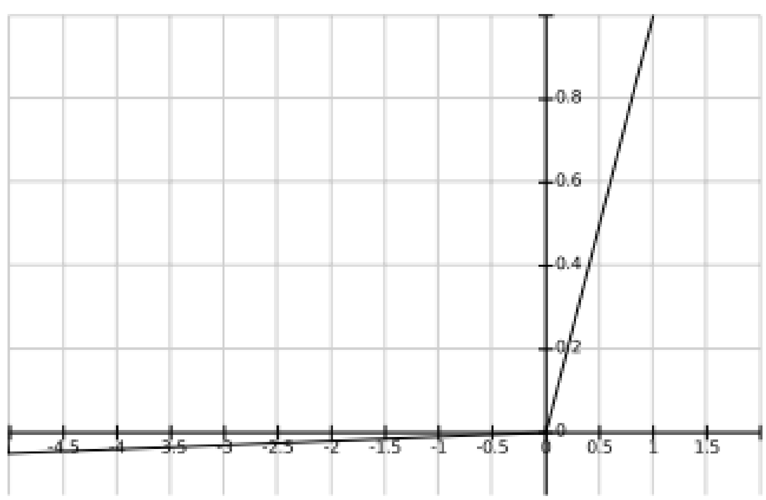



Figure 3.1-4 Activation function of Leaky ReLU

The basic layers that comprise a neural network includes an input, hidden and an output layer. The hidden layers can vary in different architectures of NN models and this variation of hidden layer number results in deep neural network. Figure 3.1-5 shows architecture of MLP having three layers [63]:

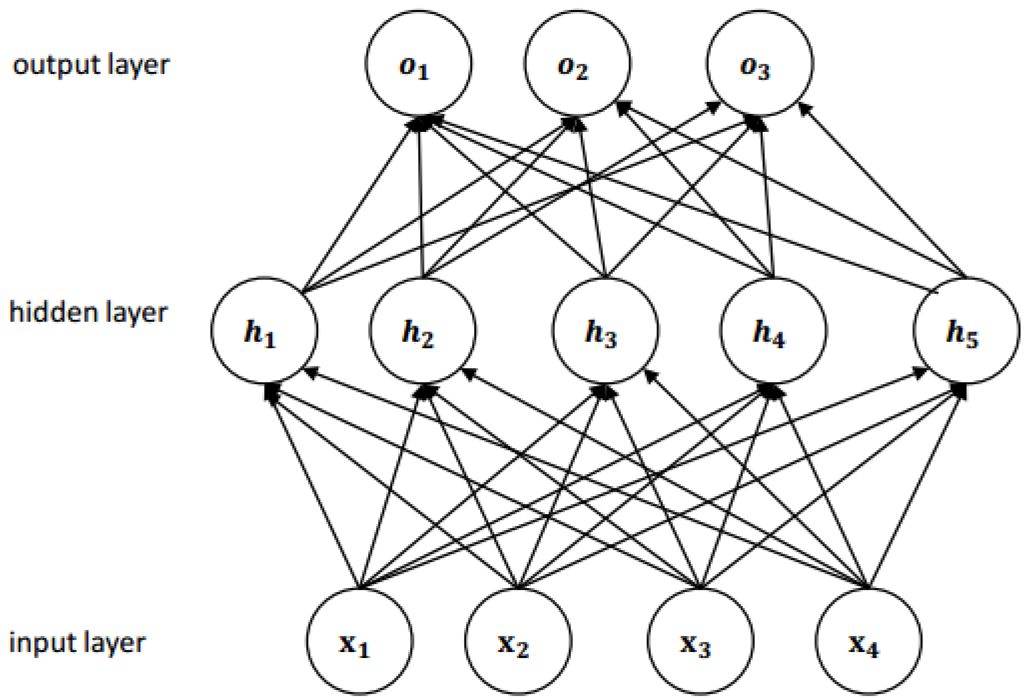



Figure 3.1-5 MLP Typical Architecture

There exist different deep neural networks that are also used in our proposed hybrid model. These models include RNN, LSTM, Bidirectional LSTM, GRU and CNN [64]. Each of this model has its own pros and cons thus giving different results for same datasets or inputs given to them. These models can have different architecture as needed to improve the accuracy.

RNN is a deep neural network. Sequence data (text data) is processed by this network.

These different types of RNNs can be seen in different frameworks. In our framework prediction is made with the use of many-to-one type, in this type the output value is only one value as predicted label in SA while every word is given as input to the model.

LSTM is an advance form of RNN that eliminates the long-distance problem i-e if the gap between two input words become large in RNN it becomes difficult to capture contextual and semantic meaning of it. BiLSTM improves the result for SA task as it captures contextual meaning using previous as well as the words appearing after the word of interest. Using both previous and forward direction the network finds the semantics and contextual meaning of the word of interest [65].

CNN or convolution neural network comprises of different layers like convolution layer, pooling layer. The convolution layer basically reduces the overall number of parameters by sharing parameters [66]. In our framework pre-trained word-embedding is used as first layer of model. For instance, for MR data this layer has initially number of parameters equals to 5718400 and after applying the convolution layer and using parameter sharing the number of parameters is reduced to 64128.

The next layer of CNN is pooling layer. It can perform pooling by either max pooling or by average pooling. Max pooling takes the maximum value out of the feature map while average pooling takes out features in a smooth way by taking all values in feature map and taking average of them [67].

## 3.2 Neural Network Optimization Techniques

Optimizing neural network model is an important factor that helps in giving better results and thus better accuracy value by reducing loss value. Loss function also has impact on model performance. The less the loss value the more is the accuracy of the model. Objective or loss function is basically an optimization function for NN model.

Training of a neural network results in also done to reduce the loss function value. Parameter tuning of model also result in loss value reduction.

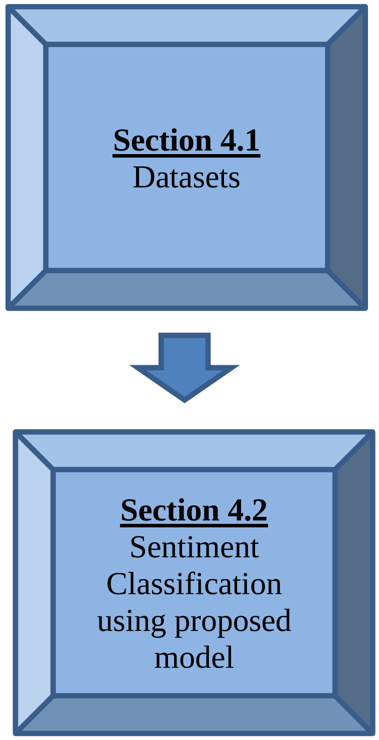
Adam optimizer is used by us in proposed framework to optimize. The training phase is the crucial phase and depending on the data and data size, training may take from several hours to a day or sometimes even more. For each parameter it adjusts the learning rate [68].

Next in neural network there are also regularization techniques to regulate the model problems. Overfitting of data is a problem that mostly occur in a model with increasing number of layers in a network. The model is well trained on training data but the general patterns are not learned so the training accuracy is good than the testing accuracy. The model does not perform well on test data. For regularization dropout layer is used in our model [69].

# Chapter 4

# Proposed Methodology

# Chapter 4: Proposed Methodology



# CHAPTER 4: PROPOSED METHODOLOGY

This chapter explains about the proposed methodology, the datasets used and the pre-processing steps involved. The different steps involved for pre-processing are explained briefly.

## 4.1 Datasets

For testing our model (RNN-LSTM-BiLSTM-GRU-CNN) different datasets specific for sentiment analysis tasks are used which are also available publicly. Following datasets are used:

- MR Dataset
- SST-1
- SST-2
- IMDB

Each dataset has its own characteristics. These are all labelled datasets. Following are the important details of each dataset:

### 4.1.1 MR Dataset:

This Movie review sentence polarity dataset is about different movies reviews and is of v1.0. It includes positive and negative reviews. Negative snippets are collected using Rotten Tomatoes by researchers and positive snippets from 345 websites pages. There are 5331 number of positive and negative reviews.

### 4.1.2 SST-1

Stanford Sentiment treebank dataset has a total of 11855 sentences. All of these sentences have originally five types of labels namely positive, very positive, neutral, negative, very negative. All the sentences are extracted from files obtained from the original pool of Rotten Tomatoes page. For experimental purposes positive and very positive both are denoted as positive label and negative and strongly negative are denoted as negative label while Neutral comments are dropped.

### 4.1.3 SST-2

This dataset has total of 9645 sentences having binary labels. For experimental purposes these numerical labels are assigned with positive and negative label. The neutral comments are already removed for experimental purposes.

### 4.1.4 IMDB

This dataset has a total of 50k entries of movie reviews of customers and is available for the purpose of sentiment analysis task. 25k data is training data and the other half is meant for testing purposes. Ratings are given on the scale of 1-10 and two classes namely positive and negative are present in it. No neutral class is present. Table 4.1-1 shows dataset description:

Table 4.1-1 Datasets Description

| **Dataset** | **Labels** | **No. of Reviews** | **Classes** | **Source Reference** |
|---|---|---|---|---|
| MR | Positive, negative | Positive=5331<br>Negative=5331<br>Total=10662 | 2 | [70] |
| SST-1 | positive, very positive, neutral, negative, very negative | 11855 | 5 | [71] |
| SST-2 | Positive, negative | 9654 | 2 | [71] |

| | | | | |
|---|---|---|---|---|
| IMDB Movie Review | Positive, negative | 50000<br>Positive= 25000<br>Negative=25000 | 2 | [72] |

## 4.2 Sentiment Classification using proposed model:

The proposed hybrid model having five different models is than applied to all the datasets and prediction is made based on the obtained results. The obtained results are than compared to state-of-the-art results.

Pre-trained word embedding helps in giving better results by improving accuracy. The pre trained embedding used in our proposed framework is Glove [73]. It is trained on large corpus that helps in finding syntactic and semantic meaning of the word. This unsupervised pre trained word embedding method is useful specifically when the data available for training is small. Figure 4.2-1 shows the overall framework for our model:

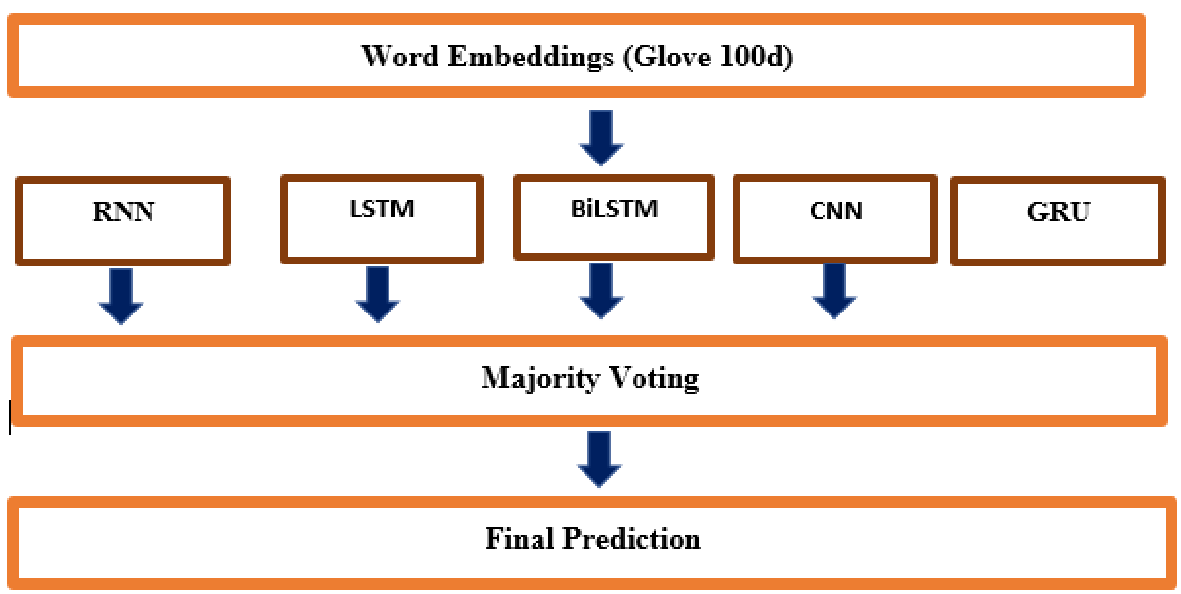



Figure 4.2-1 Proposed Framework

### 4.2.1 Datasets

Datasets used for proposed framework experimentation are IMDB, MR, SST-1 and SST-2. These datasets have reviews of movies available online. The datasets selected are available online

### 4.2.2 Pre-Processing

The next step involves conversion of un-structured data to a processed and well-structured form. Stopword removal, stemming and lemmatization etc. is also performed in this step.

### 4.2.3 Word Embedding

The pre-trained embeddings are also used. The embedding used in proposed framework is Glove 100d. It is used as input to the proposed hybrid model.

### 4.2.4 Classification

Using proposed hybrid framework classification is performed. Majority voting helps to improve the final result.

### 4.2.5 Evaluation of Result

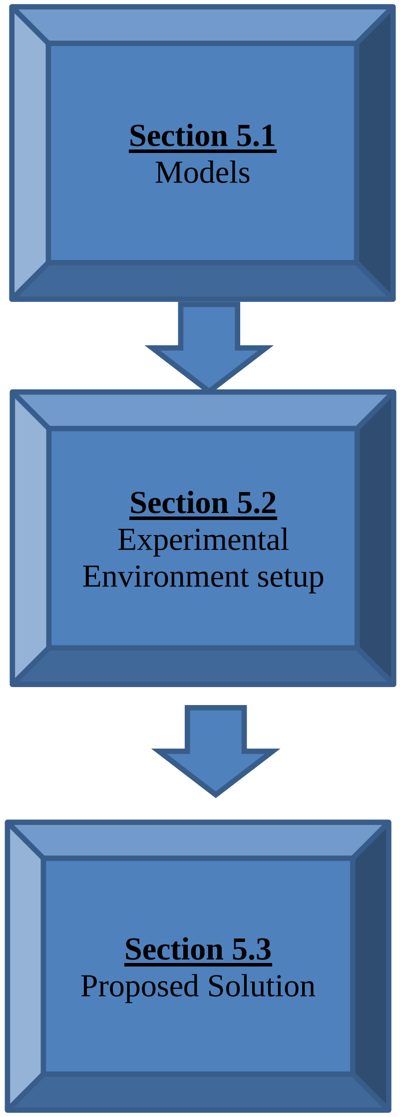
The results obtained are evaluated by using accuracy metrics. Improvement in accuracy is observed using proposed methodology.

.

# Chapter 5

# Implementation

# Chapter 5: Implementation



# CHAPTER 5: IMPLEMENTATION

This chapter explains about implementation of the proposed model and the different deep learning models involved in making hybrid. Word embeddings used are also mentioned and explained.

## 5.1 Models

Different models are used to generate the final result. Our proposed model is a combination of five models. Each model is individually trained and tested firstly. All models do not produce same results. But our proposed model is more reliable as it is generating final output by involving all the individual model outputs. So that the output that gets majority votes is the final output thus results in reliability of results.

Each individual model has its own characteristics, each of the models used are discussed below:

### 5.1.1 RNN

Natural language use Recurrent neural networks (RNN) as popular neural network technique for sentiment analysis task. RNN can utilize information from previous time steps due to its recurrent nature. RNN exist in different forms with advancement in its structure like LSTM and GRU. RNNs are widely used for text classification tasks because they can be used for text for variable length which ultimately helps in analysing reviews of variable lengths. RNNs are basically artificial NN having directed cycles. In the first step the input is passed through weight matrices U and W respectively than the previous time step hidden state and current input is fed as an input to next step. This version of RNN is called as Vanilla RNN [74].

### 5.1.2 LSTM

LSTM (Long short-term Memory) are advance form of recurrent neural networks for remembering information for short as well as long time period. This network is efficient in a way as it uses various gates to regulate the propagation of activations in the network and efficiently learns when to save information or to ignore the past state.

LSTM (Long short-term Memory) are advance form of recurrent neural networks for remembering information for short as well as long time period. This network is efficient in a way as it uses various gates to regulate the propagation of activations in the network and efficiently learns when to save information or to ignore the past state [74].

### 5.1.3 CNN

Classification and feature extraction are performed using deep neural network specifically CNN. Sequence of words from text or document is extracted and each sequence is represented using one-hot encoding into vector form. Each word is represented by a vector and thus forming a sequence of dense vector which is than fed as input to different models of Neural Networks. These sequences are than processed further for better prediction accuracy using different layers of NN model.

Hyper-parameter tuning has a major impact on increasing the accuracy of the model. Our model also has stacked Convolution and Max-pooling layer and used Rectified linear unit (ReLu). The standard layer structure of CNN is shown below having an input layer, convolution layer and a pooling layer.

Max pooling is used as a pooling layer and it helps to extract most important and relevant features. Fixed size output is also a result of this layer thus reducing dimensionality of output produced [39].

### 5.1.4 GRU

It is another deep learning model that is used for different tasks including sentiment analysis. It is an advance version of other deep learning model especially LSTM. It has a gate rather than memory cell of a LSTM. The update gate controls the updates of the other gate called as hidden gate. The control of influence of previous and current hidden state on each other is also control by the reset gate [74].

## 5.2 Experimental Environment setup

Python is a high level, object-oriented scripting language. Compared to other languages it For our model construction python libraries that are used are mentioned below [74]:

### 5.2.1 Pandas

Pandas is a python package and a tool to analyze data. It is an open source library with high-performance. It is easy to use data analysis tool.

### 5.2.2 NLTK

To work with human language, Natural language tool kit (NLTK) act as a platform to build programs. It provides libraries for text processing tasks like tokenization, stemming, parsing and tagging. For dealing with natural language it is a widely used library.

### 5.2.3 scikit-learn

It is a tool for data mining and to analyze data. It is used for different purposes like classification, clustering and Regression etc.

### 5.2.4 keras

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

### 5.2.5 Matplotlib:

It is a library of python. It is used to make graphs, histograms, pie charts, tables, scatter plots and bar charts etc. Matplotlib.pyplot is used to make metrics.

## 5.3 Proposed Solution

### 5.3.1 Input data

Three datasets are used as input to the proposed model namely SST-1, SST-2, Movie Review and MR data. The datasets are in the form of csv files. The data that is used is reviews (Text) against rating. Sampling is done to make data balanced. Varying length size is handled using padding. Padding is done using zeros. The max length set is 200. The ratings that are neutrals i-e 3 are firstly removed as they do not carry any opinion either positive or negative but both. The purpose of this research is to help business analyst in improving their business in the light of opinions given by people. Negative sentiments of people are mostly suggestion indirectly and also show demands of customers. By working on these aspects that are lacking, a handsome profit can be made.

While positive reviews actually show the strengths and positive aspects in any business that must be maintained to generate good revenue.

For our methodology the rating below 3 are categorized under negative Sentiment while above are categorized as positive sentiment. Some Sample reviews with the rating from MR dataset are shown in table 5.2-1:

Table 5.3-1 Sample Reviews (MR Dataset)

| **Review** | **Polarity/Score** |
|---|---|
| first think another Disney movie, might good, it's kids movie. watch it, can't help enjoy it. ages love movie. first saw movie 10 8 years later still love it! Danny Glover superb could play part better. Christopher Lloyd hilarious perfect part. Tony Danza believable Mel Clark. can't help, enjoy movie! give 10/10! | (positive) 1 |
| Shakespeare fan, appreciate Ken Branagh done bring Shakespeare back new generation viewers. However, movie falls short conveying overall intentions play ridiculous musical sequences. Add Alicia Silverstone's stumbling dialogue (reminiscent Keanu Reeves Much Ado Nothing) poorly cast roles, equals excruciating endurance viewing. | (Negative)0 |

### 5.3.2 Pre-processing

Data available online is mostly in un processed form. Pre-processing is used to make this data useful and to improve the accuracy and classification task. Extraction of features from reviews is also done by processing the available data by using following steps:

#### 5.3.2.1 Tokenization

Natural language toolkit (NLTK) is an open source library of python. It provides help in processing of data. It has different tasks to implement including tokenization i-e converting long sequences of text into tokens. Tokenization results in separating each word, number, stopword, phrases etc in the form of tokens [74].

Depending on the number of words tokens can be unigram, bigram, trigram etc. For instance, token having only one word is unigram, having two words are bigrams and having three words are called as trigram.

#### 5.3.2.2 Stop word Removal

Stop words are the word which make a sense of text but when tokens are made, these stop word tokens are of no use. So, these words are programmed to be ignored and they are the most common words like is, am, are etc. To make input data understandable by the computer pre-processing is done. NLTK library is used to remove stop words from the data.

#### 5.3.2.3 Stemming:

The next step is stemming of words. The input data words are stemmed to their root words. The Suffix of words is mostly removed for stemming purposes. Snowball stemmer is used in our model.

#### 5.3.2.4 Convert to Lower Case:

Next all the reviews are converted to small case. For instance, if a word is written as ‘review’ as well as ‘Review’, it will be treated as only one word ‘review’.

#### 5.3.2.5 Removing HTML tags

Using regex (re) HTML tags are removed from the input text. Everything inside < > is removed.

Punctuation marks are used to make the sense of sentences. Some examples of punctuation marks are '[. |, |) | (|\|/]' For making data useful these punctuation marks are removed using NLTK toolkit and regex.

### 5.3.3 Feature extraction

For applying Deep neural networks on sentiment classification there are three general steps. Firstly, Pre trained word embedding are applied. Glove is used in as a pre trained word embedding. Glove is an unsupervised learning algorithm for learning word vectors. It is trained using Wikipedia and Gigaword. It is of different dimensions including 50d, 100d and 300d. For our model glove of 100d and 300d are used for experimental purpose. In the next step features are extracted from glove word embedding. Against each word a vector is generated. In the third and the last step classification is performed for prediction purposes. The specification of glove embedding is given below in table 5.2-2:

Table 5.3-2 Glove Specifications

| **Model** | **Vocabulary** | **Dimension** | **Corpus** | **Tokens** |
|---|---|---|---|---|
| Glove | 400k | 100 | Wikipedia, Gigaword | 6 Billion |

The following flow chart shows complete methodology of proposed framework. All the steps are mentioned below in fig. 5.2-1:

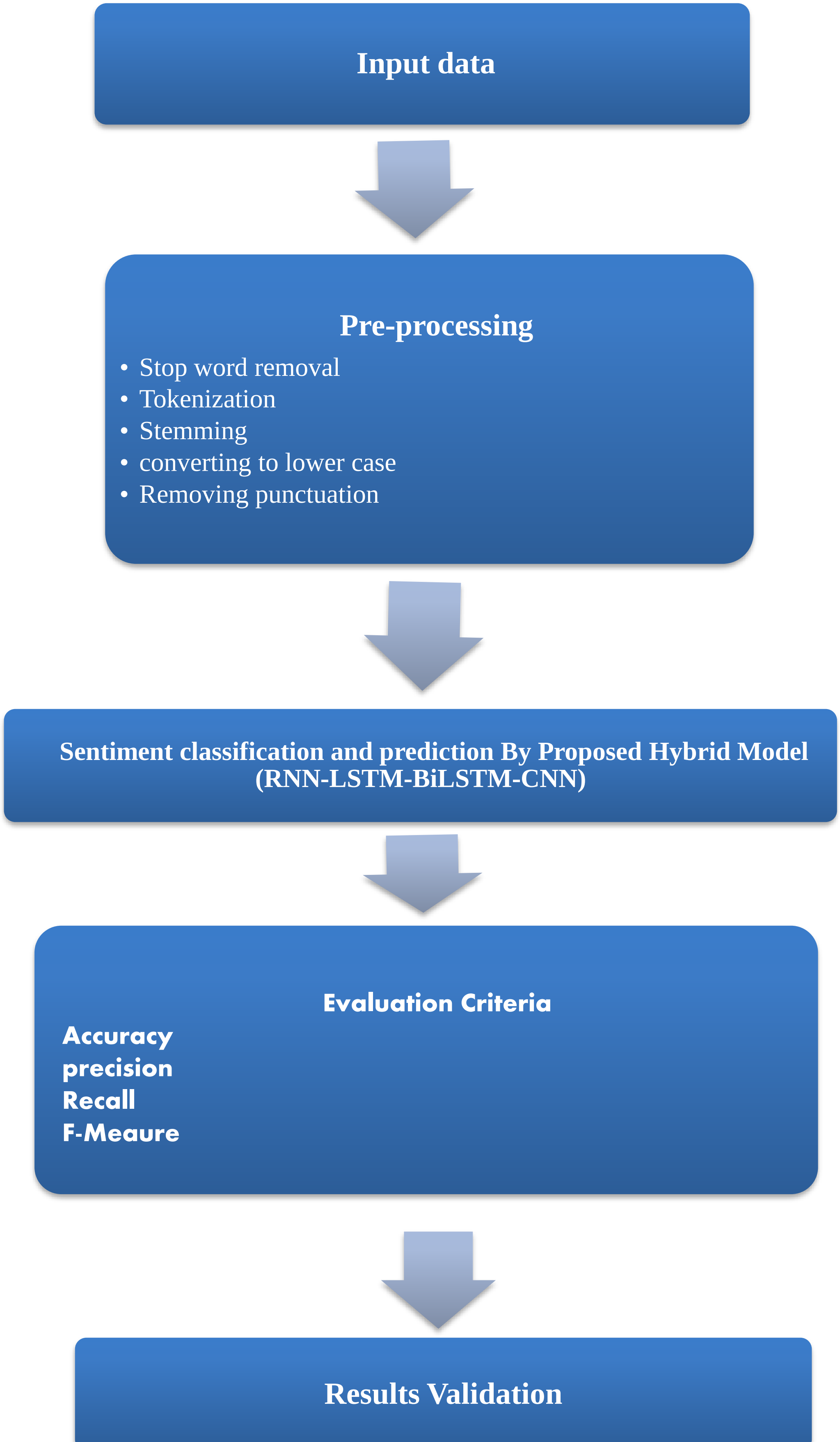

Input data
Pre-processing
• Stop word removal
• Tokenization
• Stemming
• converting to lower case
• Removing punctuation
Sentiment classification and prediction By Proposed Hybrid Model (RNN-LSTM-BiLSTM-CNN)
Evaluation Criteria
Accuracy
precision
Recall
F-Meaure
Results Validation

Figure 5.3-1 Proposed Framework

### 5.3.4 Word Embedding

ebe used to calculate the similarity between words. Cosine similarity is an example of a distance metrics. For generating word vectors two types of methods exist namely Statistical methods and Language models.

Language models considers the semantics and context information while generating word vectors. It includes Neural network language model (NNLM), word2vec, Glove, Fastex and Elmo. For our research Glove is used. Furthermore, statistical models include Bag of words, Tf-Idf and Singular value decomposition SVD. Figure 5.2-2 shows a review of word embedding techniques [76]:

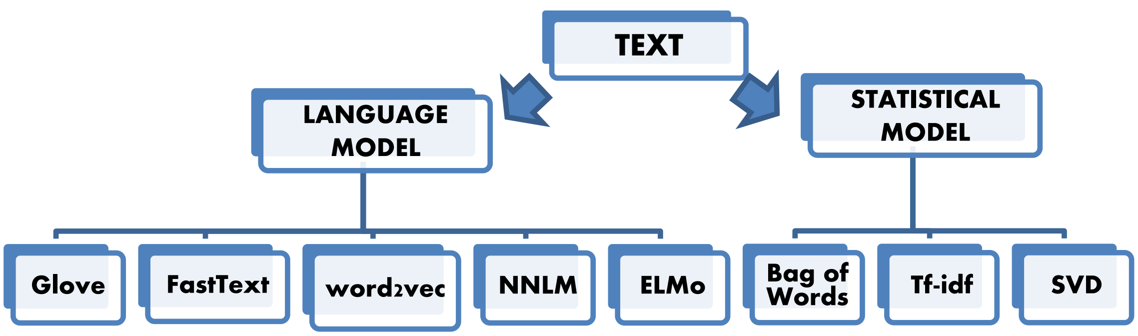



Figure 5.3-2 Word Embedding Techniques

Word2vec has two variant methods namely continuous Bag of words (CBOW) and skip gram. In skip gram method the target word can be found by using the words near to it, but in it not all the words are considered but a random and near word is chosen. While in CBOW the target word can be found using surrounding words. For example " I am purchasing these fine cotton clothes" is a complete sentence and the word that is targeted is "fine". CBOW model predicts the target word using all surrounding words like "purchasing", "cotton", "clothes". The sum of all the surrounding words vectors is used to predict the target word. While skip gram predicts using random word. Skip gram is considered to be more accurate than CBOW model for prediction [77].

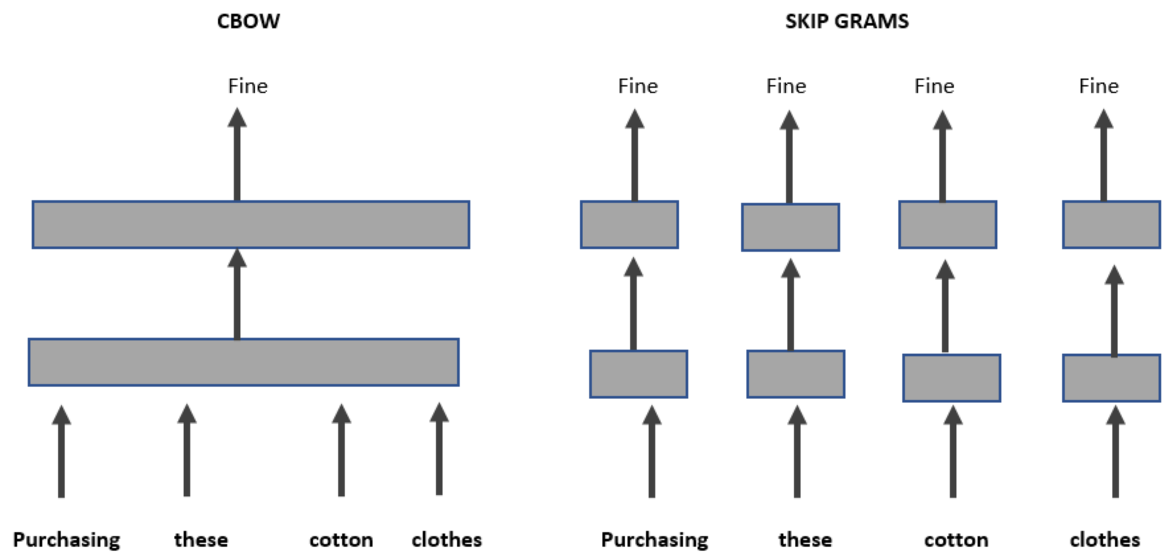



Figure 5.3-3 Word2vec Methods

Our proposed model hyper parameters are shown in table 5.2-3:

Table 5.3-3 Model HyperParameters

| **Model hyper parameters** | **Value** |
|---|---|
| Word embedding dimensions | 100 |
| Dropout rate | 0.25 |
| Epoch size | 10 |
| Batch size | 128 |

For the purpose of regularization dropout rate of 0.25 is added using dropout method. It helps deep neural network to regularize.

All the five models used and their trainable parameters are mentioned in table 5.2-4:

Table 5.3-4 Models Parameters

| **Sr. no.** | **Models** | **No. of parameters (SST-1)** | **No. of parameters (SST-2)** | **No. of parameters (MR)** |
|---|---|---|---|---|
| 1 | RNN | 9,445,902 | 9,294,702 | 13,939,602 |
| 2 | LSTM | 1,333,402 | 1,333,402 | 1,973,402 |
| 3 | BiLSTM | 1,386,802 | 1,386,802 | 2,026,802 |
| 4 | GRU | 1,320,102 | 1,320,102 | 1,960,102 |
| 5 | CNN | 1,290,140 | 1,138,940 | 5,783,840 |

The two types of parameters are [78]:

- **Trainable parameters**

These are the parameters that can be optimized during training of data. The weights of neural networks can be updated by algorithms like back propagation. For instance, the weights assigned to input can be of variable size and the weights can be estimated as for different weights value the cost function also get affected. So, weights are basically variable can be represented by w and the input matrix can be represented as x. the final function is:

$$Y = wx$$

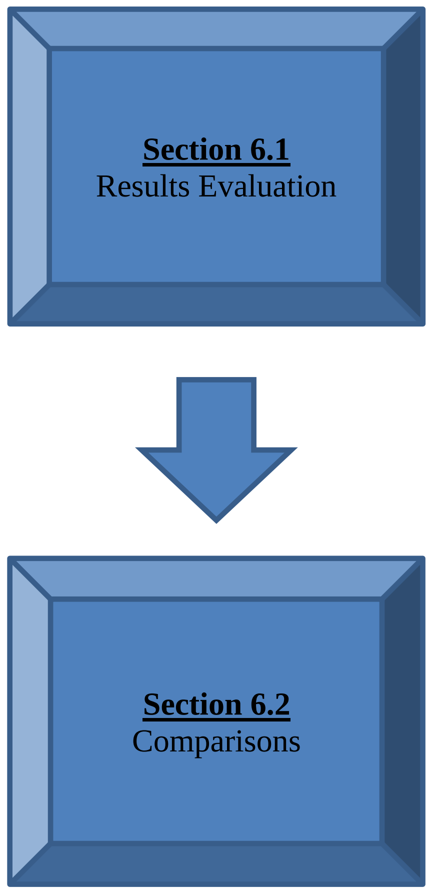
- **Non-trainable Parameters**

These types of parameters can be made trainable from non-trainable. The nodes of each layer and the number of hidden units are example of it. These types of parameters can be changed at input time but cannot be modified and updated during training.

# Chapter 6

# Results and Discussions

# Chapter 6: Results and Discussions



# CHAPTER 6: RESULTS AND DISCUSSIONS

## 6.1 Results Evaluation

Results are evaluated on our datasets including SST-1, SST-2, and MR datasets. The results obtained shows that our proposed hybrid model outperforms existing models for sentiment analysis and prediction of sentiments. Users leave a lot of free-text reviews that's why there is a need to predict the score or rating against each review accurately. This is also useful because it helps in business analysts to have a quick review of their customers opinions.

It is also observed that use of pre trained word embeddings for word vectors initializing has a positive effect on the prediction. We have used Glove in our experimentation and helps in generating better results.

The results of our proposed model along with the model to be compare on same dataset as our model are shown in the table 6.1-1. Results shows that our proposed hybrid model is the most accurate.

Table 6.1-1 Proposed Model Results

| Datasets | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| MR | 86.6 | 87.02 | 86.64 | 86.59 |
| SST-I | 88.6 | 83.64 | 88.63 | 85.89 |
| IMDB | 88.9 | 86.5 | 88.94 | 87.4 |
| SST-II | 90.6 | 89.4 | 90.6 | 87.2 |

All four datasets are run on the proposed hybrid model. The model is evaluated using different evaluation measures including Precision, Recall, F1 and Accuracy. The result shows that proposed model outperforms in terms of accuracy. SST-II dataset reached highest accuracy of 89.4%. Figure 6.1-1 demonstrates accuracies achieved using proposed model.

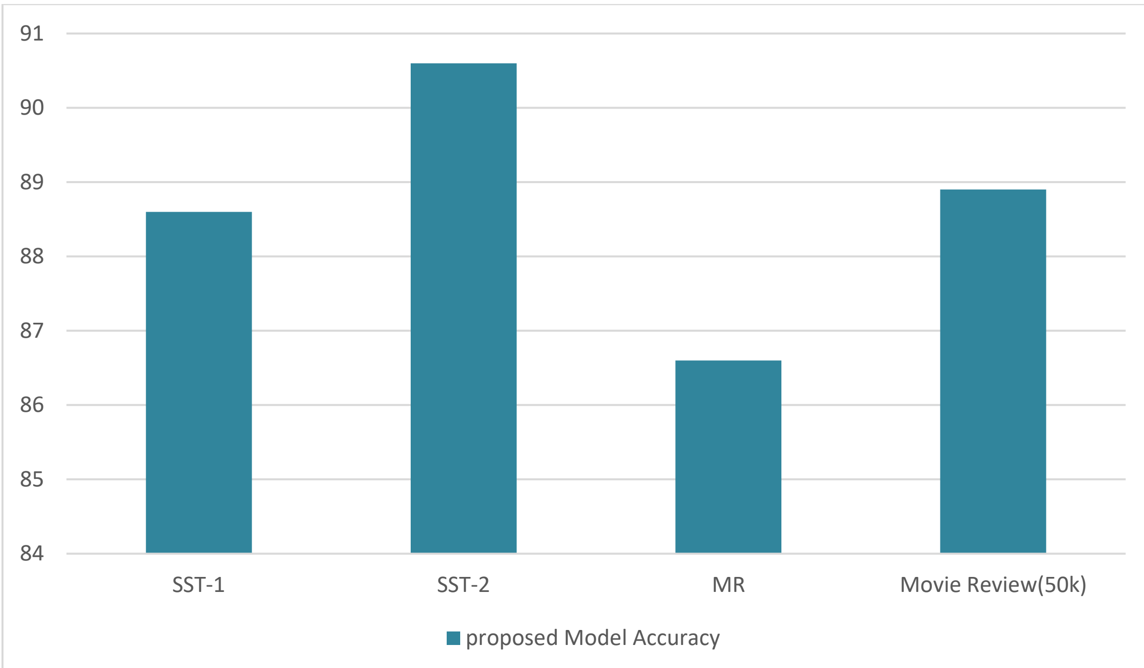



Figure 6.1-1 Proposed Model Accuracy

Our results are to be compared to the latest model of RNN-LSTM []. The datasets used are same for both models i-e the proposed model and the previous model proposed recently. Our model outperforms in accuracy when used on MR, SST-1, SST-2 dataset. The increase of 4.5%, 18.3% and 0.1% accuracy are observed using our proposed hybrid model.

For generating final results, each model is run individually and prediction is made. Then using majority voting final result is generated. Based on a fixed criteria binary labels against each review is made. If the predicted result is less than 0.5, it is declared as negative and the final label assigned to it is 0. In a same way when the value of predicted label is greater than or equal to 0.5, than the final label assigned to it is 1 (positive). Figure 6.1-2 shows review at index 2 for each dataset and prediction made against it.

Table 6.1-2 Predicted Final Labels for MR Dataset Review at Index [2]

| Sr. no. | Review at index [2] of MR Dataset | Model | Actual Label | Predicted Label | Final Label |
|---|---|---|---|---|---|
| 1 | big fan Stephen King's work, film made even greater fan King. Pet Sematary Creed family. moved new house, seem happy. pet cemetery behind house. Creed's new neighbor Jud (played Fred Gwyne) explains burial ground behind pet cemetery. burial ground pure evil. Jud tells Louis Creed bury human (or kind pet) burial ground, would come back life. problem, come back, person, they're evil. Soon Jud explains everything Pet Sematary, everything starts go hell. wont explain anymore don't want give away main parts film. acting Pet Sematary pretty good, needed little bit work. story one main parts movie, mainly original | RNN | 1.0 | 1.0 | 1.0 |
| | | LSTM | 1.0 | 0.572 | 1.0 |
| | | BiLSTM | 1.0 | 0.999 | 1.0 |

| | | | | | |
|---|---|---|---|---|---|
| | gripping. film features lots make-up effects make movie way eerie, frightening. One basic reasons movie sent chills back, fact make-up effects. one character film truly freaky. character Zelda. particular character pops film three times precise. Zelda Rachel Creed's sister passed away years before, Rachel still haunted her. first time Zelda appears movie isn't generally scary isn't talking anything, second time worst, honest, second time scares living me. absolutely nothing wrong movie, almost perfect. Pet Sematary delivers great scares, pretty good acting, first rate plot, mesmerizing make-up. truly one favorite horror films time. 10 10. | GRU | 1.0 | 0.999 | 1.0 |
| | | CNN | 1.0 | 0.4 | 0 |

Table 6.1-3 shows review of SST-2 and SST-1 dataset at index 2. Against this review the predicted labels and the final labels thus generated by individual models of hybrid are shown. In the similar way other reviews labels are also predicted and then using majority voting final label against each review is assigned.

Table 6.1-3 Predicted Final Labels for SST-2 and SST-1 Dataset Review at Index [2]

| **Sr. no.** | **Review at index [2] of SST-2 Dataset** | **Model** | **Actual Label** | **Predicted Label** | **Final Label** |
|---|---|---|---|---|---|
| 1 | singer composer bryan adams contributes a slew of songs a few potential hits a few more simply intrusive to the story but the whole package certainly captures the intended er spirit of the piece. | RNN | 1.0 | 1.0 | 1.0 |
| | | LSTM | 1.0 | 0.292 | 0.0 |
| | | BiLSTM | 1.0 | 0.998 | 1.0 |
| | | GRU | 1.0 | 0.999 | 1.0 |
| | | CNN | 1.0 | 0.994 | 1.0 |
| **Sr. no.** | **Review at index [2] of SST-1 Dataset** | **Model** | **Actual Label** | **Predicted Label** | **Final Label** |
| | | RNN | 1.0 | 1.0 | 1.0 |
| | | LSTM | 1.0 | 0.292 | 0.0 |

| | | | | | |
|---|---|---|---|---|---|
| 1 | Effective but too-tepid biopic | BiLSTM | 1.0 | 0.998 | 1.0 |
| | | GRU | 1.0 | 0.999 | 1.0 |
| | | CNN | 1.0 | 0.994 | 1.0 |

## 6.2 Comparison

Table 6.2-1 shows models that are also run on same datasets to compare the results and for experimenting with different models possible:

Table 6.2-1 Comparison of proposed Model Results with other Possible Model Results

| **Model** | **Datasets** | | | |
|---|---|---|---|---|
| | **MR** | **SST-1** | **SST-2** | **IMDB** |
| RNN-LSTM-BiLSTM-GRU | 83.3 | 83.4 | 80.8 | 87.4 |
| RNN-LSTM-BiLSTM-CNN | 83.9 | 87.5 | 83.2 | 87.7 |
| RNN-LSTM-BiLSTM | 80.80 | 87.9 | 83.4 | 86.12 |
| RNN-BiLSTM | 82.1 | 82.9 | 81.6 | 86.19 |
| LSTM-CNN | 82.92 | 78.5 | 79.22 | 86.89 |
| BiLSTM-CNN | 82.56 | 81.5 | 80.8 | 86.4 |
| RNN-LSTM-BiLSTM-CNN-GRU(Proposed Model) | **86.6** | **88.63** | **90.6** | **88.9** |

Figure 6.2- 1 shows graph showing comparison of proposed methodology results with other models accuracy which is tested for predicting sentiment of reviews.

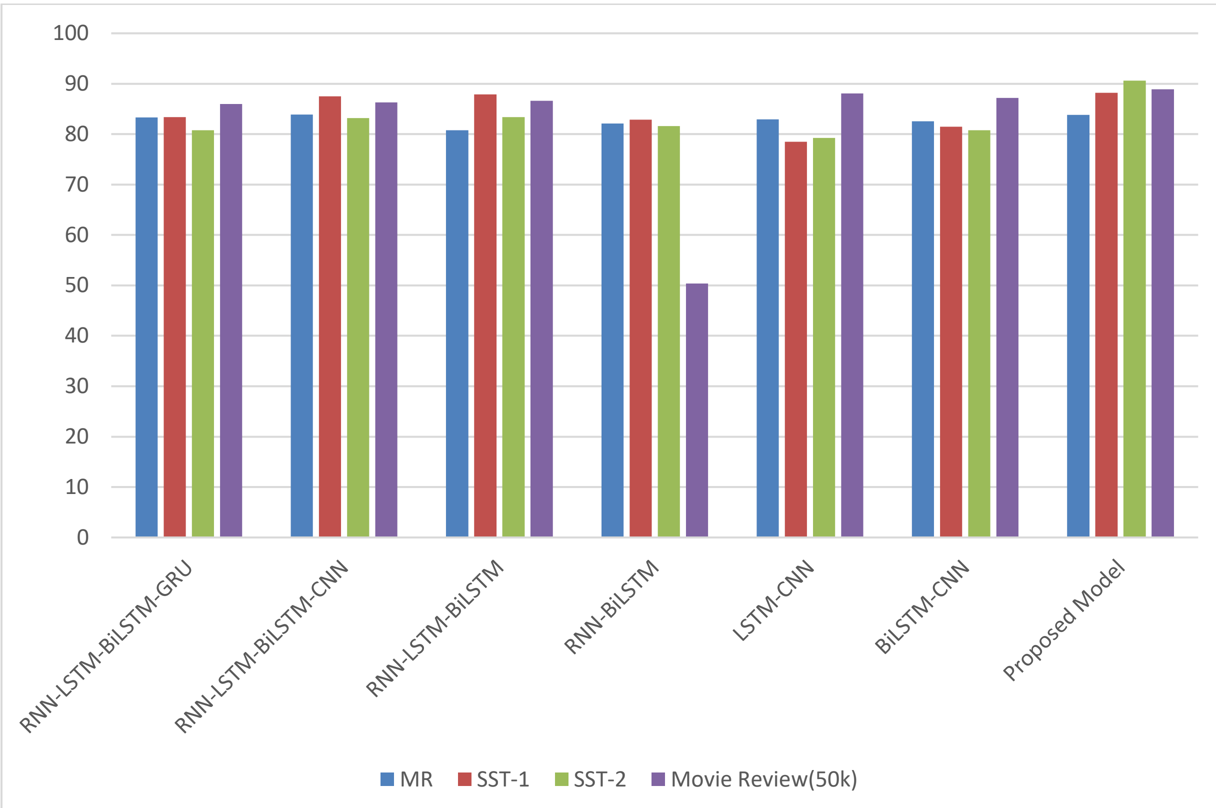



Figure 6.2-1 Comparison of Different Models Results

Table 6.2-2 shows detailed experimental results for MR dataset showing values for precision, Recall and F1

Table 6.2-2 Detailed Result For MR Dataset

| Model | Precision | Recall | F1 |
|---|---|---|---|
| RNN-LSTM-BiLSTM-GRU | 83.3 | 83.3 | 84.3 |
| RNN-LSTM-BiLSTM-CNN | 83.6 | 83.9 | 83.5 |
| RNN-LSTM-BiLSTM | 81.2 | 80.80 | 80.8 |
| RNN-BiLSTM | 83.1 | 82.1 | 81.9 |
| LSTM-CNN | 82.9 | 82.9 | 81.7 |
| BiLSTM-CNN | 82.5 | 82.56 | 82.0 |
| RNN-LSTM-BiLSTM-GRU | 84.0 | 83.3 | 83.1 |

Table 6.2-3 shows different models present in literature and their obtained accuracies. Different models use different sets of datasets. We have chosen mostly the models having same datasets as used in our model.

Table 6.2-3 Results of Different Models From Literature

| **Models** | **MR** | **SST-1** | **IMDB Movie Review** | **SST-2** | **Reference** |
|---|---|---|---|---|---|
| Ensemble Model | - | - | 86.4 | - | [34] |
| Deep CNN | - | 85.7 | - | - | [42] |
| GRNN-Sentiment Relation | 82.7 | 51.6 | - | 89.0 | [45] |
| RNNTN | - | 85.4 | - | 80.7 | [46] |
| Bi-LSTM-CRF-CNN | 82.3 | 48.5 | - | 88.3 | [48] |
| RNN | 78.0 | - | - | - | [51] |
| CNN-LSTM | 81.59 | - | 88.49 | - | [79] |
| | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| SD-RNN | 84.7 | - | - | - | [80] |
| MV-RNN | - | 44.4 | - | 82.9 | [81] |
| DCNN | - | 48.5 | - | 86.8 | [82] |
| RNN-LSTM | 82.1 | 70.3 | 82.1 | 89.3 | - |
| Proposed Approach | **86.6** | **88.6** | **88.9** | **90.6** | - |

There are different evaluation metrics used in literature like error rate for evaluation of models [83]. Some of the evaluation metrics are Accuracy, Precision, F-Measure and recall. Text classification can also include sentiment classification also known as polarity classification. And accuracy is the most important criteria to evaluate.

For evaluation purposes these evaluation metrics are used in our proposed methodology especially accuracy. Each of the evaluation metric has its own formula using which model can be evaluated. These formulas are explained below.

Confusion matrix for binary classification is shown in Fig 6.2-2:

| **Actual Label** | **Prediction** | |
|---|---|---|
| | True | False |
| True | TP | FN |
| False | FP | TN |

Figure 6.2-2 Confusion Matrix

A brief explanation of each term is following:

**True Positive (TP)**

TP is the number of labels that are actually true and are also predicted true.

**False Negative (FN)**

FN is the number of labels that are actually true but are predicted false.

**False Positive (FP)**

FP is the number of labels that are actually false but are predicted true.

**True Negative (TN)**

TN is the number of labels that are actually false and are also predicted false.

**Accuracy**

Accuracy is the number of correct predictions made divided by the total count of data. Accuracy is the most commonly used metric for evaluation purposes. The more number of correctly predicted values, results in improved accuracy.it can be shown by following equation:

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

**Precision**

The proportion of correct positive prediction in all of the positive prediction is known as precision.

It can be shown with the formula given below

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

It can be seen that precision is number of true positives divided by the sum of true positive and false positive.

**Recall**

The number of correct positive predictions divided by all of the positive cases numbers in test data is called as Recall.

$$Recall = \frac{TP}{TP+FN}$$

**F-Measure**

The weighted average of precision and recall is known as F-Measure. It results in accurately determine the model performance.

$$F\text{-}Measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

## 6.3 Discussion

Sentiment analysis is a vast field of research and different models and frameworks are proposed in literature for improving accuracy of correctly predicting polarity against reviews available online. Machine learning has another new area called deep learning and is being explored for the task of sentiment analysis. Deep learning approaches performs better perform better than tradition machine learning approaches and models exists in literature. Natural language processing (NLP) is an important task in analysing sentiments present in text. Deep learning techniques use in NLP results in better results in term of accuracy. Regardless of the use of deep learning approaches in sentiment analysis, a lot of improvements can still be made in this field and accuracy can also be improved. Existing models can be improved by improving architecture of the deep learning models and this motivation is the main reason to conduct this research and for proposing the hybrid framework.

This research study evaluates different models available in literature for sentiment analysis and to overcome the shortcomings in these models a hybrid model is proposed in this thesis. Our proposed model comprises of different deep learning models and overall accuracy is improved using hybrid having different models. The proposed framework provides better results when compared to other approaches exist in literature. Maximum accuracy achieved is of 90.6 for SST-2 dataset.

# Chapter 7

# Conclusion and Future work

# CHAPTER 7: CONCLUSION AND FUTURE WORK

## 7.1 Conclusion

For analysing and predicting sentiments from text-only reviews different models exists in literature. Some models perform better results than others. Predicting sentiments of people is an important task in many fields. Therefore, a more accurate model is proposed providing better results than existing models. Many models involving deep learning are already present in literature but they involve individual models mostly CNN. Performance of the existing models can be improved. Better models showing improved performance is still the need in the field of Sentiment Analysis. A novel hybrid model is proposed in this research study to predict sentiments against text-only reviews.

The proposed model is a hybrid of five different models including RNN-LSTM-BiLSTM-CNN-GRU. Extensive experimentation is done using different datasets and results are than compared with existing models results in literature. Five cross validation is also performed on the given datasets to generate valuable results. The results show that our model outperforms existing results I literature and provide more reliable results. The final results also show that hybrid model involving pre-trained word embedding provides better results. Different evaluation metrics are also used like accuracy, F-measure, Recall and precision to evaluate our results.

## 7.2 Future Work

The future work involves improving the performance of model for predicting the sentiments from text. Furthermore, new combination of deep learning models can also be experimented. Our proposed model can also be tested on other datasets. We have tested our model only on four datasets available online including IMDB, SST-1, SST-2 and MR dataset.

Furthermore, the proposed model is tested only on English language datasets, it can be tested on other languages datasets including Spanish, Chinese, Arabic etc.

| Sr. No | **Approach and Idea** | **Datasets** | **Model** | **Main Contribution** | **Accuracy** | **Ref.** |
|---|---|---|---|---|---|---|
| 1 | Different types of Ensemble using basic classifiers and deep learning models are proposed. | Seven public datasets are used including movie review (**IMDB**), STS-Gold, SemEval(2013) and (2014), PL04, Sentiment140 and PL04 | Ensemble Models using Bigrams | Improvement in Performance | 87.87 | Araque, Oscar, et al. "Enhancing deep learning sentiment analysis with ensemble techniques in social applications." Expert Systems with Applications 77 (2017): 236-246. |
| 2 | A neural network model that can capture behavioural information is used (CNN) | Two Datasets from twitter are used<br>- SemEval-1(2016)<br>-SemEval-2(2016) | CNN | Provides much better results than baseline method as Naïve Bayes and outperformed in capturing behavioural information than other models | 88 | Alharbi, Ahmed Sulaiman M., and Elise de Doncker. "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information." Cognitive Systems Research 54 (2019): 50-61. |
| 3 | Aspect based | Benchmark dataset of | PSO based + | Proposal of model | | Akhtar, Md Shad, et al. "Feature |

| | | | | | | |
|---|---|---|---|---|---|---|
| | sentiment analysis using cascaded framework of classifiers and feature selection is proposed. PSO is used for this purpose. | SemEval(2014) is used for two domains of: -Restaurant -Laptops | SVM | independent of domain proving efficient results than existing methods and models. | 84.52 | selection and ensemble construction: A two-step method for aspect based sentiment analysis." Knowledge-Based Systems 125 (2017): 116-135. |

| | Approach and Idea | Datasets | Model | Main Contribution | Accuracy | Ref. |
|---|---|---|---|---|---|---|
| 1 | Different types of Ensemble using basic classifiers and deep learning models are proposed. | Seven public datasets are used including movie review (**IMDB**), STS-Gold, SemEval(2013) and (2014), PL04, Sentiment140 and PL04 | Ensemble Models using Bigrams | Improvement in Performance | 87.87 | Araque, Oscar, et al. "Enhancing deep learning sentiment analysis with ensemble techniques in social applications." Expert Systems with Applications 77 (2017): 236-246. |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | A neural network model that can capture behavioural information is used (CNN) | Two Datasets from twitter are used<br>- SemEval-1(2016)<br>-SemEval-2(2016) | CNN | Provides much better results than baseline method as Naïve Bayes and outperformed in capturing behavioural information than other models | 88 | Alharbi, Ahmed Sulaiman M., and Elise de Doncker. "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information." Cognitive Systems Research 54 (2019): 50-61. |
| 3 | Aspect based sentiment analysis using cascaded framework of classifiers and feature selection is proposed. PSO is used for this purpose. | Benchmark dataset of SemEval(2014) is used for two domains of:<br>-Restaurant<br>-Laptops | PSO based + SVM | Proposal of model independent of domain proving efficient results than existing methods and models. | 84.52 | Akhtar, Md Shad, et al. "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis." Knowledge-Based Systems 125 (2017): 116-135. |
| 4 | Cascaded model having 3 stages is | Four datasets from Amazon are used:<br>-Books | Hybrid Model (PSDEE) | A hybrid model and a polarity shift elimination | | Xia, Rui, et al. "Polarity shift detection, elimination and ensemble: A three- |

| | | | | | | |
|---|---|---|---|---|---|---|
| | proposed to deal with polarity shift problem for the task of sentiment analysis. | -DVD<br>-electronic<br>-Kitchen | | method are proposed. | 87.1 | stage model for document-level sentiment analysis." Information Processing & Management 52.1 (2016): 36-45. |
| 5 | Sentiment analysis is performed using<br>• visual<br>• textual content. | Datasets using Flickr and Getty Images, visual sentiment analysis dataset. | Progressive CNN(PCNN) | Performance improvement in task of sentiment analysis for datasets of joint visual and textual information. | 78% | You, Quanzeng. "Sentiment and emotion analysis for social multimedia: Methodologies and applications." Proceedings of the 24th ACM international conference on Multimedia. ACM, 2016. |
| 6 | A novel approach for Sentiment analysis for visual analysis is proposed. | Twitter Dataset of images (1269) | CNN | This model outperforms other state-of-art models available in literature. | 86% | Islam, Jyoti, and Yanqing Zhang. "Visual sentiment analysis for social images using transfer learning approach." 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | Computing and Communications (SustainCom)(BD Cloud-SocialCom-SustainCom). IEEE, 2016. |
| 7 | Images as well as text is analysed using a deep learning model. | Twitter Dataset is used having images and text. | combination of CNN and paragraph vector model are used. | Results and experimentation show improvement in accuracy of results when both image as well as text is used for sentiment analysis | 77% | You, Quanzeng, et al. "Joint visual-textual sentiment analysis with deep neural networks." Proceedings of the 23rd ACM international conference on Multimedia. ACM, 2015. |
| 8 | Framework to find the malware sellers using deep learning to find sentiment analysis of customer along with snowball sampling | Data is collected from Russian forums containing 69,385 threads and 485,019 posts (2004-2013) | Word2Vec + RNTN(Recursive Neural Tensor Network) | Results outperform-ed using this framework of deep learning based sentiment analysis. | 88.9% | Li, Weifeng, and Hsinchun Chen. "Identifying top sellers in underground economy using deep learning-based sentiment analysis." 2014 IEEE Joint Intelligence and Security Informatics Conference. IEEE, 2014. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | and classification. | | | | | |
| 9 | Short text based sentiment analysis is perfrmed using deep CNN. | “SSTb (Stanford sentiment tree bank)and STS(stanford twitter sentiment corpora)” | Deep CNN model | Compared to traditional ML and Information Retrieval (IR), training time is less and classification performance is enhanced. | 86% | Dos Santos, Cicero, and Maira Gatti. "Deep convolutional neural networks for sentiment analysis of short texts." Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014. |
| 10 | Different sentiments of people related to a brand are analysed using Artificial neural network | Twitter corpus (10,345,184 ) | DANN (Dynamic Artificial Neural Network. | Customers reviews about a specific brand are identified to improve business. | 80% | . Ghiassi, Manoochehr, James Skinner, and David Zimbra. "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network." Expert Systems with applications 40.16 (2013): 6266-6282 |
| 11 | For the purpose of | “Reviews form four | WSDNNs(weakly shared | “Improvement in task of cross | 58.9 | Mikolov, Tomas, et al. "Efficient |

| | | | | | | |
|---|---|---|---|---|---|---|
| | cross lingual sentiment classificatio-n deep belief networks is used. | languages having 2000 reviews(1000 positive and 1000 negative reviews.” | deep neural network) RNNLMs | lingual sentiment analysis than existing models in literature.” | | estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013). |

## REFERENCES


[1] Song, Minchae, Hyunjung Park, and Kyung-shik Shin. "Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean." Information Processing & Management 56.3 (2019): 637-653.

[2] Abdi, Asad, et al. "Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion." Information Processing & Management 56.4 (2019): 1245-1259.

[3] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5.4 (2014): 1093-1113.

[4] Ramaswamy, Sridhar, and Natalie DeClerck. "Customer Perception Analysis Using Deep Learning and NLP." Procedia Computer Science 140 (2018): 170-178.

[5] Abid, Fazeel, et al. "Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter." Future Generation Computer Systems 95 (2019): 292-308.

[6] Oghina, Andrei, et al. "Predicting imdb movie ratings using social media." European Conference on Information Retrieval. Springer, Berlin, Heidelberg, 2012.

[7] "Ratings And Reviews For New Movies And TV Shows - Imdb". Imdb, 2019, https://www.imdb.com. Accessed 28 Nov 2019.

[8] "Tripadvisor: Read Reviews, Compare Prices & Book". Tripadvisor, 2019, https://www.TripAdvisor.com.

[9] "Holiday Lets, Homes, Experiences & Places - Airbnb". Airbnb, 2019, https://www.airbnb.com.

[10] Kavitha, G., B. Saveen, and Nomaan Imtiaz. "Discovering Public Opinions by Performing Sentimental Analysis on Real Time Twitter Data." 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET). IEEE, 2018.

[11] Salinca, Andreea. "Business reviews classification using sentiment analysis." 2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). IEEE, 2015.

[12] Goularas, Dionysis, and Sani Kamis. "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data." 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML). IEEE, 2019.

[13] Vo, Anh-Dung, Quang-Phuoc Nguyen, and Cheol-Young Ock. "Opinion–aspect relations in cognizing customer feelings via reviews." IEEE Access 6 (2018): 5415-5426.

[14] Idrus, Ali, and Herlambang Brawijaya. "Sentiment Analysis Of State Officials News On Online Media Based On Public Opinion Using Naive Bayes Classifier Algorithm And Particle Swarm Optimization." 2018 6th International Conference on Cyber and IT Service Management (CITSM). IEEE, 2018.

[15] Al-Moslmi, Tareq, et al. "Approaches to cross-domain sentiment analysis: A systematic literature review." IEEE Access 5 (2017): 16173-16192.

[16] Dragoni, Mauro, and Giulio Petrucci. "A neural word embeddings approach for multi-domain sentiment analysis." IEEE Transactions on Affective Computing 8.4 (2017): 457-470.

[17] Product Reviews - Phone, Computer, Electronics Reviews & More". CNET, 2019, http://reviews.cnet.com/.

[18] Balaji, Penubaka, O. Nagaraju, and D. Haritha. "Levels of sentiment analysis and its challenges: A literature review." 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC). IEEE, 2017.

[19] Agarwal, Basant, and Namita Mittal. "Machine learning approach for sentiment analysis." Prominent feature extraction for sentiment analysis. Springer, Cham, 2016. 21-45.

[20] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.

[21] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5.4 (2014): 1093-1113.

[22] Muhsina, N. K., and Nabizath Saleena. "Redundancy detection: A comparison of lexical and value based approaches." 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT). IEEE, 2017.

[23] Dey, Umid Kumar, and Alaa Noor. "Comparative Exploration Of Prediction Algorithms For Sentiment Analysis Using NLP." 2019 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2019.

[24] Rathor, Abhilasha Singh, Amit Agarwal, and Preeti Dimri. "Comparative Study of Machine Learning Approaches for Amazon Reviews." Procedia computer science 132 (2018): 1552-1561.

[25] Wang, Jin, et al. "Dimensional sentiment analysis using a regional CNN-LSTM model." Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2016.

[26] Wan, Yun, and Qigang Gao. "An ensemble sentiment classification system of twitter data for airline services analysis." 2015 IEEE international conference on data mining workshop (ICDMW). IEEE, 2015.

[27] Nguyen, Huy Tien, and Minh Le Nguyen. "An ensemble method with sentiment features and clustering support." Neurocomputing (2019).

[28] Abdi, Asad, et al. "Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion." Information Processing & Management 56.4 (2019): 1245-1259.

[29] Çano, Erion, and Maurizio Morisio. "A deep learning architecture for sentiment analysis." Proceedings of the International Conference on Geoinformatics and Data Analysis. ACM, 2018.

[30] Zharmagambetov, Arman S., and Alexandr A. Pak. "Sentiment analysis of a document using deep learning approach and decision trees." 2015 Twelve international conference on electronics computer and computation (ICECCO). IEEE, 2015.

[31] Ramadhani et al. "Twitter sentiment analysis using deep learning methods." 2017 7th International Annual Engineering Seminar (InAES). IEEE, 2017.

[32] Jianqiang, Zhao, Gui Xiaolin, and Zhang Xuejun. "Deep convolution neural networks for twitter sentiment analysis." IEEE Access 6 (2018): 23253-23260.

[33] Fu, Xianghua, et al. "Lexicon-enhanced LSTM with attention for general sentiment analysis." IEEE Access 6 (2018): 71884-71891.

[34] Araque, Oscar, et al. "Enhancing deep learning sentiment analysis with ensemble techniques in social applications." Expert Systems with Applications 77 (2017): 236-246.

[35] Alharbi, Ahmed Sulaiman M., and Elise de Doncker. "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information." Cognitive Systems Research 54 (2019): 50-61.

[36] Akhtar, Md Shad, et al. "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis." Knowledge-Based Systems 125 (2017): 116-135.

[37] Xia, Rui, et al. "Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis." Information Processing & Management 52.1 (2016): 36-45.

[38] You, Quanzeng. "Sentiment and emotion analysis for social multimedia: Methodologies and applications." Proceedings of the 24th ACM international conference on Multimedia. ACM, 2016.

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

[40] You, Quanzeng, et al. "Joint visual-textual sentiment analysis with deep neural networks." Proceedings of the 23rd ACM international conference on Multimedia. ACM, 2015.

[41] Li, Weifeng, and Hsinchun Chen. "Identifying top sellers in underground economy using deep learning-based sentiment analysis." 2014 IEEE Joint Intelligence and Security Informatics Conference. IEEE, 2014.

[42] Dos Santos, Cicero, and Maira Gatti. "Deep convolutional neural networks for sentiment analysis of short texts." Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014.

[43] Ghiassi, Manoochehr, James Skinner, and David Zimbra. "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network." Expert Systems with applications 40.16 (2013): 6266-6282

[44] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).

[45] Chen, Chaotao, Run Zhuo, and Jiangtao Ren. "Gated recurrent neural network with sentimental relations for sentiment classification." Information Sciences 502 (2019): 268-278.

[46] Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the 2013 conference on empirical methods in natural language processing. 2013

[47] Heikal, Maha, Marwan Torki, and Nagwa El-Makky. "Sentiment analysis of Arabic Tweets using deep learning." Procedia Computer Science 142 (2018): 114-122.

[48] Chen, Tao, et al. "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN." Expert Systems with Applications 72 (2017): 221-230.

[49] Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Improving deep neural network design with new text data representations." Journal of Big Data 4.1 (2017)

[50] Parlar, Tuba, Selma Ayşe Özel, and Fei Song. "QER: a new feature selection method for sentiment analysis." Human-centric Computing and Information Sciences 8.1 (2018)

[51] Rong, Wenge, et al. "Structural information aware deep semi-supervised recurrent neural network for sentiment analysis." Frontiers of Computer Science 9.2 (2015): 171-184.

[52] Chachra, Anmol, Pulkit Mehndiratta, and Mohit Gupta. "Sentiment analysis of text using deep convolution neural networks." 2017 Tenth International Conference on Contemporary Computing (IC3). IEEE, 2017.

[53] Zhang, Jingren, Fangai Liu, and Weizhi Xu. "The Application of An Optimized Convolutional Neural Network Model in Film Criticism." 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT). IEEE, 2019.

[54] Che, Wanxiang, et al. "Sentence compression for aspect-based sentiment analysis." IEEE/ACM Transactions on audio, speech, and language processing 23.12 (2015): 2111-2124.

[55] Zhou, Shusen, Qingcai Chen, and Xiaolong Wang. "Fuzzy deep belief networks for semi-supervised sentiment classification." Neurocomputing 131 (2014): 312-322.

[56] Rong, Wenge, et al. "Structural information aware deep semi-supervised recurrent neural network for sentiment analysis." Frontiers of Computer Science 9.2 (2015): 171-184.

[57] Poria, Soujanya, Erik Cambria, and Alexander Gelbukh. "Aspect extraction for opinion mining with a deep convolutional neural network." Knowledge-Based Systems 108 (2016): 42-49.

[58] Dongsuk, O., et al. "Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph." Proceedings of the 27th international conference on computational linguistics. 2018.

[59] Nakagawa, Tetsuji, Kentaro Inui, and Sadao Kurohashi. "Dependency tree-based sentiment classification using CRFs with hidden variables." Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010.

[60] Heikal, Maha, Marwan Torki, and Nagwa El-Makky. "Sentiment analysis of Arabic Tweets using deep learning." Procedia Computer Science 142 (2018): 114-122.

[61] Manda, Kundan Reddy. “Sentiment Analysis of Twitter Data Using Machine Learning and Deep Learning Methods.” 2019. Web. 21 Oct 2019.

[62] Grossi, Enzo, and Massimo Buscema. "Introduction to artificial neural networks." European journal of gastroenterology & hepatology 19.12 (2007): 1046-1054.

[63] Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow In-text: (Géron)

[64] DiPietro, Robert, and Gregory D. Hager. "Deep learning: RNNs and LSTM." Handbook of Medical Image Computing and Computer Assisted Intervention. Academic Press, 2020. 503-519

[65] Wang, Jin, Bo Peng, and Xuejie Zhang. "Using a stacked residual LSTM model for sentiment intensity prediction." Neurocomputing 322 (2018): 93-101.

[66] Liao, Shiyang, et al. "CNN for situations understanding based on sentiment analysis of twitter data." Procedia computer science 111 (2017): 376-381.

[67] Chen, Tao, et al. "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN." Expert Systems with Applications 72 (2017): 221-230.

[68] Zhao, Huizhen, et al. "Research on a learning rate with energy index in deep learning." Neural Networks 110 (2019): 225-231.

[69] Tušar, Tea, et al. "A study of overfitting in optimization of a manufacturing quality control procedure." Applied Soft Computing 59 (2017): 77-87.

[70] "Data". Cs.Cornell.Edu, 2019, https://www.cs.cornell.edu/people/pabo/movie-review-data/.

[71] "Deeply Moving: Deep Learning For Sentiment Analysis". Deeply Moving: Deep Learning For Sentiment Analysis, 2019, http://nlp.stanford.edu/sentiment/.

[72] Maas, Andrew L., et al. "Learning word vectors for sentiment analysis." Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1. Association for Computational Linguistics, 2011."

[73] Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014

[74] Zhou, Jie, et al. "Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision and Challenges." IEEE Access (2019).

[75] Hatcher, William Grant, and Wei Yu. "A survey of deep learning: platforms, applications and emerging research trends." IEEE Access 6 (2018): 24411-24432.

[76] Henry, Sam, Clint Cuffy, and Bridget T. McInnes. "Vector representations of multi-word terms for semantic relatedness." Journal of biomedical informatics 77 (2018): 111-119.

[77] Ling, Yuan, et al. "Integrating extra knowledge into word embedding models for biomedical nlp tasks." 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017.

[78] FOSS, GREG. MODDERMAN. PRACTICAL DATA SCIENCE WITH SAP. O'reilly MEDIA, INC, USA, 2019.

[79] Rehman, Anwar Ur, et al. "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis." Multimedia Tools and Applications (2019): 1-17.

[80] Mozer, Michael C., Denis Kazakov, and Robert V. Lindsey. "State-Denoised Recurrent Neural Networks." arXiv preprint arXiv:1805.08394 (2018).

[81] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning,C., Ng, A., and Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In EMNLP, 2013.

[82] Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "A convolutional neural network for modelling sentences." arXiv preprint arXiv:1404.2188 (2014).

[83] Bespalov, Dmitriy, et al. "Sentiment classification with supervised sequence embedding." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2012.