Multi-Modal Emotion Extraction using Neural Network Hybrid Activation Function

Ravikumar Thallapalli, Dr.G.Narsimha
Research Scholar, Faculty of Informatics, Department of CSE, Osmania University, Hyderabad-500007, Telangana, India.
Professor, Department of Computer Science and Engineering, JNTUH University College of Engineering, Sultanpur, Sangareddy, Telangana, India.

Abstract:
The higher availability of the online platforms for sharing feedback for various aspects or products, the demand for analyzing the contents to estimate customer satisfaction level is increasing. The service providers are keen to understand the product sentiment from these informal reviews, which as per them demonstrate the true sentiment for the product. Thus, a good number of research attempts can be seen for analyzing the texts for extracting the sentiments. Nonetheless, these methods ignorea few foundations facts by using standard methods for tokenization, lemmatization and further extraction of the sentiments using tagging methods and cause either underfitting or overfitting problems. Hence, this proposed method demonstrates a few unique strategies such as differential analysis for tokenization, reduced time complex lemmatization, threshold-based sentiment extraction and further summarization. Finally, this work outcomes into a 99% accuracy using enhanced sigmoid based neural network activations and unique strategy for weight correction in the neural networks.

Keywords: Text Pre-Processing, Differential Analysis, Dictionary Based Lemmatization, Categorization, Neural Network, Sigmoid Activation, Sentiment Extraction
DOI: 10.24297/j.cims.2023.1.25

1. Introduction

The proliferation of Internet-enabled devices has led to a rise in the number of people who prefer to do their shopping online. Sentiment analysis of online reviews on a massive scale can improve customer satisfaction. An existing body of work has provided a novel model for sentiment analysis, dubbed SLCABG, which makes use of a sentiment lexicon, a convolutional neural network, and a Bidirectional Gated Recurrent Unit for attention (BiGRU). The SLCABG model is able to better analyze the tone of product reviews since it blends a sentiment lexicon with deep learning. Combining deep learning with sentiment lexicons is what SLCABG is all about. The reviews are made more positive by using the sentiment lexicon. The main sentiment and context elements of the reviews are retrieved using a GRU network and the attention technique to weight. Finally, sort qualities according to their emotional weight. In this work, we prepare and scrape data from dangdang.com to use in actual evaluations and tests. Researching the sentiments of the Chinese population has never been easier, thanks to this massive dataset. Experiments suggest that the method enhances text sentiment analysis [1].

Too far, there have been minimal efforts made to construct linguistic "resources for the Urdu language" and assess user attitude, despite the fact that over 169 million people speak Urdu and a great deal of Urdu data is uploaded on social networks every day. With this effort, we hope to develop a resource-constrained Urdu benchmark dataset for sentiment analysis and assess "machine and deep learning" approaches to the problem. Each feature type was investigated using a variety of different algorithms, including RF, NB, SVM, AdaBoost, MLP, and LR (1D-CNN and LSTM). For sentiment analysis, the author recommends using LR with word n-gram features [2].

The attention mechanism of long-term "short-term memory (LSTM) neural networks" is used to learn and detect text sentiment. When analyzing text for sentiment, most deep learning approaches ignore the moderating effect of emotion on feature extraction. Sentiment classification performance and "learning text sentiment features" may be negatively impacted if higher-level abstractions are ignored. A model for text sentiment detection is presented in this research; it is called AEC-LSTM and it uses "emotional intelligence (EI) and attention" to enhance the LSTM network and solve the problem. Using EI, LSTM networks' feature learning skills are improved, yielding an emotion-enhanced LSTM (ELSTM) with emotion-modulated learning. To better capture structural patterns in text sequences, ELSTM combines convolution, pooling, and concatenation. We provide a framework at the level of topics for adjusting the hidden representational relevance of texts in real time. Effective representation and categorization of sentiment is made possible by EI and the attention mechanism via the use of latent sentiment semantic information in the subject and context of the text. The author claims that his method can outperform current deep learning-based systems used for sentiment categorization [3].

Further, the rest of the paper is organized such that in Section – II, the recent literature reviews are discussed to identify the bottlenecks in the present research, the identified problems are summarized in Section – III, the proposed solutions are furnished using mathematical models in Section – IV, the proposed algorithms with automated framework is furnished in the Section – V, the obtained results are furnished and compared with the benchmarked researches in Section – VI and VII respectively. Finally, the research conclusion is presented in Section VIII.

2. Recent Research Reviews – Literature Review

The proliferation of content on the internet in non-standard languages has accelerated in recent years. When people are able to communicate without linguistic obstacles, everyone benefits. A total of 170,2 million individuals can communicate in Urdu. An individual's or a group's sentiments can be analyzed through sentiment analysis. Researchers have recently taken an interest in Urdu sentiment analysis. We don't make nearly enough use of deep learning for sentiment analysis in Urdu. Urdu's complex morphology makes it difficult to comprehend written text. A "framework for Urdu Text Sentiment Analysis (UTSA)" is proposed in this research, which makes use of deep learning techniques and word vector representations to determine the tone of written Urdu text. There is a comparison of deep learning algorithms for sentiment analysis. Among them are "Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Long Short LSTM, BiLSTM-ATT, and C-LSTM all make use of layered architectures. The convolution layer of a CNN employs numerous filters. An analysis of both supervised and "unsupervised self-trained embedding" approaches is performed in the sentiment classification task. When compared to other deep learning models, BiLSTM-ATT outperformed them all with an accuracy of 77.9 percent and an F1 score of 72.7 percent" [4].

Since sentiment analysis is so useful in politics, the media, and other areas, it has become a significant area of study for NLP scholars. Assisting sentiment analysis are word embeddings. One significant shortcoming of current sentiment embeddings is that they include sentiment lexicons directly into conventional word representation. This method is limited in that it can only accurately interpret a word's emotional connotation in a single setting. According to this study, using public opinion might help find a solution. The writers factored in setting to establish which Microsoft Concept Graph feeling was more salient. To ensure accurate embedding, the authors of this study compiled a vocabulary of sentiment intensity terms that draws from many semantic domains. Two improved word embedding techniques were employed by the authors. This research [5] verifies the efficacy of sentiment-based word embeddings by contrasting them with traditional and sentiment embeddings on six widely used datasets.

Analyzing, inferring, and extracting inferences from subjective writings is what sentiment analysis is all about. Sentiment analysis is used by businesses to get insight into public opinion, market trends, brand reputation, customer experiences, and the impact of social media. Document, phrase, and feature-based breakdowns are all possible. This article discusses the most up-to-date methods for aspect-based sentiment analysis. Popular methods include dictionaries, ML, and deep learning. The methods of deep learning are the subject of this article's review. The current state of deep learning is discussed along with the available techniques, benchmark datasets, and evaluation criteria. We wrap up by addressing the problems we've encountered and the next directions of our investigation [6].

Recent years have demonstrated that deep neural networks may be quite useful for this task. Since deep neural networks can automatically extract features from data, the features may be found in the networks' intermediate representations. The distinct architectures of deep neural networks make it possible to extract a wide variety of feature classes. The authors used multi-view classifiers to combine data from several types of neural networks, allowing for more accurate document-level sentiment analysis. For classification purposes, the multi-view deep network makes use of "convolutional and recursive neural" features. Multi-view deep neural networks have been shown to be more effective in terms of "efficiency and generalization" [7].

To ascertain the author's or speaker's perspective on a topic, sentiment analysis employs computationally automated cognitive techniques. More and more thoughts are being shared in text form on social networks, making it more challenging to identify emotional trends. Learning a "pretrained language model contextual "'s representation is more effective than learning words by themselves. Although feature-based and fine-tuning procedures are the two most fundamental approaches to apply pretrained language models to downstream tasks, they are generally studied independently. Many sentiment analysis tasks cannot be adequately handled by using only task-specific contextual representations. The authors weighed the advantages and disadvantages of the various approaches and proposed a "broad multitask transformer network (BMT-Net)." For optimal results, BMT-Net employs feature-based and fine-tuning techniques. The goal of this study was to investigate effective and contextual representation. In this work, we offer a system that uses multitask transformers to generalize learnt representations. Because of its extensive and thorough search for critical attributes, BMT-Net is able to pick up the broad learning system's rich contextual representation. There was a comparison between the Stanford Sentiment Treebank version 2 (SST-2) and SemEval Sentiment Analysis on Twitter (Twitter). The combined deep and broad augmentation of the representation leads to a $F1$-score of 0.778

on Twitter and 94.0 percent accuracy on the SST-2 dataset. These empirical findings validate the usefulness of sentiment analysis recognition and stress the significance of contextual factors across both narrow and wide domains [8].

Several years have been spent creating Arabic sentiment analysis machine learning algorithms (ASA). A majority of these efforts have used shallow machine learning strategies rather than deep learning. "The low performance of ASA's deep learning models can be attributed to their use of noncontextualized embedding techniques. To do so, we propose a unique deep learning-based MPAN model that can simultaneously compute contextualized embeddings at the character, word, and sentence levels. In order to achieve state-of-the-art precision, the MPAN model combines attention vectors from many layers. MPAN shows state-of-the-art performance by outperforming all previous ASA baselines on 34 publicly available datasets. For two multidomain datasets, the proposed model achieves state-of-the-art accuracy (95.61 percent for a binary classification collection and 94.2 percent for a tertiary classification collection)". Using the public IMDB movie review dataset, MPAN is shown to perform to 96.13 percent accuracy [9]. Every day, millions of words, hours of audio, and countless hours of video are added to the internet. These numbers can shed light on popular sentiment and international tendencies. Ads are pitched by companies depending on a user's internet habits. Urdu has fewer tools available, hence it is more challenging to analyze raw data for trends. In the previous study, we introduced a novel multimodal dataset in Urdu, consisting of 1372 phrases. Second, the authors developed an MSA framework that takes into account both verbal and nonverbal cues to assess the underlying emotional state of a given situation. Fusion methods have also been employed by the authors to enhance sentiment polarity prediction. Experiments showed that using the suggested technique improved polarity identification accuracy from 84.32 to 95.35 [10].

Here, we look at how deep learning may be used for aspect-level sentiment classification (ASC), the process by which the polarity of an aspect's sentiment is established inside a text. The recent widespread success of ASC based on deep learning has attracted the attention of academics and professionals. By providing a taxonomy of existing methodologies and evaluating their performance, this study intends to bridge knowledge gaps. Improve the ability to objectively assess the performance of different methodologies by standardizing evaluation processes and making use of shared datasets. To demonstrate ASC's development, this article gives a thorough analysis of state-of-the-art deep learning-based methods. Recent studies on ASCs are analyzed using deep learning. Using deep learning, the authors categorized existing ASC methods and provided an overview of future directions. The authors have collected all publicly available benchmark ASC datasets, giving researchers a comprehensive collection of data with which to assess their own results. As a conclusion, certain thorny unanswered topics are addressed, and some avenues for future study are given [11].

The emotional tone of a document may be automatically determined by using sentiment analysis. To enhance online review sentiment categorization, conventional deep learning sentiment classification models place a premium on algorithm improvement. To put it another way, the model's classification performance will suffer if there isn't enough quality control in the manual labeling of the sample data. The use of a deep learning sentiment classification system that relies on weak tagging can help cut down on the time and effort spent categorizing data manually. In order to improve sentiment classification performance, the model mitigates the effect of noise in weak tagging information, which may indicate review sentiment trends but also contains noise. The experimental findings demonstrate that the weak tagging information-based deep learning sentiment classification model achieves better outcomes than the standard deep model without raising human cost [12].

Twitter and Facebook have made it easier for people to voice their opinions on current events, products, and services. Nevertheless, identifying how users feel about given topics is essential for making conclusions. Deep learning methods including LSTM, GRU, BiLSTM, and convolutional neural networks are used for sentiment categorization (CNN). Word2Vec and FastText, two word embedding methods, are being investigated for their potential to map text to real number vectors. There are advantages and disadvantages to using deep learning with word embedding. Sentiment classification in NLP is aided by a combination of deep learning and word embedding (NLP). This research suggests a new type of deep learning model that incorporates Word2Vec, FastText, and character-level embeddings (LSTM, GRU, BiLSTM, CNN). In order to classify texts according to their emotional tone, the suggested method incorporates deep learning word embedding features. Utilizing basic deep learning methods, we were able to evaluate the efficacy of the suggested approach. Compared to previous research [13], the proposed method performs significantly better when classifying emotions.

Analysis of sentiment reveals opposing views in written material. This method of analysis may be used to spot weird OS log events, which are indicators of negativity. The methods now in use either rely on human exploration, rigid criteria, or automated discovery. In this research, we provide a sentiment analysis method that can be used to examine OS logs for out-of-the-ordinary behavior, and it is based on deep learning. Sentences serve as the foundation for log messages. To understand the feelings conveyed in a book, authors employ GRU networks. Operating system log files tend to reflect more negativity than positivity. In order to address the issue of class imbalance, the authors built a GRU layer on top of a Tomek link-based approach. The suggested method has an F1 of 99.84 and an accuracy of 99.93 percent when used to analyze OS logs in search of anomalous occurrences [14].

The popular microblogging service Twitter has become a central forum for people to broadcast their views and sentiments to the world. Consequently, the field of tweet tone analysis was born. The accuracy and speed of sentiment analysis provided by deep learning models like CNN and Bi-LSTM has surpassed those of older, less sophisticated algorithms. While CNN's convolutional and max-pooling layers are effective at eliciting relevant local input, the network is unable to learn sequential correlations. In deep learning scenarios, Bi-use LSTM's of two LSTM orientations improves performance, but it is unable to extract local features in parallel. It is not recommended to do sentiment analysis with just one CNN or Bi-LSTM. The CNN-Bi-LSTM architecture is presented in this research. To measure the volume of tweets, ConvBiLSTM employs a word embedding model, a CNN layer that takes feature embeddings as input and outputs lower-dimensional features, and a Bi-LSTM model that analyzes CNN input to deliver a classification result. The proposed model was evaluated using both Word2Vec and GloVe. ConvBiLSTM used data from Tweets and SST-2. On the dataset of recovered Tweets, ConvBiLSTM had the highest accuracy at 91.13 percent.

E-commerce has made buying goods online as commonplace as buying groceries or clothing. Big data sentiment analysis of shopper feedback from online stores may make a noticeable difference for consumers. While deep neural network-based sentiment analysis algorithms achieve more accuracy than traditional methods without the need for human-designed features, they still fall short and are hindered during training by over-fitting and vanishing gradients. The present study's authors developed MBGCV to address these issues and enhance its precision. Multichannel concepts like BiGRUs, CNNs, and VILs are used in this approach (VIB). When using many channels, you can get nuanced expressions of emotion. Afterwards, CNN collects channel-specific attributes using the context data gathered by BiGRU. By combining the VIB and Maxout activation functions, this model is able to deal with overfitting and vanishing gradient. The authors of this study use real-world review datasets to undertake extensive experiments demonstrating that the suggested methodology may enhance text sentiment analysis [16].

How a culture responds to a crisis is indicative of its values and political will. Often, choices are driven not by the will of the country but by circumstances, cultural pressure, or urgent need. As a result, some people will be happy, and others will be sad. As a result of COVID-19, other governments made similar decisions. In recent months, COVID-19, the pandemic, lockdowns, and related hashtags have dominated online conversation. Although geographically close, countries often react in various ways to events that occur nearby. The positions taken by the governments of Denmark and Sweden on this issue could hardly be more divergent. There was largely unbroken support at home, in contrast to the anxious and hostile reception they received among their South Asian neighbors. This study evaluates the government's response to the novel Coronavirus from many cultural perspectives. On the sentiment140 dataset, state-of-the-art accuracy has been developed using deep long short-term memory (LSTM) models. Validating supervised deep learning models with emoticons from scraped tweets is novel [17].
Research into Twitter's multilingual data sets is expanding. Using a “recurrent neural network, a convolutional neural network, and a hierarchical attention network, the authors identify stemmed Turkish Twitter data for sentiment analysis (HAN). The training data is then expanded by one of three data augmentation methods (Shift, Shuffle, or Hybrid)”. The results from both DL and TML models were compared. Augmentation tactics help DL models whereas stemmed data hampers TML models. TML models outperform DL models in terms of training-time (TTM) and runtime (RTM) complexity, while DL models score higher in the most critical performance variables and in the average performance rankings [18].

For a complete “Aspect-Based Sentiment Analysis (ABSA), the duties of Aspect Mining (AM) and Aspect Sentiment Classification (ASC) must be completed (ASC). Recent state-of-the-art results in AM and ASC have been accomplished using supervised deep sequence learning models, since the problems can be formulated as a sequence labeling problem to predict the aspect or emotion label for each word in the sequence. These supervised models need excessive labeled evaluations and can complete just one of the two tasks. An ABSA SEmi-supervised Multi-task Learning (SEML) paradigm is proposed in this study. There are three primary characteristics of SEML. A small number of labeled reviews and a huge collection of unlabeled reviews on the same topic are used in SEML to provide semi-supervised sequence learning. For (2), SEML performs both tasks concurrently by using three layers of bidirectional recurrent neural layers to learn review representations. When it comes to improving representation learning and prediction accuracy, SEML employs a Moving-window Attentive Gated Recurrent Unit (MAGRU)" . Semantic information that is close at hand in a review is crucial for ABSA's prediction task. There are four SemEval datasets that are used to assess ABSA. Compared to the best existing models, SEML performs better [19].

Processing negation is a part of sentiment analysis. You must submit your rejection in writing. Machine learning-based language models can make mistakes when the polarity of an entire phrase changes, such as when a negation word is used. The use of automated opinion mining has made it crucial to properly handle negative terms. Natural languages typically use implicit negations rather than explicit ones. One of the fields that most frequently employs rule-based negation handling is healthcare. It is challenging to develop a generic machine learning model for negation due to its peculiar syntactic structure. These studies investigate the factors (sentence cue and scope) that influence polarity reversal. The authors detail a deep neural network model for negation handling that is based on long short-term memory (LSTM). This model picks up negation characteristics from a labeled training dataset. The authors employed ConanDoyle, a narrative corpus annotated with negation information, to train and evaluate their models. By initially locating negation cues, the proposed model uses bidirectional LSTM to infer the cue's association with other words. In order to improve polarity modeling, the authors developed features at the word level. When compared to SVM, HMM, and CRF models, LSTM-based nonlinear language models perform better. As demonstrated by the negation test (F1 = 93.34%), BiLSTM is superior to the rule-based model [20].

Opinions on a wide range of issues are discussed and debated in web 2.0 communities, such as blogs and Twitter. With the use of sentiment analysis, we can put people's varied feelings into distinct buckets. The scope of sentiment analysis extends far beyond the traditional corporate realm. In order to quantify sentiment, we need to know how often each interest class occurs. More research is needed to analyze the performance of classifiers and the effect of new features on sentiment analysis. The results of deep learning methods are superior to those of traditional machine learning algorithms [21].

Vaccines are being developed, and their efficacy is being tested in communities all over the world as COVID-19 continues to spread. The current state of vaccination campaigns is hampered by public hostility and skepticism of the vaccine. The purpose of "this study was to examine the opinions of Twitter users in the United States, United Kingdom, Canada, Turkey, France, Germany, Spain, and Italy during COVID-19 with regards to vaccination and vaccine kinds. This study utilized a dataset of 928,402 tweets categorized by location and language from 195 countries. A total of four BERT models (mBERT-base, BioBERT, ClinicalBERT, and BERTurk) were used to classify tweets into various policy, health, media, and other categories" . Twitter posts were analyzed for sentiment on the top six COVID-19 vaccines mentioned in the available datasets. Until According to the author, this is the first study of its kind to investigate how the general population feels about vaccinations. Different vaccines and the levels of support for them vary from country to country, and these differences were highlighted. The F1 Score success of the method used in this investigation ranged from 84% to 88%, with an average accuracy of 87% [22].

3. Problem Identifications

After analyzing the recent research improvements for sentiment extraction, the authors conclusively furnish the following challenges as,

- Firstly, most of the work does not implement the parts of speech tagging in order to reduce the time complexity. Hence during the threshold detection for statical models or machine learning driven neural network models, the priority for weight adjustments is completely overlapped. This clearly results in overfitting of the sentiment detection models.
- Secondly, due to the writing style for each and every author, the formation of the sentence changes, which is primarily reflected in the use of the verb forms. These changes in the verb forms can further influence the categorization of the sentiments. Thus, neutralization of the verb forms or any other variable grammar forms is a must.
- Finally, the many parallel research works are influencing the use of standard classification methods for the detection of the sentiments. Nevertheless,if the corpus size is smaller, then the accuracy of these standard classification models is bounded with a lower accuracy. Hence the use of deep learning driven neural networks models are recommended.

Henceforth, in the next section of this work, the proposed models using mathematical formulations are furnished.

4. Proposed Solutions

After the detailed understanding of the existing systems and the bottlenecks of the parallel research outcomes, in this section of the work, the proposed method is furnished.

Assuming that, the complete text corpus is T[], which contains two parts as ID as unique identifier of the text and the text for sentiment extraction as t. Thus, this can be formulated as,

$$T[] = < ID, t > \quad \text{(Eq.1)}$$

Further, based on this initial formulation, the rest of the proposed systems are formulated.

A.Sentence Level and Word Level Tokenization

Firstly, this proposed system proposes the tokenization using available dictionary based connectors as Con[]. The Con[] collection houses the standard sentence connector words and each word can be presented as Ci. Thus, this can be formulated as,

$$Con[] = < C_1, C_2, C_3, ...C_n > \quad \text{(Eq.2)}$$

Further, the initial text, t, must be divided into individual parts of the sentences, t1[] using the standard differential questions as,

$$t1[] = \frac{d\{t\}}{d\{Con[]\}} \quad \text{(Eq.3)}$$

Thus, the text corpus can be re-written as,

$$T[] = < ID, t1[] > \quad \text{(Eq.4)}$$

Once, the text is separated into parts of the sentences, each parts can now be separated into words, W[], using the stop symbols collection, st[]. Where st[] is the standard collection of stop symbols with each stop symbol, including spaces, can be represented as sti. Thus, this can be formulated as,

$$st[] = < st_1, st_2, st_3, ...st_m > \quad \text{(Eq.5)}$$

And,

$$w[] = \frac{d\{t1[]\}}{d\{st[]\}} \quad \text{(Eq.6)}$$

Hence, finally,

$$T[] = < ID, w[] > \quad \text{(Eq.7)}$$

B.Lemmatization

Secondly, this work also proposes the lemmatization of the text for overcoming the overfitting problem during the sentiment analysis. The general grammar, G[] form usually contains different verb forms, F[], and the source verb, S, as a collection. Thus, this can be represented as,

$$G[] = < FD[], S > \quad \text{(Eq.8)}$$

Further, each w[] must be converted to the original form of the verbs as,

$$S[] = \prod_{G.F[i]=T[j].w[k]} S \quad \text{(Eq.9)}$$

Thus, finally, the text corpus can be re-build as,

$$T[] = < ID, S[] > \quad \text{(Eq.10)}$$

C.Parts of the Speech Identification

Thirdly, the identification of the parts of the speech is also highly crucial for deciding the thresholds for each type of word categories. The proposed method again utilizes the dictionary, D[], for building the categorization of the parts of the speech. The dictionary houses the part of the speech, p, and the category, d, for each part. This can be formulated as,

$$D[] = < p[], d[] > \quad \text{(Eq.11)}$$

Thus, with the help of this dictionary, each word can be associated with speech categories as,

$$< s[], p[] > = \prod_{D[i].p[j]=T[k].S[l]} d \quad \text{(Eq.12)}$$

Further, the final corpus can be re-written as,

$$T[] = < ID, s[], d[] > \quad \text{(Eq.13)}$$

D.Sentiment Extraction using Neural Network Model

Finally, this proposed method furnishes an activation function based on the sigmoid function class for detection of the sentiments, sc[], from the text corpus as,

$$sc[] = \frac{1}{1+e^{-\{\sum_{i=0}^{n} s_i w_i \pm d_i\}}} \quad \text{(Eq.14)}$$

Thus, sc[] collection contains the final sentiment classes for each text available in the text corpus.

Further, based on these proposed methods, in the next section of this work, the proposed algorithms are furnished.

5. Proposed Algorithms And Framework

After the detailed analysis of the proposed method, this section of the work is dedicated to discussing the proposed algorithms and the automated framework for detection of the sentiments.

Firstly, the Text Pre-Processing with Differential Analysis (TP-DA) Algorithm is furnished.

Algorithm - I: Text Pre-Processing with Differential Analysis (**TP-DA**) Algorithm
Input: Text corpus as T[]
Output: Text corpus as T1[] with tokenized words
Process: Step - 1. For each text element in T[] as T[i] a. For each word in T[i] as T[i].t[j] i. If T[i].t[j] is connector word ii. Then, extract each part as P[k++] = {T[i].t[] - T[i].t[j]} b. End For c. For each parts in P[] as P[k] i. For each element in P[k] as P[k][l] 1. If P[k][l] is " " 2. Then, R[m++] = P[k][l-1] ii. End For d. End For e. Build T1[i] = R[m] Step - 2. Return T1[]

Preprocessing your text may be done in several ways. You should be familiar with the following methods, and I will do my best to stress the relevance of each. It's best practise to use lowercase letters, but I've run across circumstances when it was crucial to keep the capitalization. Foreseeing a file's source language is one such application. Java's "System" is not at all like Python's "system." To make matters worse for the classifier, converting them to lowercase renders them indistinguishable. Lowercasing is advantageous in most situations, although it may not always be the best choice. As a rough heuristic, stemming attempts to reduce words to their roots by removing their terminal affixes. It's possible that the terms "trouble," "troubled," and "troubles" will be changed to "troublin" due to the aforementioned severing of their terminals.

Secondly, the Dictionary Based Lemmatization with Categorization (DBLC) Algorithm is furnished.

Algorithm - II: Dictionary Based Lemmatization with Categorization (**DBLC**) Algorithm
Input: Text corpus as T1[] with tokenized words
Output: Text corpus as T2[] with lemmatized and categorized
Process: Step - 1. Initialize the grammar collection G[] with FD[] and S using API access Step - 2. Initialize the dictionary collection D[] with parts of speech, P and category d using API access Step - 3. For each text element in T1[] as T1[i] a. For each word in T1[i] as T1[i].w[j] i. If G[k].FD[] contains T1[i].w[j] ii. Then, S[k] = G[k].S iii. If S[k] == D[m].P iv. Then, T2[i] = <T1[i].w[j], D[m].P> b. End For Step - 4. End For Step - 5. Return T2[]

In the branch of computer science known as "natural language processing," software is used to analyze human-written text. This may be used for a wide variety of purposes, including "sentiment analysis, language translation", the identification of bogus news, the correction of grammatical errors, etc.

Natural language processing takes text as its input. This text's information comes from a wide variety of resources. Before the data can be utilized for analysis, it has to go through a lengthy processing and cleaning phase.

Thirdly, the Detection of Sentiments using Neural Network with Sigmoid Activation (DS-NN-SA) Algorithm is furnished.

Algorithm - III: Detection of Sentiments using Neural Network with Sigmoid Activation (**DS-NN-SA**) Algorithm
Input: Text corpus as T2[] with lemmatized and categorized
Output: Sentiment collection as SC[]
Process: Step - 1. For each corpus as T2[i] a. For each word in T2[i] as T2[i].w[j] i. Calculate the SC[k] using Eq. 14 ii. Calculate the summary for SC[k] iii. k++ b. End For Step - 2. End For Step - 3. Return SC[]

Further, based on the proposed algorithms, the final automated framework is furnished here [Fig – 1].

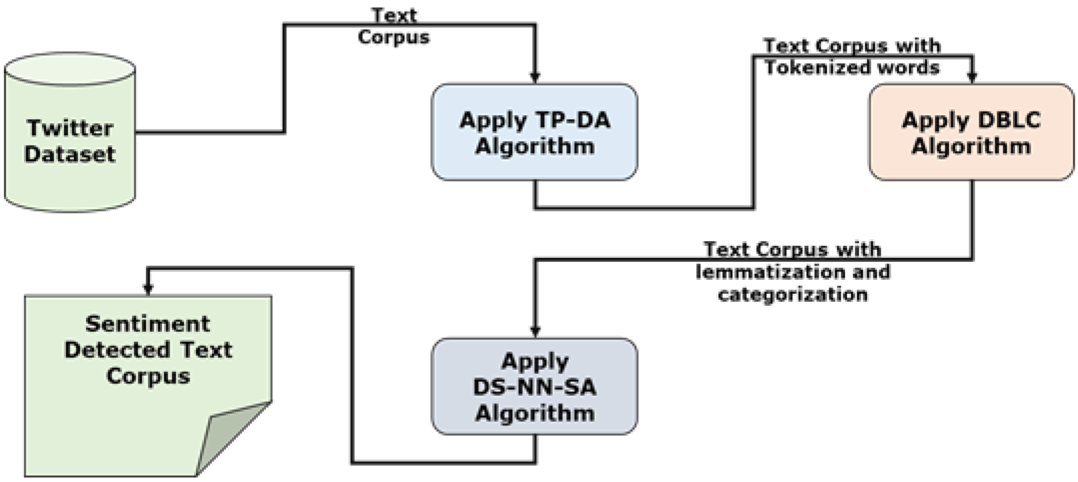

Fig. 1. Automated Framework for Sentiment Extraction

Henceforth, in the next section of this work, the obtained results are discussed.

6. Results And Discussions

After the detailed discussion on the proposed solution, the proposed algorithms, and the automated framework, in this section of the work, the obtained results are furnished.

Firstly, the dataset [23] description is furnished here [Table – 1].

TABLE I. DATASET DESCRIPTION

Parameter Name	***Value***
Number of Columns	17
Number of Rows	14,139
Positive Sentiment Rows	2363
Negative Sentiment Rows	9178
Neutral Sentiment Rows	2598
Missing Values	0
Outliers	0

It is natural to observethat the dataset contains no missing values and no outliers.

The description is visualized graphically here [Fig – 2].

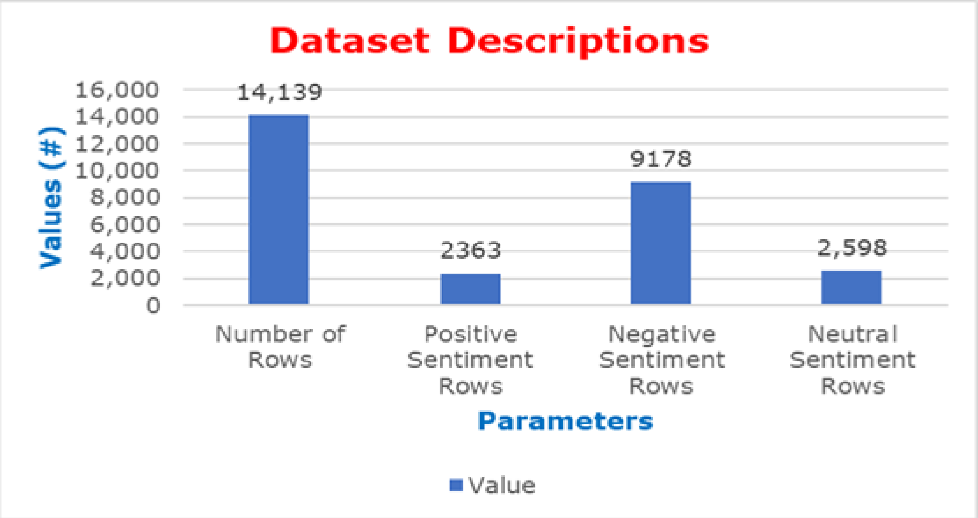

Fig. 2. Dataset Description

The complete dataset is tested for all 14139 rows, however for the presentation purposes only 15 are listed for all the phases.

Secondly, the outcome of the tokenization algorithm is furnished here [Table – 2].

TABLE II. TOKENIZATION PROCESS OUTCOMES

Tweet ID	*Actual Number of Words (A)*	*Extracted Number of Words (B)*	*Difference [Abs(A-B)]*
1	132	131	1
2	169	169	0
3	94	93	1
4	102	101	1
5	60	59	1
6	106	105	1
7	176	174	2
8	190	188	2
9	114	113	1
10	172	172	0
11	141	140	1
12	200	198	2
13	103	102	1
14	67	66	1
15	165	165	0

The proposed algorithm has performed a nearly 100% extraction of the words without the stop words or connector words. The mean deviation for the entire dataset is 1.4 words. The outcome is again visualized graphically here [Fig – 3].

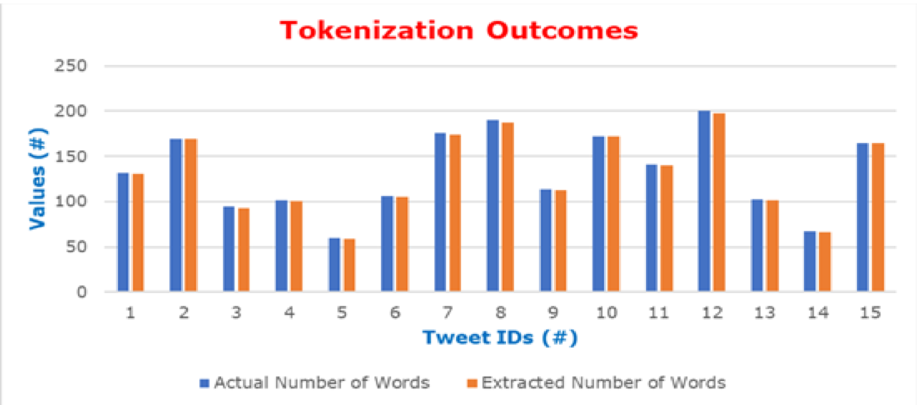

Fig. 3. Tokenization Process Outcomes

Thirdly, the lemmatization process outcomes are furnished here [Table – 3].

TABLE III. LEMMATIZATION PROCESS OUTCOMES

Tweet ID	*Actual Number of Verbs (A)*	*Lemmatized Number of Verbs (B)*	*Difference [Abs(A-B)]*
1	25	25	0
2	30	29	1
3	17	17	0
4	18	18	0
5	9	9	0
6	17	17	0
7	32	32	0
8	36	35	1
9	21	21	0
10	26	25	1
11	25	25	0
12	36	35	1
13	18	18	0
14	11	11	0
15	31	31	0

The proposed algorithm extracts the original verb forms from the text for nearly 100% of the cases. The mean deviation is only 0.27.

The outcome is visualized graphically here [Fig – 4].

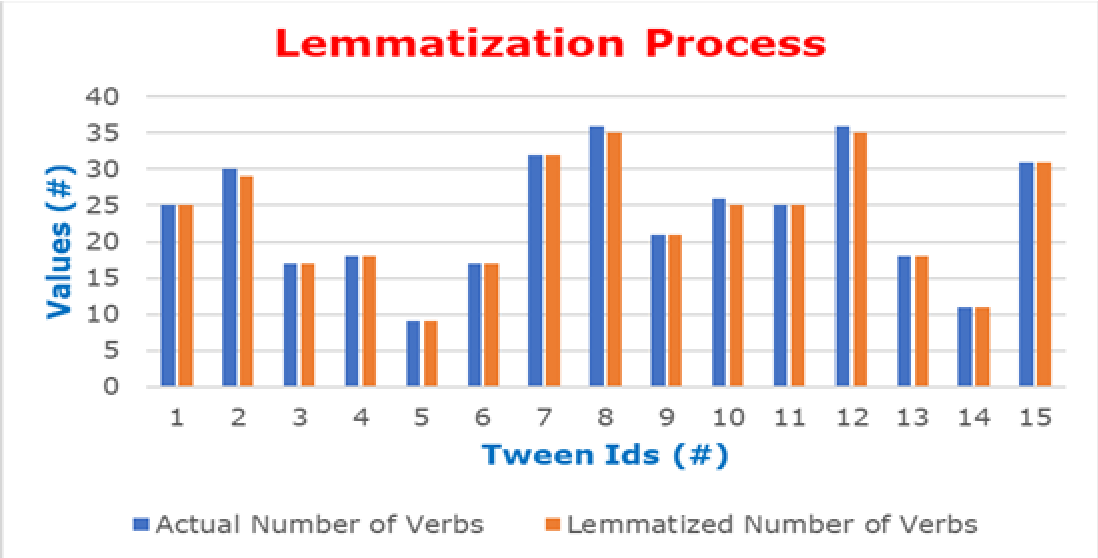

Fig. 4. Lemmatization Process Outcomes

Fourthly, the weight adjustment for 10 epoch is furnished here [Table – 4].

TABLE IV. WEIGHT CORRECTION OUTCOMES

Epoch	*Initial Weight for the Neural Model*	*Corrected Weight After the Epoch*
1	0.0870	0.1987
2	0.0771	0.2061
3	0.0666	0.2211
4	0.0597	0.2317
5	0.0563	0.2343
6	0.0499	0.2653
7	0.0487	0.2845
8	0.0462	0.3048
9	0.0440	0.3068
10	0.0417	0.2798

The proposed neural network model has demonstrated nearly 98% accuracy and the details are furnished here [Table – 5].

TABLE V. ACCURACY ANALYSIS

Epoch	*Accuracy (%)*	*Time (sec)*
1	96.84	68
2	97.22	68
3	97.63	72
4	97.73	68
5	97.88	68
6	98.30	72
7	98.33	68
8	98.30	68
9	98.36	72
10	98.41	68

The mean accuracy of the proposed model is nearly 99% for the complete dataset.
Further, the accuracy of this model is visualized graphically here [Fig – 5].

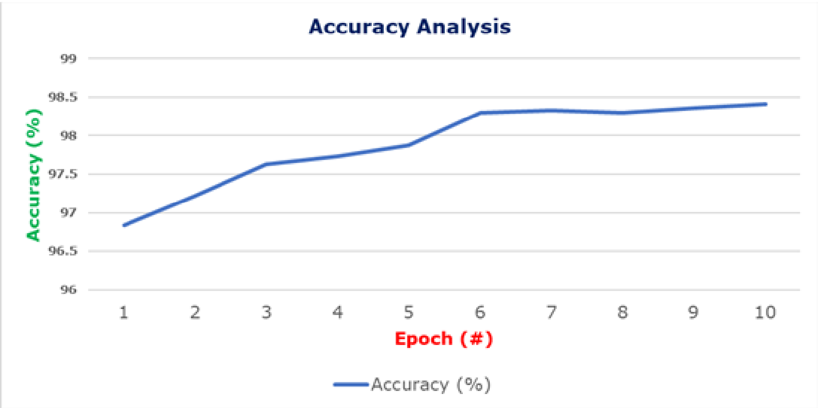

Fig. 5. Accuracy Analysis

Further, in the next section of this work, the obtained results are compared with the parallel research outcomes.

7. Comparative Analysis

After the discussions on the obtained results in the previous section of this work, in this section, the framework is compared with the benchmarked parallel research outcomes [Table – 6].

TABLE VI. COMPARATIVE ANALYSIS

Author, Year	*Framework Components*	*Model Complexity*	*Accuracy (%)*	*Time (sec)*
F. Huang et al. [3], 2022	Sentiment Extraction	$O(n^2)$	95	88
T. Zhang et al. [8], 2022	Sentiment Extraction	$O(n^2)$	92	112
İ. Aygün et al. [22], 2022	Sentiment Extraction	$O(n^2)$	96	95
C. Wang et al. [12], 2021	Speech Tagging, Sentiment Extraction	$O(n^2)$	84	72
Proposed Framework	Tokenization, Lemmatization, Speech Tagging, Sentiment Extraction	$O(n^2)$	99	69

Henceforth, it is natural to observe that the proposed model has outperformed the majority of the parallel research outcomes with higher components in the framework but with lesser sentiment extraction time.

Further, in the final section of this work, the research conclusion is presented.

8. Conclusion

Due to the proliferation of online forums for venting dissatisfaction with a certain service or product, there is a growing need to sift through the comments to gauge consumer sentiment. Service providers are interested in learning how customers feel about their products based on these evaluations, which they believe more accurately reflect consumers' real opinions. As a result, there have been a few studies aimed at evaluating texts in order to extract the emotions conveyed by the authors. However, by relying on tried-and-true techniques for tokenization with the proposed Text Pre-Processing with Differential Analysis (TP-DA) Algorithm, lemmatization, and sentiment extraction using the proposed Detection of Sentiments using Neural Network with Sigmoid Activation (DS-NN-SA) Algorithm through tagging approaches with proposed Dictionary Based Lemmatization with Categorization (DBLC) Algorithm, these strategies run the risk of underfitting or overfitting the data. As a result, the suggested approach exemplifies many novel approaches, including differential analysis for tokenization, complicated lemmatization with decreased processing time, threshold-based sentiment extraction, and subsequent summarization. In the end, the improved sigmoid-based neural network activations and one-of-a-kind weight correction approach used in this study lead to an accuracy of 99%.

References

1. L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020.
2. L. Khan, A. Amjad, N. Ashraf, H. -T. Chang and A. Gelbukh, "Urdu Sentiment Analysis With Deep Learning Methods," in IEEE Access, vol. 9, pp. 97803-97812, 2021.
3. F. Huang, X. Li, C. Yuan, S. Zhang, J. Zhang and S. Qiao, "Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 9, pp. 4332-4345, Sept. 2022.
4. U. Naqvi, A. Majid and S. A. Abbas, "UTSA: Urdu Text Sentiment Analysis Using Deep Learning Methods," in IEEE Access, vol. 9, pp. 114085-114094, 2021.
5. Y. Wang, G. Huang, J. Li, H. Li, Y. Zhou and H. Jiang, "Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis," in IEEE Access, vol. 9, pp. 37075-37085, 2021.
6. H. Liu, I. Chatterjee, M. Zhou, X. S. Lu and A. Abusorrah, "Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods," in IEEE Transactions on Computational Social Systems, vol. 7, no. 6, pp. 1358-1375, Dec. 2020.
7. H. Sadr, M. M. Pedram and M. Teshnehlab, "Multi-View Deep Network: A Deep Model Based on Learning Features From Heterogeneous Neural Networks for Sentiment Analysis," in IEEE Access, vol. 8, pp. 86984-86997, 2020.
8. T. Zhang, X. Gong and C. L. P. Chen, "BMT-Net: Broad Multitask Transformer Network for Sentiment Analysis," in IEEE Transactions on Cybernetics, vol. 52, no. 7, pp. 6232-6243, July 2022.
9. M. A. El-Affendi, K. Alrajhi and A. Hussain, "A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis," in IEEE Access, vol. 9, pp. 7508-7518, 2021.

10. U. Sehar, S. Kanwal, K. Dashtipur, U. Mir, U. Abbasi and F. Khan, "Urdu Sentiment Analysis via Multimodal Data Mining Based on Deep Learning Algorithms," in IEEE Access, vol. 9, pp. 153072-153082, 2021.
11. J. Zhou, J. X. Huang, Q. Chen, Q. V. Hu, T. Wang and L. He, "Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision, and Challenges," in IEEE Access, vol. 7, pp. 78454-78483, 2019.
12. C. Wang, X. Yang and L. Ding, "Deep Learning Sentiment Classification Based on Weak Tagging Information," in IEEE Access, vol. 9, pp. 66509-66518, 2021.
13. M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," in IEEE Access, vol. 8, pp. 58080-58093, 2020.
14. H. Studiawan, F. Sohel and C. Payne, "Anomaly Detection in Operating System Logs with Deep Learning-Based Sentiment Analysis," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 5, pp. 2136-2148, 1 Sept.-Oct. 2021.
15. S. Tam, R. B. Said and Ö. Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification," in IEEE Access, vol. 9, pp. 41283-41293, 2021.
16. T. Gu, G. Xu and J. Luo, "Sentiment Analysis via Deep Multichannel Neural Networks With Variational Information Bottleneck," in IEEE Access, vol. 8, pp. 121014-121021, 2020.
17. A. S. Imran, S. M. Daudpota, Z. Kastrati and R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets," in IEEE Access, vol. 8, pp. 181074-181090, 2020.
18. H. A. Shehu et al., "Deep Sentiment Analysis: A Case Study on Stemmed Turkish Twitter Data," in IEEE Access, vol. 9, pp. 56836-56854, 2021.
19. N. Li, C. -Y. Chow and J. -D. Zhang, "SEML: A Semi-Supervised Multi-Task Learning Framework for Aspect-Based Sentiment Analysis," in IEEE Access, vol. 8, pp. 189287-189297, 2020.
20. P. K. Singh and S. Paul, "Deep Learning Approach for Negation Handling in Sentiment Analysis," in IEEE Access, vol. 9, pp. 102579-102592, 2021.
21. K. Ayyub, S. Iqbal, E. U. Munir, M. W. Nisar and M. Abbasi, "Exploring Diverse Features for Sentiment Quantification Using Machine Learning Algorithms," in IEEE Access, vol. 8, pp. 142819-142831, 2020.
22. İ. Aygün, B. Kaya and M. Kaya, "Aspect Based Twitter Sentiment Analysis on Vaccination and Vaccine Types in COVID-19 Pandemic With Deep Learning," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2360-2369, May 2022.

23. KaggleNLP Tweet Dataset, URL - https://www.kaggle.com/code/serkanp/nlp-tweet-accuracy-0-94-data-viz/data