J. Electrical Systems 20-6s (2024): 2786-2810

[1]Anuj Kumar

[2]Shashi Shekhar

Hybrid Framework Integrating Lexicon and Learning Methods for Enhancing Sentiment Analysis Based on Patients' Tweets on Medicines

Abstract: - Sentiment analysis is the act of locating and categorizing the emotions conveyed in text data using text analysis tools. Previous studies have demonstrated the great potential for sentiment analysis of medicine reviews to provide useful data that will aid healthcare professionals and businesses in assessing the safety of pharmaceuticals after they have been sold. These details protect patients and strengthen their confidence in healthcare professionals. Existing frameworks for opinion investigation within the restorative field either take two types of methods one is lexicon methods and the other one is machine learning models. Learning-based approaches need annotated data, whereas lexicon-based approaches are domain-specific and have a smaller range of applications. To improve outcomes, this study employs a hybrid methodology that fusion machine and deep learning models with lexicon techniques. The reviews are annotated using all-purpose emotion lexicons like TextBlob and SenticNet. To extract meaningful features, we are using feature engineering methods TF and TF- IDF. Last but not least, classification tasks are carried out using learning models such as machine learning models and deep learning models for biomedical text. The execution of the prospective combined technique is appraised using performance metrics. According to experimental findings, combining lexicon- and learning-based techniques yields superior outcomes over using them alone. Additionally, TextBlob has demonstrated impressive results, providing an accuracy of 97% with LSTM-CNN and the Bio Bert model when used to a dataset of medication reviews, as well as 95% accuracy when used with TF and the logistic regression model. Additionally, TextBlob provides an accuracy of 94% when combined with TF and LSTM, and provides an accuracy of 97% when using with Bio Bert Model on a dataset including tweets.

Keywords: Sentiment analysis, Lexicon, SenticNet, review classification, Term Frequency, machine learning, deep learning.

I. INTRODUCTION

One of the most serious worldwide health crises is diabetes. It is a chronic disorder that develops when the body is unable to utilize insulin or create enough of it, and it is identified by the presence of elevated blood glucose levels. In a diabetic, glucose is still circulating in the blood due to the absence or inefficiency of insulin. Over time, the associated elevated blood glucose levels (known as hyperglycemia) harm numerous bodily tissues, culminating in the emergence of incapacitating and fatal health consequences [1]. There are currently thought to be more than 5, children with type 1 diabetes who are 14 years old or younger. Additionally, Impaired glucose tolerance is anticipated to affect 318 million people, placing them at a high risk of developing the condition in the future., while another 415 million persons are predicted to have diabetes. By 24, there would be more than 642 million individuals worldwide suffering from the disease if this upward trend is not stopped [2].

People with chronic diseases, such as diabetes, may frequently interact with medical personnel, but self-management of their condition requires special knowledge, a positive attitude, and support. As a result, social media sites like Twitter are a great resource for patients as they enable them to interact with others who share their disease and experiences. It offers the setting and resources necessary for peer assistance and knowledge exchange. However, finding perspectives might be challenging; for instance, a straightforward Twitter search for "diabetes" generates hundreds of replies. Users have trouble locating pertinent information via straightforward queries. As a result, methods for summarizing opinions that employ sentiment analysis (SA) or opinion-mining techniques are required. At the moment, there are not many papers that investigate the utilization of opinion mining in connection with diabetes. Understanding the influence that information like this might have on persons afflicted with diabetes and the members of their families requires first and foremost an awareness of the attitudes that are prevalent among users of social media platforms towards this health condition. The goal of this research is to assess the feelings

[1] Department of Computer Engineering & Application, GLA University, Mathura, Uttar Pradesh, India

[2] Department of Computer Engineering & Application, GLA University, Mathura, Uttar Pradesh, India

*Corresponding author: Anuj Kumar

conveyed in tweets and drug reviews that are hyper-focused on diabetes and other types of diseases and medicine uses that are posted on social media platforms. Sentiment analysis is used on the Twitter reviews of patient tweets and a standard dataset of drug review are used that was taken from the UCI Machine Learning Repository. The dataset includes the user-posted analyses of different diseases under certain conditions and their findings. The contributions made by this study are as follows:

- This study looks at the feasibility of using sentiment analysis on two different datasets: medicine reviews and Twitter tweets. as well as assesses the efficacy of publicly accessible sentiment lexicons in the medical arena.
- A hybrid framework is proposed for precise and efficient sentiment analysis in the healthcare domain. The developed hybrid model is based on Lexicons and uses both machine learning and deep learning model algorithms to achieve high accuracy.
- TextBlob and SenticNet are two sentiment lexicons that are used to annotate medicine reviews into positive, negative, and neutral groups. The efficacy of the two feature engineering approaches is evaluated on the Medicine reviews and Twitter tweets datasets. TF and TF-IDF feature engineering approaches are utilized for this purpose.
- To accomplish sentiment classification, five machine learning models and four deep learning models are used: Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Naive Byes (NB), and Support Vector Machine (SVM), and Deep learning models are Convolutional Neural Network

(CNN), Long Short Term Memory (LSTM), Hybrid of CNN-LSTM, and Bio-BERT.

- Several experiments are run to see how well the suggested technique performs on two publicly accessible datasets. In order to confirm its effectiveness, the execution of the suggested technique is contrasted with state-of-the-art approaches to validate its efficiency.

The remaining sections of the essay are structured as follows. Segment II examines studies that are comparable to the one under consideration. The pretreating techniques, lexicon methods, feature engineering methodology & machine learning, and deep learning algorithms selected for this project are covered in Segment III. The suggested approach is explained in the next Segment. The facts, analysis, and discussion are included in Segment V, and the ending is found in Segment VI.

II. RELATED WORK

Wherever This pilot study aims to categorize tweets about diabetes into several groups that are appropriate and pertinent to health. A deep learning algorithm and a classifier were used to split the 13,668 diabetes-related tweets into five groups. It was revealed that around 25.7% of the tweets were related to health, with 9.3% classified as Treatment and Medication, 9.9% classified as Preventive Measures, and 6.5% classified as Symptoms and Causes. Others accounted for more than 7% of all respondents. The investigation of hashtags taken from tweets arranged in each category found that the information they contained had a significant connection to health [1]. Through the use of the Twitter standard API, tweets that included the terms "diabetes," "t1d," and/or "t2d" were collected for a period of one week. The only information that was gathered was the user's text message as well as the number of followers. The analysis of sentiment was carried out with the help of SentiStrength. By using methods that analyze sentiment on social media, we may be able to get a better understanding of how the use of social media impacts people living with diabetes and their families, in addition to how it influences public health indicators [2] .In this study proposed a diabetes-specific ontology-based approach to sentiment analysis at the aspect level. Aspects' emotions are determined by analyzing the context words provided using N-gram algorithms. In order to test how well our approach works, we collected a corpus from Twitter that has been annotated with positive, negative, and neutral labels at the aspect level [3].Using machine learning techniques, we want to better comprehend the perspectives, emotions, and observations expressed within the diabetes online community by extracting both explicit and implicit cause-effect linkages from patient-reported, diabetes-related tweets [4].The current research endeavors to detect tweets on the patient experience as a supplementary useful tool for public health monitoring. Using a mixture of supervised machine learning and natural language processing, we extracted tweets on breast cancer patients' experiences [5]. The goal of this study is to get a better understanding of the general public's perspective on diabetes, namely types 1 and 2 [6]. The issues and challenges associated with improving sentiment accuracy and predicting the dynamic nature of sentiment evolution were the focus of this survey. This article provides a quantitative and qualitative comparison of the amount of user participation on Twitter relating to certain topics. The tweet volume for each health concern is compared with the actual death rate for that condition, and the results are given [7][8]. Examine the theory and practice of sentiment analysis to see how it might be put to use in a variety of contexts. Then, it examines the pros and drawbacks of the different methods by comparing them, contrasting them, and evaluating [9]. Dissect this notion, highlighting the gaps and under-explored, but critical, components of this topic that are required to achieve genuine sentiment comprehension. We examine the big changes that have led to its present significance. In addition, we make an effort to plot a future for this area of study that addresses many of the concerns that have been asked but not yet addressed [10].Article demonstrates how sentiment analysis may be used and explains how the process works in general. Then, it examines the methods that have been used in order to get a complete picture of their benefits and downsides by reviewing, comparing, and investigating them. Next, the difficulties inherent in sentiment analysis are laid forth in an effort to shed light on potential solutions[11]. esearch looks at the practicality, breadth, and applicability of a group of machine learning (ML) approaches for consumer sentiment analysis (CSA) of online reviews in the travel and hotel industries. The early 2s saw the beginning of research into sentiment analysis. Since that time, a number of approaches to analyzing the thoughts and feelings expressed by people through online opinion sources (such as blogs, forums, or commercial websites) have been developed [12-14].. People are sharing their thoughts on a variety of subjects on social networks like Twitter, which has resulted in an increased level of interest in this area in modern times [15,17]. The majority of these activities are based on the semantic orientation (SO) strategy and the machine learning strategy, which are both key approaches. They both fall under the category of "strategies." On the one hand, sentiment lexicons like SentiWordNet used [4] and [20]iSOL and eSOL , and Ml-Senticon are used [16,18,19] to extract ideas from the SO method. These methods go through the dictionary and give each word either a positive or negative rating based on their findings. SentiWordNet is the vocabulary that has been used by the scientific community as a whole to a greater extent. In spite of the fact that promising results have been reached using lexicons, certain suggestions have not produced satisfactory results since some words might have a varied meaning depending on the area in which they are used, which can be either positive or negative. Domain-dependent lexicons are one solution that has been offered [21,22,23], to cope with this problem. On the other hand, there are other techniques that are based on supervised machine learning [24][25] In these works, the classification algorithm requires a training set in order to create a model based on the various attributes of the corpus documents. In addition to this, the efficiency of the feature extraction approach that is used is directly proportional to the performance of the classification algorithms. Other proposals dissect and assess approaches based on term recurrence converse record recurrence (TF-IDF), reliance characteristics [27-33]) , POS-related feature [34] and in certain situations, a mix of these factor [35]. Both the SO and machine learning methodologies, although having been effectively used in a number of different fields, have a few drawbacks that should be considered. On the one hand, a SO method necessitates the use of linguistic resources, many of which are in short supply for languages like Spanish. On the other hand, the supervised machine learning technique calls for a big dataset that has been labeled, which is tough to come by in the research community, and the process of developing the model calls for a significant amount of time and effort to be invested [36-41]. We have taken a semantic orientation strategy in this study via the use of the SentiWordNet lexicon. This method has been successfully used in a number of other research. Regarding the fields to which the aforementioned works are directed, the majority of them have proven their methods in settings such as movies [4].

III. PROPOSED FRAMEWORK

A. Before Problem statement

We know that lots of diseases and their respective drugs are available in the market but when a new disease spread in our community health agencies create medicine to respect that disease and launch it in the market is then prescribed by physicians to the patients. Despite the medicine having undergone testing before release, users have reported experiencing certain negative effects. Numerous healthcare web gatherings offer places for patients to audit the pharmaceutical and examine their positive or negative encounters for this reason. ADR identification, post-market monitoring, absorption of patient attitudes and opinions, and other empirical applications can all benefit from mining these evaluations. The corporation, doctors, and other individuals who share the same medical issues can use these reviews to assess the efficacy of a certain medicine in treating underlying diseases. In other words, the setting of these medication evaluations offers clues as to the patient's contentment or discontent along a certain medicine in a disease.

B. Prospective fusion system for sentiment analysis

Natural Language Processing (NLP) takes sentiment analysis very seriously. Both learning-based and lexical-based methods are used in the proposed model to perform sentiment analysis on drug assessments. Each revision to the dataset is preprocessed before being classified as three class of sentiments applying lexical approaches. TF and TF-IDF were used for extracting relevant feature from the document. The collected features are then used to train and evaluate machine learning and deep learning systems. Briefly, model performance was evaluated using endpoints such as accuracy, precision, recall, and F1 score.

This study used a composite method to classify the sentiments of drug reviewers. This method involves building a model that execute opinion mining of medicine reviews utilizing a learning algorithm. and annotating the input data using a sentiment vocabulary. Drug reviews were labeled with a vocabulary-based method because the dataset used in this study did not contain any labels. SenticNet and TextBlob sentiment vocabularies derive information from pre-processed assessments. First, each word is given a rating before giving each word a related emotional score. The appraisal is at that point divided into a single among three, i.e. positive, negative or neutral, concurring to the characterized parameters as appeared within the table, accumulating the estimation score of each word within the survey. The practice and test set percentages for the labeled assessments are 75% and 25%, respectively. On this labeled data, the selected classifiers are trained and tested. Let us see all process step by step.

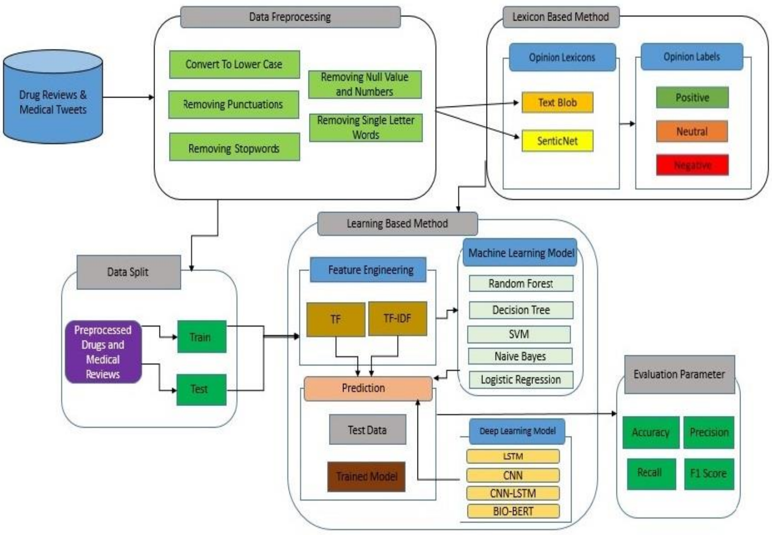


FIGURE 1: Architecture of the proposed approach.

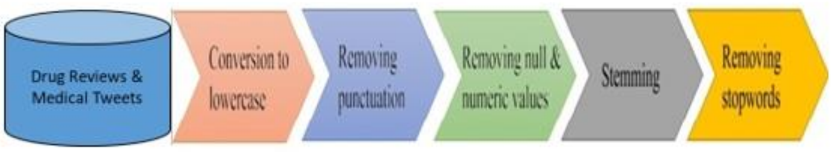


FIGURE 2. Schematic of preprocessing

B.1 Pre-Processing

B.1.1 Within the to begin with stage, known as "Pre-processing," information is organized in arrangement for the Classification Stage. Taking care of refutations, expelling accentuation, expelling single-character words, expelling purge spaces, evacuating numbers, stemming, changing over to lower case, and expelling halt words are all portion of the pre-processing methods in this ponder. The TF, TF-IDF model is then given the pre-processed dataset for analysis.

B.1.2 Handle Negations

Negative connotations in tweets need to be eliminated.

B.1.3 Convert to Lower Case

Any tweets that start with an uppercase letter will be changed to lowercase.

B.1.4 Removing Punctuations

Punctuation has been stripped out of the Tweets.

B.1.5 Removing Single Letter Words

Singular-letter terms that were deemed offensive in previous tweets have been deleted.

B.1.6 Removing Blank Spaces

The tweets that had extra spaces between the words were deleted.

B.1.7 Removing Numbers

Tweets containing irrelevant or offensive figures have been deleted.

B.1.8 Removing Stopwords

Tweets containing Stopwords have been deleted.

B.1.9 Stemming

Words may be "stemmed," or broken down to its root forms, or "lemma," by attaching a suffix, prefix, or other part of the word to the original, unaltered word's morpheme. Natural language processing and natural language understanding rely heavily on stemming since it allows for more available memory.

TABLE 1. Sample preprocessing techniques

Pre-processing Techniques	**Input**	**Output**
Convert To Lower Case	SINCE diagnosed TYPE 2 DIABETES learned food least 1 CAULIFLOWER mashed POTATOES okay enough butter salt 2 CAULIFLOWER pizza crusts utter shit 3 carrots salsa better chips salsa	since diagnosed type 2 diabetes learned food least 1 cauliflower mashed potatoes okay enough butter salt 2 cauliflower pizza crusts utter shit 3 carrots salsa better chips salsa
Removing Punctuations	In fact, having lots of fat in the <abdominal area> is strongly linked to diseases like {type 2 diabetes} and {heart disease }.	In fact, having lots of fat in the abdominal area is strongly linked to diseases like type 2 diabetes and heart disease .
Removing Single Letter Words	I have a proof that higher number pregnancy loss tied subsequent type diabetes	have a proof that higher number pregnancy loss tied subsequent type diabetes

Removing Blank Spaces	mental health important part type diabetes care take care mind take care body check mental health resource kind issue mental health diabetes matter	mental health important type diabetes care take care mind take care body check mental health resource kind issue mental health diabetes matter
Removing Numbers	see Instagram 2 photo Astra beauty cure type 2 diabetes naturally diabetes type reversed reverse diabetes naturally	see Instagram photo Astra beauty cure type diabetes naturally diabetes type reversed reverse diabetes naturally
Removing Stopwords	Manage your type 2 diabetes without ever leaving your house. With 9amhealth! https://t.co/XyTNlZofQL	Manage type 2 diabetes without ever leaving house. With 9amhealth! https://t.co/XyTNlZofQL

B.2 Lexicon-based methods

The dictionary procedure joins assumption vocabularies, which consist of a set of guidelines for categorising words from the content among three positive, negative and neutral [37]. By looking at the extremity of the sentiment-bearing words in a certain content, one may decide the extremity of a content, which serves as the premise for lexicon-based assumption categorization.

It is common to refer to a vocabulary that records the polarity values of words as a sentiment lexicon. The dictionary indicates next to each pair of words or phrases the emotion polarity score that goes with them. The (word, sentiment polarity score) format is widely used to represent the tuples of the sentiment lexicon database. Each vocabulary in a lexicon-driven method encompasses a extremity score, It may be favorable, negative, or impartial. This run is utilized to categories the content beneath thought.

$$Rp = [ps, nts, ns] \qquad (1),$$

Where Rp denotes the polarity score range. ps denotes a positive score, ns denotes a negative score, and nts denotes a neutral score.

B.2.1 Text Blob

Natural language processing (NLP) operations are carried out on content input utilizing the well-known Python-based toolkit TextBlob. It gives straightforward APIs for a assortment of NLP assignments, counting conclusion examination, thing state extraction, parts-of-speech naming, translation, and classification. TextBlob is utilized by the architects of [35] as an conclusion analyzer to expect the extremeness of the tweets. TextBlob supports both the normal dialect toolkit (NLTK) and design preparation for opinion research [38]. TextBlob's emotion dictionary has 2,918 words. Utilizing Text Blob’s sentiment investigation include, you will classify the content as either genuine data or an individual's conclusion based on its extremity and subjectivity appraisals. When TextBlob returns a tuple for a word or express, it has the taking after arrange

$$\text{Sentiment (polarity \_score, subjectivity \_score)} \qquad (2)$$

where the float values for the polarity and subjectivity scores are Rp = [+1., 1.] and Rp = [., 1.], respectively.

B.2.2 Senticnet

SenticNet is a lexical tool made for sentiment analysis at the concept level. Sentic Computing, a cutting-edge, multidisciplinary method to sentiment analysis, serves as its foundation. SenticNet has the capacity to attribute polarity and emotional information to complex ideas, such completing objectives or remembering significant occurrences, in contrast to other resources. SenticNet now provides sentiment scores for over 14, common ideas, with values ranging from -1 to 1.

TABLE 2. Super settings for sentimental vocabulary.

Parameter Name	Parameter Value
Positive	Score is greater than zero
Negative	Score is less than zero
Neutral	Score is equal to zero

The hyper-parameter settings of the opinion dictionaries utilized in this think about are appeared in Table 2.

C. Feature Engineering

To enhance the performance of prediction models on unobserved data, a feature selection approach is utilized to extract important features from the pre-processed data [38]. This technique improves the outputs of the learning model by identifying features that relate to the issue statement. The authors of the research [39] demonstrated that feature engineering led random formed out performed conventional models. Textual data may be feature engineered using a variety of techniques. In this work, we individually extract key attributes from the data for model training using TF and TF-IDF.

C.1 TF(Term Frequency)

The assessment of a word's frequency inside a certain document is referred to as a "term feature" [44]. You may calculate it by dividing the quantity of words in document D by the frequency with which the word w appears. It may be represented in maths as

$$tf(w,D) = (\text{Occurrences of } w \text{ in } D)/(\text{total number of words in } D) \qquad (3)$$

The Count Vectorizer function from Python is used in this study to extract word frequencies from the dataset.

C.2 TF-IDF(Term Frequency-Inverse Document Frequency)

Using TF-IDF, a word in a given document is quantified. A word's weight is often calculated based on how relevant it is to the given material. The relevance of a word to a document will increase with its weight score, and vice versa [34]. This is accomplished by combining the metrics of inverse document frequency (IDF) and term frequency (TF), where IDF measures a word's frequency over the whole corpus of texts while TF measures a word's relevance within a given document. The word "w" in document D has the following mathematical representation in IDF:

$$Idf(w,D)= \log(N/df(w)+1) \qquad (4)$$

where N represents the total number of documents in the corpus and df (w) is the number of documents that include the word w.

$$tf\ idf\ (w,D) = tf\ (w,D) * idf\ (w,D) \qquad (5)$$

$$tf\ idf\ (w,D) = tf\ (w,D) * \log(N/df(w)+1) \qquad (6)$$

Where: The term frequency of the word "w" in document "D" (TF(w, D)), which counts the frequency of the word "w" appearing in document "D."

The Inverse Document Frequency (IDF) of the word "w" evaluates the significance of the term over the whole document collection.

D. Machine Learning Models for Sentiment Classification

1. Naive bayes: Naive Bayes is a probabilistic classifier that makes use of the Bayes theorem and the "naive" assumption of feature independence. By figuring out the posterior probability of a sentiment class given the document's properties, it may be used for sentiment classification

$$P(\text{sentiment} \mid \text{features}) = (P(\text{features} \mid \text{sentiment}) * P(\text{sentiment})) / P(\text{features}) \qquad (7)$$

2. Support vector machines (SVM): SVM is a binary classification model that seeks to identify the optimum hyperplane for dividing data points into several groups. SVM may be trained for sentiment classification utilizing a feature representation of the text input. The foundation of SVM is the margin maximization issue, which entails optimizing the decision boundary.

3. Logistic regression: Used for binary classification issues, logistic regression is a linear model. A logistic function is used to model the likelihood that an instance belongs to a given class.

$$P\ (\text{sentiment} \mid \text{features}) = 1 / (1 + \exp(-z)) \qquad (8)$$

where z is the result of combining the characteristics' corresponding weights linearly.

4. Random forest: Using a combination of several decision trees, Random Forest is an ensemble learning technique that produces predictions. Each decision tree's foundation is a randomly selected collection of characteristics, which are then combined to yield the final forecast.

5. Decision tree: For the categorization of sentiment, decision trees are a popular machine learning paradigm. Based on the characteristics of the supplied data, they decide to use a hierarchical structure of nodes and branches. A feature or characteristic is represented by each internal node, and each branch is a potential value or result of that feature. The formulae used with decision tree models are listed below:

Information Gain: Information gain is a metric used to select the most appropriate split or attribute at each decision tree node. By dividing the data based on a certain property, it may quantify the decrease in entropy or impurity that was made possible.

$$\text{Information Gain} = \text{Entropy(parent)} - \text{Sum(weighted\_entropy(child))} \qquad (9)$$

$$\text{Entropy(parent)} = -\text{Sum}(p\_i * \log2(p\_i))\ \text{Entropy(child)} = -\text{Sum}(p\_i * \log2(p\_i)) \qquad (10)$$

where p_i represents the proportion of samples in each sentiment class in the parent or child node.

By navigating the tree based on the feature values until a leaf node (indicating a sentiment class) is reached, the decision tree may be used to predict the sentiment of fresh, unknown documents once it has been trained.

E. Deep learning models for sentiment classification

Three deep learning models, namely CNN, LSTM, and CNN-LSTM, are employed for Sentiment Analysis (SA) to analyze the provided datasets.

1. CNN model:

A multi-layer neural network that uses supervised learning is known as a convolutional neural network. The layer of convolutions, as well as the pool. The sample modules in the hidden layer of the convolutional neural network are the essential components that enable the network to accomplish its function of feature extraction. The gradient descent approach is used in the network model to achieve the goal of minimizing the loss function. This is accomplished by backward adjusting the weight parameters in the network layer by layer. In addition, the accuracy of the network may be increased by regular iterative training, which also helps cut down on the number of times the network has to be retrained. A hidden layer and a logistic regression classifier make up the top layer of the convolutional neural network. This layer is analogous to the conventional multi-layer perceptron that is used in other neural networks. A maximum pool sampling layer plus an alternative convolutional layer make up the convolutional neural network's bottom hidden layer. This layer is also known as the "hidden layer." The yield of the method of extricating highlights from the convolutional layer and the subsampling layer is the included picture, which is utilized as the input for the primary fully connected layer. The ultimate yield layer may be a classifier, and it may categorize the input picture utilizing calculated relapse, Softmax relapse, or indeed bolster vector machine in case essential. the structure of a convolutional neural network consists of three layers: the convolutional layer, the down-sampling layer, and the fully-linked layer. Each layer has several feature maps, a convolution filter is used by each feature map to extract a feature from the input, and many neurons are included inside each feature map.

Layer of convolutional filters: The original signal properties may be increased via the convolution operation, and noise can be decreased through the convolution operation, which is why the convolutional layer is used. Another essential property of the convolution operation is that it can be used to reduce noise.

2. LSTM MODEL:

An improved variant of the RNN network structure is the Long Short-Term Memory, or LSTM, network structure. The LSTM is able to successfully retain the history information in prolonged sequences by including a memory cell and three control gates. This aids in reducing the loss of previous data and gradient disappearance brought on by too much layer RNN training.

Figure 3 shows The hierarchical structure of the LSTM. To hold historical data, a memory cell is introduced to the framework. The three gates that regulate the updating, deletion, and output of historical information are input, forget, and output, respectively. Additionally, a memory cell for historical data is part of the structure. To determine how incoming vectors impact the state of the memory cell, an input gate is required. The memory cell is given the capacity to affect the outputs through the usage of the output gate. In conclusion, the memory cell has the power to either recall or forget the data it has previously stored thanks to the forget gate.

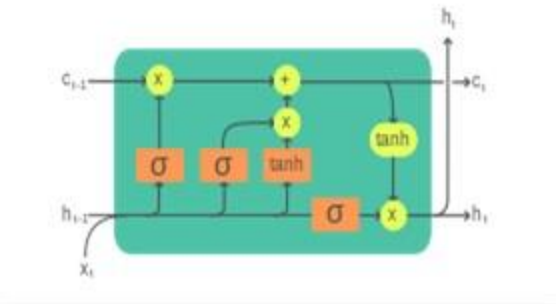


Figure. 3. Structural representation of LSTM [42]

3. CNN- LSTM Model:

The CNN-LSTM Model was the first model that I attempted to use. An initial convolution layer is part of our CNN-LSTM model combo. This layer's input will be the word embedding that was previously generated. Following that, the output will be aggregated into a smaller dimension before being sent into an LSTM layer. Following that, an LSTM layer will be applied. The idea behind this method is that the convolution layer will be able to extract local characteristics, which the LSTM layer will then utilize to learn about the text ordering of the input. This is the thinking behind this model. In actual use, the performance of this model is inferior to that of our other suggested LSTM-CNN model.

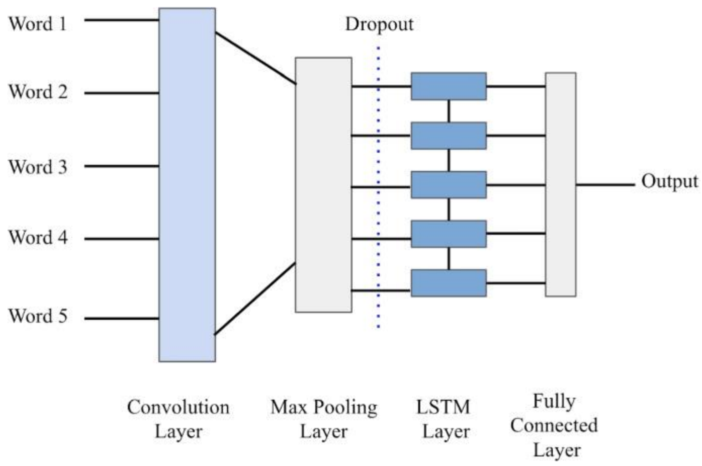


Figure. 4. Structural representation of CNN LSTM [43]

4. BIO- BERT Model:

BioBERT is a specialized variant of the BERT (Bidirectional Encoder Representations from Transformers) model that has been fine-tuned specifically for biomedical and clinical text. While BERT is a powerful language model pre-trained on a large corpus of general-domain text, BioBERT takes advantage of domain-specific biomedical literature and data to improve its performance in the biomedical field, including tasks such as sentiment analysis. To perform sentiment analysis using BioBERT, the model is first pre-trained on a large corpus of biomedical text, which could include research articles, clinical notes, or other relevant biomedical literature. During pre-training, BioBERT learns to understand the contextual relationships and meaning of words and sentences. After pre-training, BioBERT is fine-tuned on specific sentiment analysis tasks using labeled biomedical data. Fine-tuning involves training the model on a smaller dataset with sentiment labels specific to the biomedical domain. The model learns to associate the contextual representations learned during pre-training with sentiment classes (such as positive, negative, or neutral). During inference, BioBERT takes in a biomedical text as input and generates contextualized word embeddings. These embeddings capture the meaning and relationships of words within the given context. The sentiment analysis task can then be performed by feeding these embeddings into a classifier (e.g., a linear layer or a support vector machine) to predict the sentiment class associated with the input text. The advantage of using BioBERT for sentiment analysis in the biomedical domain is that it leverages the domain-specific knowledge acquired during pre-training, allowing it to better understand and interpret sentiment in biomedical texts. This can be particularly beneficial for sentiment analysis tasks involving biomedical research papers, clinical notes, patient reviews, or any other biomedical text sources

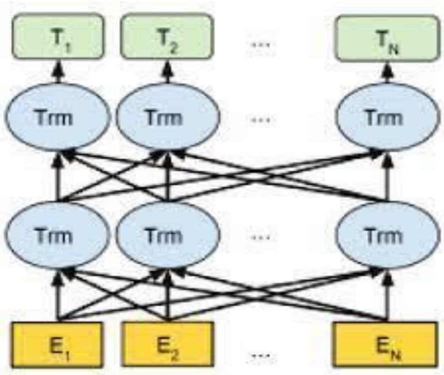


Figure. 5. Structural representation of Bio BERT[44]

F. Evaluation Parameters

After testing and training the model, it is crucial to assess how the system is doing. A categorization model can have four different results.

- True positives (TP): Situations where a good outcome is predicted but turns out to be true.
- True negatives (TN) are situations that are projected to be bad and fall under the category of negative events.
- False positives (FP) are situations that are expected to be positive but are instead in the negative category.
- False negatives (FN) are situations when the outcome is projected to be negative but is really positive.

Accuracy, precision, recall, and F1-score are the four metrics used to assess the performance of the classifiers.

1. Accuracy

The proportion of tweets that are assigned the appropriate categories, expressed as a fraction of the total number of tweets, is the standard measure of accuracy.

$$\text{Accuracy}=(TP+TN)/(TP+TN+FP+FN) \tag{11}$$

Here TP, TN, FP and FN are True Positives, True Negatives, False Positives, and False Negatives respectively.

2. Precision

The term "precision" refers to the ratio of the proportion of "positive" tweets that accurately forecast the outcome to the entire number of tweets that may be categorized as "positive."

$$P = (TP/TP+FP) \tag{12}$$

3. Recall

The term "recall" refers to the measurement that is obtained by dividing the total number of tweets that are really positive by the proportion of tweets that are classified as positive.

$$P = (TP/TP+FN) \tag{13}$$

4. F1-Score

A model's performance is thought to be best represented by the F1-score, which combines recall and accuracy. Precision and recall have a harmonic mean. An improved model's performance will be determined by the F1 score. In order to compute it,

$$F1 \text{ - } score = 2 * (PR * RC)/(PR + RC) \tag{14}$$

In this ponder, we utilized a cross breed method to classify assumption in medicate audits. It contains semantic lexicon that has been appeared to be vital for assumption investigation. The proposed strategy is utilized to construct a framework that consequently clarifies unannotated information with a estimation lexicon and classifies feelings for drugs audits and patient audits evaluation employing a prepared show.

IV. Dataset Description

For the Purpose of this Research Work two types of dataset were used First dataset is the medicate surveys dataset was gotten from the UCI machine learning store, a well-known location for benchmark datasets and the second one was tweets that were downloaded from twitter.

- Medicine Review Dataset: In this dataset patient underlying ailments at the time of drug use, the material includes patient reviews of drugs. The star rating for the drug is on a scale of 1 to 1, with 1 being the highest degree of satisfaction. As seen in Table 1, the dataset contains 161,291 data entries, each with 7 wavering. In trials, the only variable used is reviews. The evaluations show the 3,436 drug records where consumers stated their levels of enjoyment and dissatisfaction. The assessments contain user testimonies of negative effects, side effects, and a drug's positive conclusion, which displays the user's contentment.

TABLE 3. Depiction of qualities of dataset [41] for hypothesis testing.

Variables	Description
Unique ID	User Identification number
Drug Name	Drug brand name evaluated by the user
Condition	Name of the Ailment
Review	Review authored by user
Rating	1 stellar patient evaluation
Date	Date of Review arrival
Useful Count	The number of people who considered the review informative.

- Tweets: Diabetes and Type 2 diabetes tweets are used as datasets for finding sentiments. 32047 tweets downloaded from different Twitter accounts about diabetes using Python code it includes some then identified the sentiments of tweets on the basis of polarity (Positive and Negative). It includes distinct fields such as

  - Tweet. Date,
  - Tweet.id,
  - Tweet. Content,
  - Tweet. User.
  - Username

TABLE 4. Depiction of qualities of dataset 2 for hypothesis testing.

Variables	Descriptions
Tweet. Date	Date of Tweet entry
Tweet. Id	Unique Id of each user
Tweet. Content	Review written by user
Tweet. User	Twitter account holder name
Username	Twitter account user name

V. RESULT AND EVALUATION

This region gives the unpretentious components for the tests tallying the portrayal of the system utilized for tests and the hyper-parameters of distinctive classification calculations. Tests are performed utilizing Intel Center i7 8th period machine outlined with a quad-core processor, 8GB sporadic get to memory on Windows 1. arrange. Utilization is done utilizing Python tongue with Boa constrictors 3. computer program and Spyder note cushion and Google COLAB.

In this segment, we talk about the comes about of the half breed approach for each highlight building method with two diverse datasets. Results when we are using Drug Reviews as a data set.

1. Drug Reviews used as a dataset

A. Explanation of medicate surveys utilizing lexicon-based approach

The first experiment looks into the use of a lexicon-based method for sentiment labeling of 161291 unannotated medicine reviews. Drug reviews are labeled as good or negative depending on the sentiment score, with a sentiment score of denoting neutral feeling. The amount of reviews attributed to each sentiment label is displayed in Table 5.

TABLE 5. Number of instances per label for each sentiment lexicon.

Approach	Positive	Neutral	Negative
TextBlob	98255	8426	5461
SenticNet	9762	182	61869

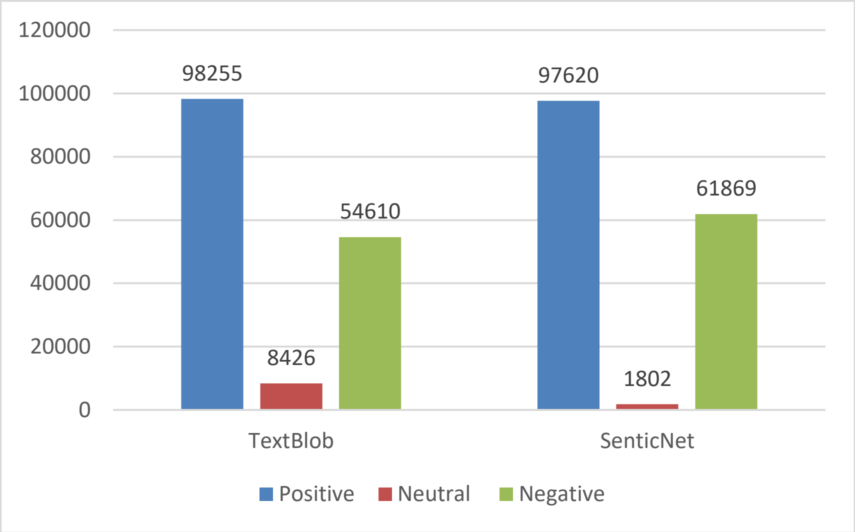


FIGURE 6. Number of instances per label corresponding to eachsentiment lexicon.

Figure 6 demonstrates that TextBlob tends to extract fewer negative feelings than SenticNet (5461). The SenticNet vocabulary has a lot of negative terms, however, the polarity scores given to each positive or negative word are rarely high, allowing the opportunity for 2 and 2 as the most typical polarity scores. Comparing TextBlob's scores to SenticNet's scores, which are in standardized order, yields 98255 positive feelings, which is more than previous sentiment lexicons were able to extract. The amount of neutral sentiment recovered by SenticNet is at its lowest, whilst positive sentiment is considerably more evenly distributed.

B. Execution examination of machine learning models with textblob opinions

Results from experiments using sentiments that were derived from TextBlob show that the suggested TF feature as compared to the hybrid feature engineering strategy Figure 7 demonstrates that TF's classification accuracy is on par with or better than that of other feature engineering strategies. Higher TF accuracy is produced by TextBlob with LR obtaining an accuracy of .95 while TF-IDF obtains an accuracy of .93.

Table 6 demonstrates that TextBlob delivers the maximum precision for TF of .95 when combined with LR. Contrarily, TF-IDF features exhibit a maximum recall of .99 when combined with TextBlob and NB. When TF features are combined with TextBlob and LR, the resulting F1 score is .96, which is the highest.

TABLE 6. Results for classification models with TextBlob sentiments

Learning Model	TF			
	A	P	R	F
TextBlob+LR	.95	.96	.97	.96
TextBlob+NB	.77	.79	.86	.82
TextBlob+DT	.89	.92	.91	.92

Learning Model	TF			
	A	P	R	F
TextBlob+RF	.88	.89	.88	.87
TextBlob+SVM	.89	.93	.92	.92
Learning Model	TF-IDF			
	A	P	R	F
TextBlob+LR	.93	.92	.97	.95
TextBlob+NB	.68	.66	.99	.79
TextBlob+DT	.89	.9	.92	.91
TextBlob+RF	.87	.89	.87	.87
TextBlob+SVM	.88	.89	.94	.92

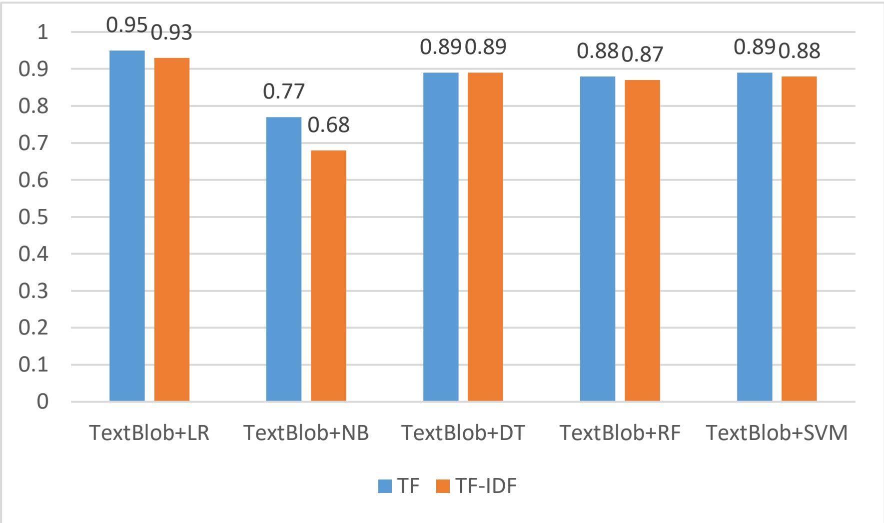


FIGURE 7. Accuracy score of classification models TextBlob sentiments

C. Execution examination of machine learning models with senticnet opinions

Opinion testing comes about extricated from SenticNet demonstrating that the proposed combined include designing strategy performs way better on normal. Figure 8 appears that the classification exactness of TF is prevalent in other include designing strategies. SenticNet with LR produces higher exactness for TF with .93 and TF-IDF with .91 Precision. Table 8 appears that SenticNet when combined with LR, produces the most elevated exactness of .94 for TF. On the other hand, the TF-IDF highlights when utilized with SenticNet and NB appeared the most noteworthy review of .98. For the F1 score, the TF highlights utilizing SenticNet with LR gave the most elevated F1 score of .95.

TABLE 7. Results for classification models with SenticNet sentiments.

Learning Model	TF			
	A	P	R	F
SenticNet+LR	.93	.94	.95	.95
SenticNet+NB	.79	.81	.86	.84
SenticNet+DT	.81	.85	.85	.85
SenticNet+RF	.87	.87	.87	.86
SenticNet+SVM	.87	.91	.89	.9
Learning Model	TF-IDF			
	A	P	R	F
SenticNet+LR	.91	.91	.94	.93
SenticNet+NB	.7	.68	.98	.8

SenticNet+DT	.82	.85	.85	.85
SenticNet+RF	.87	.87	.87	.86
SenticNet+SVM	.86	.88	.9	.89

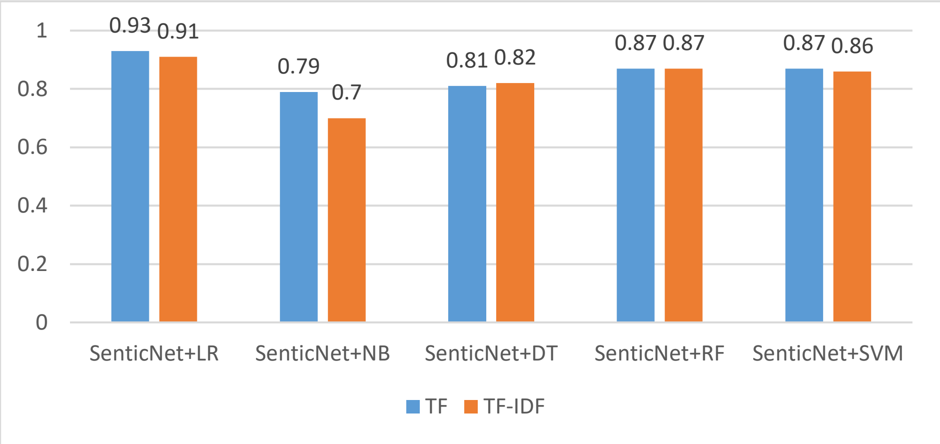


FIGURE 8. Accuracy score of classification models SenticNet sentiments

TABLE 8. Comparative performance analysis of Machine Learning classification models with SenticNet and TextBlob sentiments.

Models	TF		TF-IDF	
	TextBlob	SenticNet	TextBlob	SenticNet
LR	.95	.93	.93	.91
NB	.77	.79	.68	.7
DT	.89	.81	.89	.82
RF	.88	.87	.87	.87
SVM	.89	.87	.88	.86

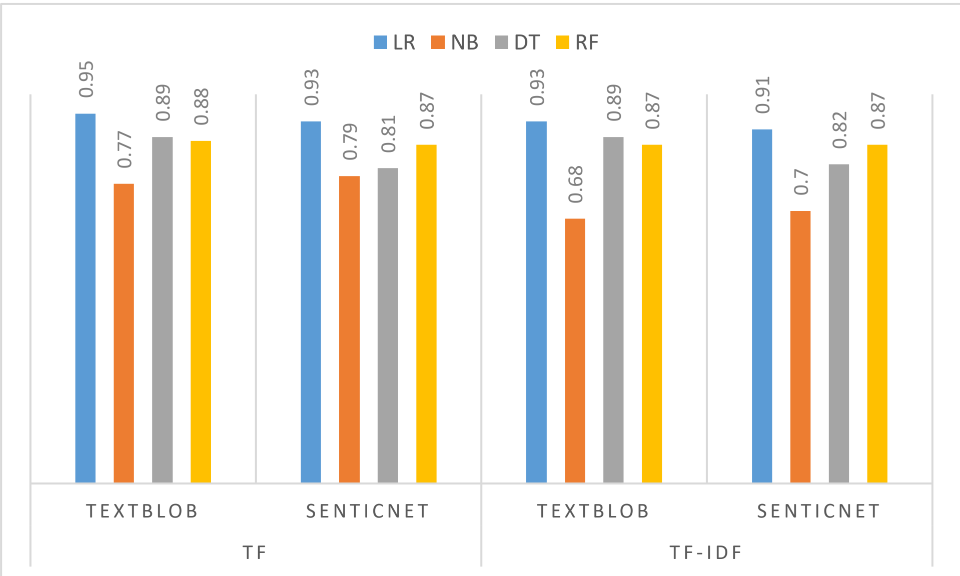


FIGURE 9. Accuracy comparison of machine Learning sentiment classification models with SenticNet and TextBlob sentiments.

D. Execution Investigation of deep learning models with SenticNet and TextBlob opinions

TABLE 9. Results for Deep Learning classification models with SenticNet sentiments.

Deep Learning Model	A	P	R	F
SenticNet+LSTM	.91	.91	.95	.93
SenticNet+CNN	.92	.86	.87	.86
SenticNet+CNN+LSTM	.93	.93	.93	.93
SenticNet+Bio Bert	.95	.96	.94	.96

TABLE 10. Results for Deep Learning classification models with TextBlob sentiments.

Deep Learning Model	A	P	R	F
TextBlob+LSTM	.96	.97	.97	.97
TextBlob+CNN	.96	.95	.93	.94
TextBlob+CNN+LSTM	.97	.96	.96	.96
TextBlob+Bio Bert	.97	.97	.96	.96

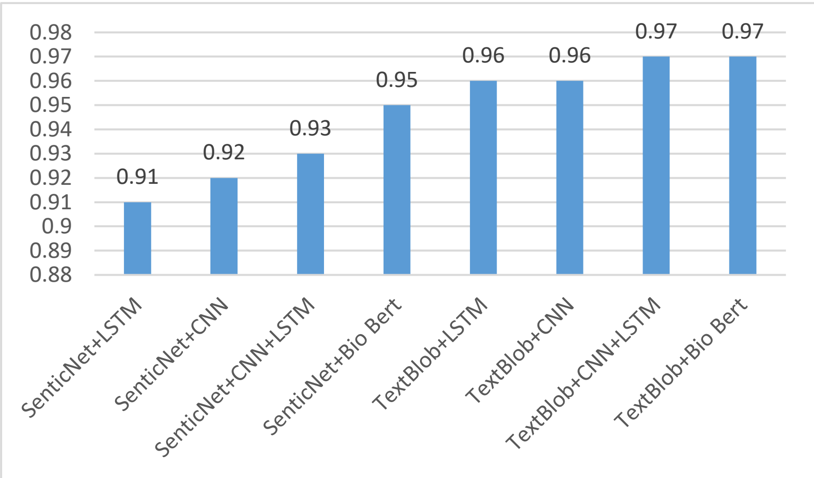


FIGURE 10. Accuracy score of Deep Learning classification models using TextBlob and SenticNet sentiments

E. Comparative Execution Investigation Of Machine Learning And Deep Learning Models With Senticnet And Textblob Opinion

According to Table 1,9,8, the greatest sentiment classification accuracy of .97 is attained using TextBlob with Bio-BERT, followed by TextBlob with CNN+LSTM for .97 and TextBlob with LR with TF for .95. It demonstrates the effectiveness of utilizing a hybrid strategy that combines learning- and lexicon-based methodologies (Machine and Deep Learning). When combined with any of the two feature engineering strategies, TextBlob leverages word sense clarification with more precision and outperforms in terms of performance regarding its pattern analyzer qualities. SenticNet, on the other hand, has underperformed other sentiment lexicons when it comes to its restriction to micro-blog length phrases.

Due of its durability, LR outperforms the learning models However, compared to LR, tree-based models like RF, DT, and NB perform poorly. If the training data and testing data differ considerably, the tree-based models will produce poor results due to their susceptibility to sampling mistakes and strong tendency to over-fit. Additionally, a correlation exists between the performance of these models and the size of the feature collection. In contrast to tree-based models, models like LR may be trained more effectively when the size of the feature set is greater than the training examples. As a result of the greater feature set, TF has demonstrated higher performance.

2. Twitter tweets used as a dataset:

A. Explanation of medicate surveys utilizing lexicon-based approach

The first experiment looks into the use of a lexicon-based method for sentiment labelling of 3247 unannotated medicine reviews. Drug reviews are labelled as good or negative depending on the sentiment score, with a sentiment score of denoting neutral feeling. The amount of reviews attributed to each sentiment label is displayed in Table 11.

TABLE 11. Number of instances per label for each sentiment lexicon.

Approach	Positive	Neutral	Negative
TextBlob	1612	5278	1667
SenticNet	2232	1395	142

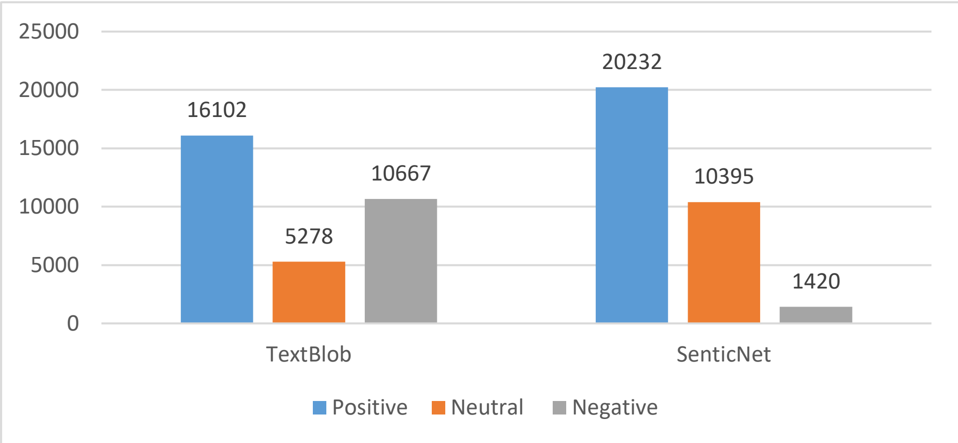


FIGURE 11. Number of instances per label corresponding to each sentiment lexicon.

Figure 11 demonstrates that SenticNet tends to extract less negative feelings than TextBlob (1667). The TextBlob vocabulary has a lot of negative terms, however the polarity scores given to each positive or negative word are rarely high, allowing opportunity for 2 and 2 as the most typical polarity scores. Comparing SenticNet's scores to TextBlob's scores, which are in standardised order, yields 2232 positive feelings, which is more than previous sentiment lexicons were able to extract. The amount of neutral sentiment recovered by TextBlob is at its lowest, whilst positive sentiment is considerably more evenly distributed.

B. Execution examination of machine learning models with TextBlob opinions

Results from experiments using sentiments that were derived from TextBlob show that the suggested hybrid feature engineering strategy performs better overall. Figure 6 demonstrates that TF's classification accuracy is on par with or better than that of other feature engineering strategies. Higher TF accuracy is produced by TextBlob with SVM, while TF-IDF obtains an accuracy of .92.

Table 15 demonstrates that TextBlob delivers the maximum precision for TF of .95 when combined with SVM. Contrarily, TF-IDF features exhibit a maximum recall of .99 when combined with TextBlob and NB. When TF features are combined with TextBlob and SVM, the resulting F1 score is .94, which is the highest.

TABLE 12. Results for classification models with TextBlob sentiments.

Learning Model	TF			
	A	P	R	F
TextBlob+LR	.89	.93	.92	.93
TextBlob+NB	.67	.63	.93	.75
TextBlob+DT	.90	.91	.92	.91
TextBlob+RF	.83	.84	.83	.81
TextBlob+SVM	.92	.95	.93	.94
Learning Model	**TF-IDF**			
	A	P	R	F
TextBlob+LR	.85	.87	.91	.89
TextBlob+NB	.58	.55	.99	.71
TextBlob+DT	.78	.85	.81	.83
TextBlob+RF	.79	.81	.79	.77
TextBlob+SVM	.90	.92	.92	.92

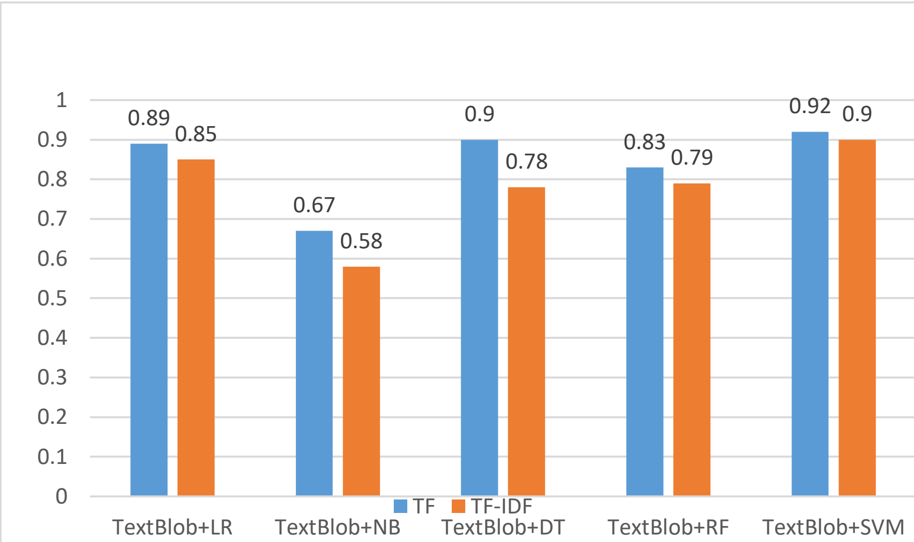


FIGURE 12. Accuracy score of classification models TextBlob sentiments

C. Execution examination of machine learning models with senticnet opinions

Exploratory comes about of extricated assumptions from SenticNet illustrated that the proposed half-breed highlight building approach performs superior on normal. Figure 13 appears that the classification precision of TF is higher than other include designing approaches. SenticNet with SVM and with LR produces higher precision for TF and TF-IDF accomplishes an precision of .87. Table 16 appears that SenticNet when joined with SVM produces the most elevated exactness of .92 for TF. On the other hand, TF-IDF highlights when utilized with SenticNet and NB appear the most noteworthy review of .99. For F1-score, TF highlights when utilized SenticNet with SVM appear the most elevated F1 score of .91.

TABLE 13. Results for classification models with SenticNet sentiments.

Learning Model	TF			
	A	P	R	F
SenticNet+LR	.87	.9	.92	.91
SenticNet+NB	.75	.79	.87	.83

SenticNet+DT	.72	.79	.78	.79
SenticNet+RF	.78	.79	.78	.78
SenticNet+SVM	.87	.92	.9	.91
Learning Model	TF-IDF			
	A	P	R	F
SenticNet+LR	.83	.82	.95	.88
SenticNet+NB	.68	.67	.99	.80
SenticNet+DT	.69	.77	.75	.76
SenticNet+RF	.78	.78	.78	.77
SenticNet+SVM	.85	.86	.93	.89

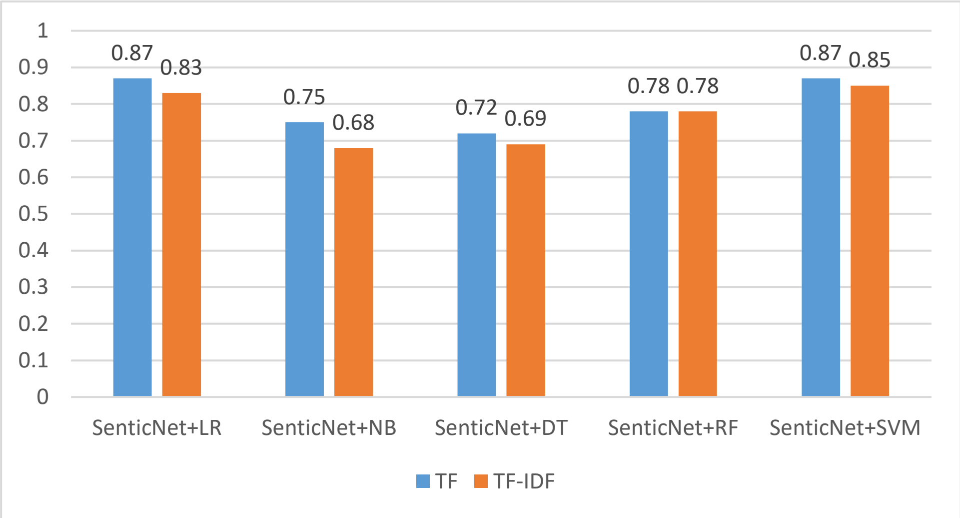


FIGURE 13. Accuracy score of classification models SenticNet sentiments

TABLE 14. Comparative performance analysis of classification models with SenticNet and TextBlob sentiments.

Models	TF		TF-IDF	
	TextBlob	SenticNet	TextBlob	SenticNet
LR	.89	.87	.85	.83
NB	.67	.75	.58	.68
DT	.90	.72	.78	.69
RF	.83	.78	.79	.78
SVM	.92	.87	.90	.85

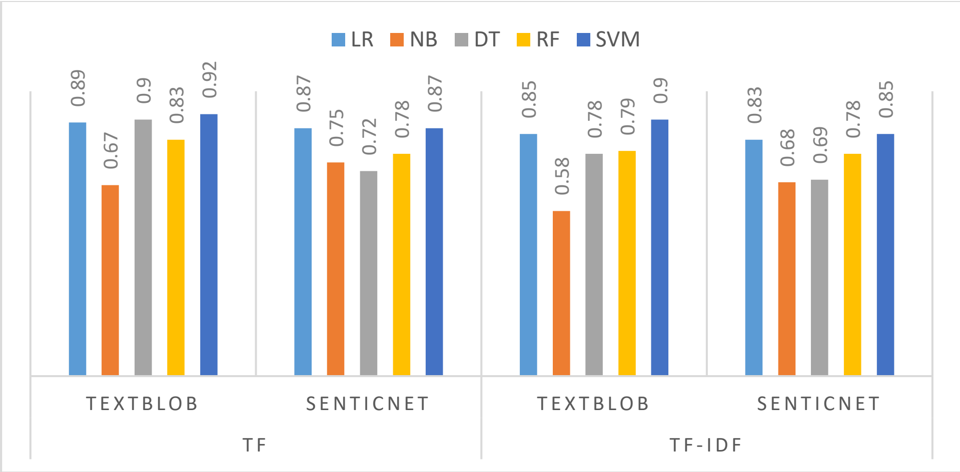


FIGURE 14. Accuracy comparison of machine Learning sentiment classification models with SenticNet and TextBlob sentiments.

D. Execution investigation of profound learning models with Senticnet and TextBlob opinions

TABLE 15. Results for Deep Learning classification models with SenticNet sentiments.

Deep Learning Model	A	P	R	F
SenticNet+LSTM	.82	.84	.93	.91
SenticNet+CNN	.85	.77	.85	.8
SenticNet+CNN+LSTM	.85	.86	.85	.85
SenticNet+Bio Bert	.87	.84	.83	.84

TABLE 16. Results for Deep Learning classification models with TextBlob sentiments.

Deep Learning Model	A	P	R	F
TextBlob+LSTM	.94	.94	.97	.95
TextBlob+CNN	.93	.9	.91	.92
TextBlob+CNN+LSTM	.92	.92	.92	.92
TextBlob+Bio Bert	.97	.96	.97	.96

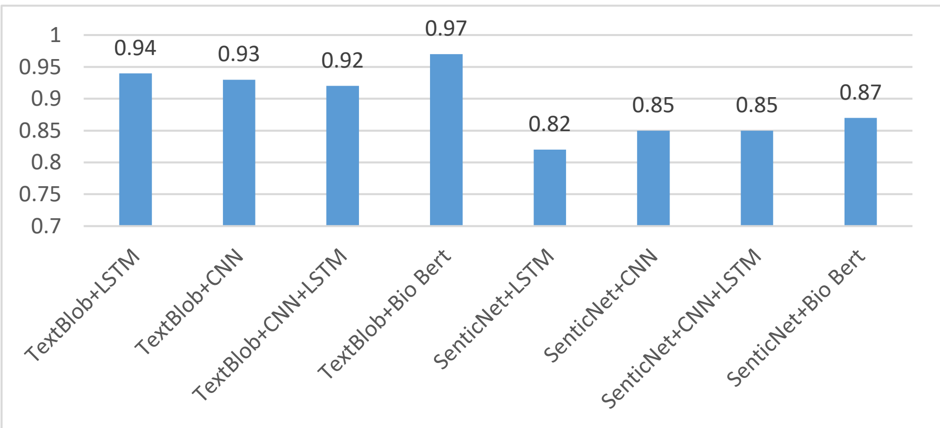
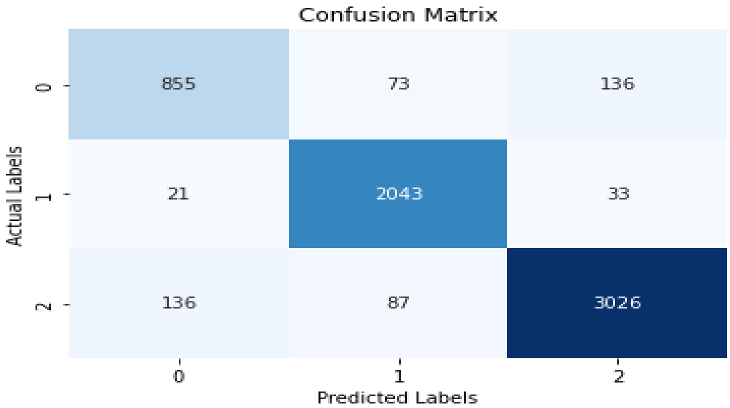


FIGURE 15. Accuracy score of Deep Learning classification models using TextBlob and SenticNet sentiments

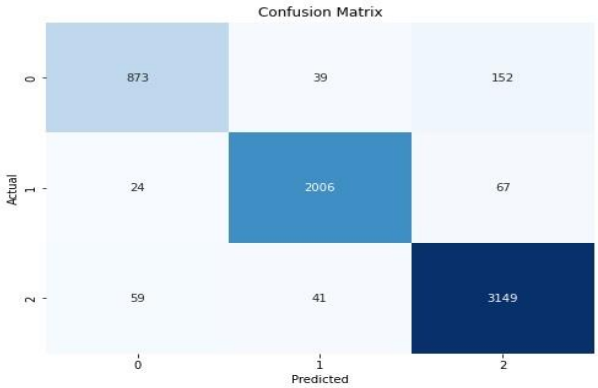
E. Comparative execution investigation of machine learning and deep learning models with senticnet and textblob opinion

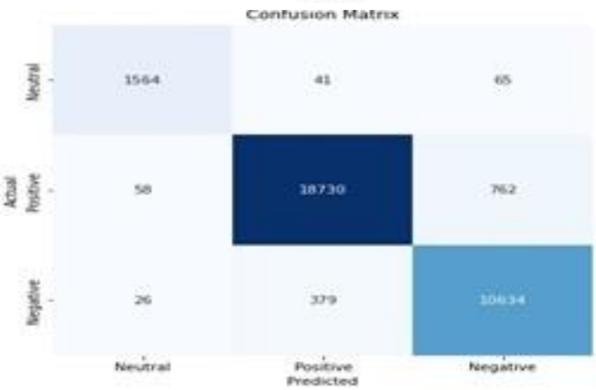
According to Table 14,15,16, the greatest sentiment classification accuracy of .97 is attained using TextBlob with Bio Bert, followed by TextBlob with LSTM for .94 and TextBlob with SVM with TF for .92. It demonstrates the effectiveness of utilizing a hybrid strategy that combines learning- and lexicon-based methodologies (Machine and Deep Learning). When combined with any of the two feature engineering strategies, TextBlob leverages word sense clarification with more precision and outperforms in terms of performance regarding its pattern analyzer qualities. SenticNet, on the other hand, has underperformed other sentiment lexicons when it comes to its restriction to micro-blog length phrases.

TextBlob+SVM(Tweets)



(TextBlob+LSTM) (Tweets) (TextBlob+CNN+LSTM)(Drug Review)





(TextBlob+BioBert) (Tweets) (TextBlob + LR)(Drug Review)

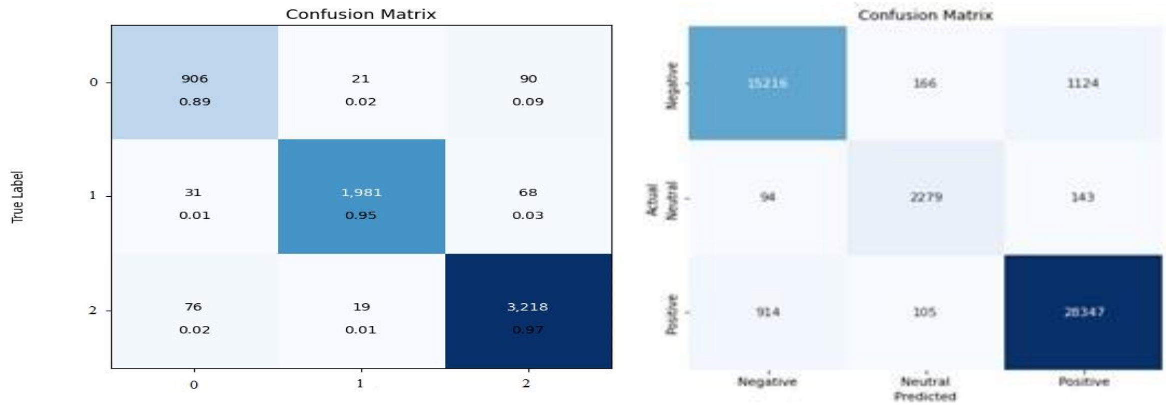


FIGURE 16. Confusion matrix for the most accurate classification models to analyze positive, neutral, and negative predictions for drug reviews and tweets.

F. Comparative investigation of proposed framework with past ponders

In order to assess the hybrid approach's effectiveness for sentiment categorization, its performance is compared to that of other cutting-edge methodologies. In order to compare different research that analyses the sentiment of medication reviews, this is done. The effectiveness of various sentiment analysis methodologies is shown in Table 17. To improve classification accuracy, the authors in [37] suggest an adaptation method with minimum pre-processing and a bag of words characteristics. Additionally, they contrast term weighting approaches to score the emotion of domain-specific phrases. The adaptive method that uses the delta TF-IDF score performs well. For drug review sentiment analysis, TF-IDF is utilized in conjunction with count vectorizer feature extraction approaches [38] For sentiment classification, a number of classifiers are employed, including ANN, LSTM, GRU, SVM, LR, and RF. In order to assign sentiment ratings to phrases linked to health, research [38] suggests a hybrid strategy that blends generic lexicons with domain-specific lexicons. Results of experiments indicate positive results. For the sentiment analysis of drug reviews, medical SWN characteristics are employed [39]. Additionally, to improve sentiment classification accuracy, the SWN lexicon's position encoding and domain enhancement are combined with a variety of machine learning models, including RF, SVM NB, and RBFN. Finally, the current analysis is contrasted with another hybrid technique suggested in [4]. Using different feature concepts with boot-strapping and corpus-based sentiment grouping, experiments are conducted on a dataset that contains the health-related text. Suggest a fusion method that merges three different lexicons: AFFIN, VADAR, and TextBlob. Various classifiers, such as LR, ADA, MLP, and ETC, are also used for sentiment classification, as well as feature engineering techniques like term frequency (TF), term frequency-inverse document frequency (TF-IDF), and union of TF and TF-IDF (TF U TF-IDF) [41]. Results in Table 17 show that the present strategy performs better than previous works.

TABLE 17. Comparative investigation of Existing approach with past ponders.

References	**Dataset**	**Classes**	**Feature**	**Method**	**Results**			
					A	**P**	**R**	**F**
[37]	Drug Review (215063)	Multi	TF-IDF, Count vectorizer	ANN,LSTM,GRU,SVM, LR,RF	0.93	0.95	0.95	0.9
[38]	Drug Review (50000)	Multi	Uni and bigram	Lexicon combination and information gain	0.91	0.76	0.53	0.62
[39]	Drug Review (5600)	Binary	SWN and Position encoding	RF,SVM,NB,RBFN		0.65	0.58	0.62

[40]	Drug Review (26060)	Multi	Uni , bigram and trigram	Corpus based sentiment classification		0.89	0.79	0.83
Proposed	Drug Review (161297)	Multi	TF and TF-IDF	TextBlob+BioBert, TextBlob+CNN-LSTM	0.97, 0.97	0.97, 0.96	0.96, 0.97	0.98, 0.96
Proposed	Medical Review (32047)	Multi	TF and TF-IDF	Textblob+BioBert, TextBlob+LSTM	0.97, 0.94	0.97, 0.93	0.97, 0.94	0.97, 0.93

VI. Conclusion

In order to analyze users' feelings, this work tackles the issues of the need for manual annotation in a learning-based method and the domain specificity of sentiment lexicons. For labeling and classification, it combines lexicon-based and learning-based methodologies to achieve this. For two datasets—drug reviews and patient tweets—TextBlob and SenticNet are tested as data annotation methods. According to experimental analysis, TextBlob often produces superior outcomes when annotating medication reviews. There are two established methods for feature engineering, TF, and TF-IDF. In order to classify sentiment, five machine learning models—LSTM, SVM, NB, RT, and DT—as well as four deep learning models—LSTM, CNN, CNN-LSTM, and Bio BERT—are utilized. Their effectiveness is assessed using a variety of annotation strategies and feature engineering techniques. the results indicate that SVM and LR machine learning models perform well at the time performed on term frequency and term frequency Inverse document frequency with TextBlob lexicons, whereas CNN-LSTM and Bio Bert Model provide the most accurate results when compared to all other machine and deep learning models in a deep learning model. The recommended strategy in this study's empirical findings improved accuracy outcomes over earlier methods for drug review sentiment analysis by 4%. Additionally, tests on datasets from different areas demonstrate the usefulness of the suggested strategy. Sarcasm and negation detection may be a future focus. The sentiment categorization and other lexicons we can employ, such MediSent, UMLS, etc., may be impacted by the review data's inclusion of negative feelings that are utilized in a positive context.

Declaration of Interest Statement

There will be no declaration of Interest.

Data Availability

Data is available from corresponding author.

References

[1] H. Osop, R. Hasan, CS. Lee, CY. Neo, CK. Foo and A. Saurabh, "Diabetweets: Analysis of Tweets for Health-Related Information," *HCI International 2020–Late Breaking Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II*, vol. 22, pp. 500-508, 2020.

[2] E. Gabarron, E. Dorronzoro, O. Rivera-Romero and R .Wynn, "Diabetes on Twitter: a sentiment analysis," *Journal of Diabetes Science and Technology*, vol. 13, no. 3, pp. 439-444, 2019.

[3] J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. Angel Rodriguez-Garcia and Rafael Valencia-Garcia, "Sentiment analysis on tweets about diabetes: an aspect-level approach," *Computational and Mathematical Methods in Medicine*, 2017. Available: https://www.hindawi.com/journals/cmmm/2017/5140631/

[4] A. Ahne, V. Khetan, X. Tannier, Md I. H. Rizvi, T. Czernichow, F. Orchard, C. Bour, A. Fano, and G. Fagherazzi,"Identifying causal relations in tweets using deep learning: Use case on diabetes-related tweets from 2017-2021," *arXiv preprint arXiv:2111.01225*, 2021.

[5] E. Clark, T. James, C.Jones, A .Alapati, P. Ukandu, C.M. Danforth and P. Dodds, "A sentiment analysis of breast cancer treatment experiences and healthcare perceptions across Twitter," *arXiv preprint arXiv:1805.09959*, 2018.

[6] H. Singh and V. Dhir, "Comparative Study of Sentiment Analysis of Type 1 and Type 2 Diabetic Patients Using Apache Flume and Hive," *Webology*, vol. 19, no. 1, pp. 4035-4054, 2022.

[7] A .Nazir, Y. Rao, L .Wu, and L .Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive

survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 845-863, 2020.

[8] B.N. Harish, K. Reena, S. Kumar, and J. Zhong, "How Much do You Care?: Mining and Analysis of Tweets Pertaining to Health Issues," *SoutheastCon 2018*, pp. 1-8, IEEE.

[9] M. Wankhade, A.C.S. Rao, and C .Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731-5780, 2022.

[10] S. Poria, D. Hazarika, N. Majumder, and R .Mihalcea, "Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research," *IEEE Transactions on Affective Computing*, 2020.

[11] M . Birjali, M. Kasri, and A .Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, 107134.

[12] P.K .Jain, R .Pamula and G. Srivastava , "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews," *Computer Science Review*, vol. 41, 100413.

[13] M.F.Gavilanes, T. A.Lopez , J. J. Martínez , E.C.Montenegro, and F.J.G. Castano, "Unsupervised method for sentiment analysis in online texts," *Expert Systems with Applications*, vol. 58, pp. 57-75, 2016.

[14] P. Nakov, S.Rosenthal, S. Kiritchenko, S.M. Mohammad, Z. Kozareva, and A. Ritter, "Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts," *Language Resources and Evaluation*, vol. 50, pp. 35-65, 2016.

[15] A. Montejo-Raez, E. Martinez-Camara, M.T Martin-Valdivia, and A . L Urena-Lopez,"Ranked wordnet graph for sentiment polarity classification in Twitter," *Computer Speech & Language*, vol. 28, no. 1, pp. 93-107, 2014.

[16] M.D .Molina-Gonzalez, E.Martínez-Camara, M.T. Martin-Valdivia and J.M .Perea-Ortega,"Semantic orientation for polarity classification in Spanish reviews," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7250-7257, 2013.

[17] N . Ofek, C. Caragea, L. Rokach, P .Biyani, P. Mitra, J .Yen, K. Portier, and G .Greer,"Improving sentiment analysis in an online cancer survivor community using dynamic sentiment lexicon," in *2013 International Conference on Social Intelligence and Technology*, pp. 109-113, IEEE, 2013.

[18] M.D. Molina-Gonzalez, "Generacion de Recursos para An ´ alisis ´ de Opiniones en Espanol," Universidad de Jaen, Ja ´ en, Spain, 2016.

[19] S. Park, W. Lee, and I.C. Moon, "Efficient extraction of domain specific sentiment lexicon with active learning," *Pattern Recognition Letters*, vol. 56, pp. 38-44, 2015.

[20] S. Deng, A.P. Sinha, and H. Zhao, "Adapting sentiment lexicons to domain-specific social media texts," *Decision Support Systems*, vol. 94, pp. 65-76, 2017.

[21] M.R. Saleh, M.T. Martin-Valdivia, A. Montejo-Ráez, and L.A. Ureña-López, "Experiments with SVM to classify opinions in different domains," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14799-14804, 2011.

[22] M. del Pilar Salas-Zarate, M.A. Paredes-Valverde, J. Limon, D.A. Tlapa and Y.A. Báez, "Sentiment Classification of Spanish Reviews: An Approach based on Feature Selection and Machine Learning Methods," *J. Univers. Comput. Sci.*, vol. 22, no. 5, pp. 691-708, 2016.

[23] I. Habernal, T. Ptáček, and J. Steinberger , "Supervised sentiment analysis in Czech social media," *Information Processing & Management*, vol. 50, no. 5, pp. 693-707, 2014.

[24] T. Ali, D. Schramm, M. Sokolova, and D. Inkpen, "Can i hear you? sentiment analysis on medical forums," in *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 667-673, 2013.

[25] P. Biyani, C . Caragea, P.Mitra, C. Zhou, J . Yen, G.E. Greer, and K. Portier, "Co-training over Domain independent and Domain-dependent features for sentiment analysis of an online cancer support community," in *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*, pp. 413–417, ACM, Sydney, Australia, August 2013.

[26] P. Smith and M. Lee, "Cross-discourse development of supervised sentiment analysis in the clinical domain," in *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 79–83, Stroudsburg, Pa, USA, 2012.

[27] V. Bobicev and M. Sokolova,"Sentiment analysis in health related forums," in *Proceedings of the 8th International Conference on Microelectronics and Computer Science*, pp. 212–216, Chisinau, Moldova, October 2014.

[28] M.P.Salas-Zarate, E. Lopez, R. García, N. Gilles, A. Almela, and G. Hernandez, "A study on LIWC categories for

opinion mining in Spanish reviews," *Journal of Information Science*, vol. 4, no. 6, pp. 749– 76, 2014.

[29] H. Saif, Y . He, and H. Alani, "Alleviating data sparsity for Twitter sentiment analysis," in *Proceedings of the 2nd Workshop on Making Sense of Microposts (#MSM '12): Big Things Come in Small Packages at the 21st International Conference on the World Wide Web (WWW '12)*, pp. 2–9, Lyon, France, 2012.

[30] B. Agarwal, S. Poria, N. Mittal ,A. Gelbukh and A. Hussain, "Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach," *Cognitive Computation*, vol. 7, no. 4, pp. 487–499, 2015.

[31] I. Habernal , T. Ptacek and J. Steinberger, "Reprint of "supervised sentiment analysis in Czech social media"," *Information Processing and Management*, vol. 51, no. 4, pp. 532–546, 2015.

[32] B. Agarwal, and N. Mittal "Prominent feature extraction for review analysis: an empirical study," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 3, pp. 485–498, 2014.

[33] H. Sharif , F. Zaffar, A. Abbasi, and D. Zimbra, "Detecting adverse drug reactions using a sentiment classification framework," in *Proceedings of the 6th ASE International Conference on Social Computing (SocialCom '14)*, Stanford, Calif, USA, May 2014.

[34] J.C Na, WYM. Kyaing, C.S.G. Khoo, S. Foo, Y.K. Chang and Y.L .Theng, "Sentiment classification of drug reviews using a rule-based linguistic approach," in *The Outreach of Digital Libraries: A Globalized Resource Network*, pp. 189–198, Springer, Berlin, Germany, 2012.

[35] N.M. Alharbi, N.S. Alghamdi, E.H. Alkhammash, and J.F. Al Amri, "Evaluation of sentiment analysis via word embedding and RNN variants for Amazon online reviews," *Mathematical Problems in Engineering*, 2021.

[36] X.L. Fu, J. Wu, J. Chen, and S. Liu, "Attribute-Sentiment pair correlation model based on online user reviews," *Journal of Sensors*, 2019.

[37] S. Vijayaraghavan and D. Basu, "Sentiment analysis in drug reviews using supervised machine learning algorithms," arXiv:23.11643, 2022.

[38] L. Goeuriot, J.C. Na, W.Y. Min. Kyaing, C .Khoo, Y.K. Chang, Y.L. Theng,and J.J. Kim, "Sentiment lexicons for health-related opinion mining," in *Proc. 2nd ACM SIGHIT Symp. Int. Health Informat. (IHI)*, 2012, pp. 219-226.

[39] M.Z. Asghar, S. Ahmad, M. Qasim, S.R. Zahra and F.M. Kundi, "Senti-Health: Creating health-related sentiment lexicon using a hybrid approach," *SpringerPlus*, vol. 5, no. 1, p. 1139, Dec. 2016.

[40] S. Liu and I. Lee "Extracting features with medical sentiment lexicon and position encoding for drug reviews," *Health Inf. Sci. Syst.*, vol. 7, no. 1, p. 11, Dec. 2019.

[41] E. Saad, S. Din, R. Jamil, F. Rustam, A. Mehmood, I. Ashraf, and G.S. Choi, "Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums," *IEEE Access*, vol. 9, pp. 85721-85737, 2021.

[42] https://en.wikipedia.org/wiki/Long_short-term_memory#/media/File:LSTM_Cell.svg

[43] https://link.springer.com/article/10.1007/s11227-021-03838-w

[44] J. Lee, W. Yoon,., S. Kim, D.Kim,, S. Kim, H. C. So and J. Kang," BioBERT: a pre-trained biomedical language representation model for biomedical text mining". *Bioinformatics*, vol .4. pp. 1234-1240. doi: 10.1093/bioinformatics/btz682