MACHINE INTELLIGENCE RESEARCH
ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

EFFICIENT ANALYSIS OF SENTIMENTS IN TWITTER DATA BY USING STACKED BI-DIRECTIONAL LSTM

M.Srisankar[1] Dr.K.P.Lochanambal[2]

[1]Research Scholar, Department of Computer Science, Government Arts College, Udumalpet, Bharathiar University, Tamilnadu, India. Shankar276@gmail.com
[2]Assistant Professor, Department of Computer Science, Government Arts College, Udumalpet, Bharathiar University, Tamilnadu, India. kplmca@rediffmail.com

Abstract
With increase in social media popularity and different platforms permitting people to show opinion on diverse subjects, Sentiment Analysis (SA) as well as opinion mining is becoming a subject which attracts attention of researchers. SA has gained much popularity amid people with varying interests as well as motivations. In this paper, Sentiment140 dataset with tweets extracted using Twitter API is used. Pre-processing is performed using tokenisation by stemming and lemmatization. Extraction of features is carried out using Term Frequency-Inverse Document Frequency (TF-IDF), Word to Vector (Word2Vec) and word embedding using BERT. Tweets are categorizedinto positive and negative using Long Short-Term Memory (LSTM), Bi-directional LSTM (Bi-LSTM) and proposed Stacked Bi-LSTM (SBi-LSTM)model.SBi-LSTM offers better performance due to its modest structure with stacked LSTM layers. SBi-LSTM offers improved results based onAccuracy, Recall, Precision and F1-Score.
Keywords: Twitter Sentiment Analysis, Tokenisation, TF-IDF, Word2Vec, BERT, LSTM, Bi-directional LSTM, Stacked Bi-directional LSTM

1. Introduction

Microblogging is predominant among Internet fraternity.Microblogging websites are social media sites in which users post small and frequent posts. People post their opinions as well as sentiments about products, institutions, news, etc., Twitter is a highly used microblog, wherein clients post their opinions. It permits users to post about 140 characters. This enables users to be brief about their opinions and includes a huge collection of sentiments to investigate. Itoffers developer-based streaming Application Programming Interface (API) for retrieving data,thus permitting analysts to search tweets from several users in real-time.With rapid increase in social applications, people use these platforms to post their opinions related to daily issues. Recently, researchers are seen to highly focus on analysing opinions on diverse topics.Collecting and analysing peoples' opinions on a product, public services etc.,have become vital. Tweets represent raw data. A method which automatically classifies tweets based on positive or negativesentiments is essential.Sentiment Analysis (SA) plays a dominant role in forecasting, mining opinions etc., SA or opinion mining aims at determining opinions related to objects or subjects under discussion. Initially, sentiments are analysed based on lengthy texts. It is used in pre-crime and post-crime analysis. It aids customersto get feedback about product or services before purchasing the same. Companies use SA to get customers' opinion about products, such that customer satisfaction may be analysed and their product can be improved. Customers discuss,

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

comment as well as criticize several topics. They write reviews and recommendations and user-produced data includesvalued information pertaining to product, people, event etc.SAis becoming popular in computational linguistics due to inclusion of sentiment-based information in social websites, online forums as well as blogs. It seems to be a challenge in areas ofcomputational linguistics, Natural Language Processing (NLP) as well astext analytics. Automated techniques are essential for analysing and mining user-generated content as crowd-sourced mining as well as analysis ischallenging. In this paper, tweets are analysed and classified. Initially, the posts are taken and pre-processed using tokenisation through stemming and lemmatization. Features are extracted using Term Frequency-Inverse Document Frequency (TF-IDF), Word to Vector (Word2Vec) and word embedding using BERT. Tweets are classified by using Long Short-Term Memory (LSTM), Bi-directional LSTM (Bi-LSTM) and proposed Stacked Bi-LSTM (SBi-LSTM).

2. Related Work

Marketers perform SA for analysing services, business, public opinions or assess customer satisfaction. SA is used for collecting substantial feedback on challenges associated with new products. The stages include collection of text, cleaning, pre-processing along with feature extraction. Noise, missing data, tags, emoticons and punctuations are removed from data. By focussing on twitter posts, domain-based data can be extracted. Data can be categorised using SA methods.

2.1 *Pre-processing*

It is noted that the classification accuracy improves once text is pre-processed and dimensionality is reduced. Text pre-processing facilitates predictive analysis.Arora &Kansalet al [2019]have focussed on processing information in Twitter dataset and determining message polarity. Text involving unstructured data is normalised using deep Convolutional character level embedding (Conv-char-Emb) Neural Network (NN) model. Noisy sentences are processed for identifying sentiments. Memory space in embedded learning at word-level is also handled. Initial pre-processing for carrying out normalization of text includes the ensuing steps namely, tokenization, identification and replacement of Out Of Vocabulary (OOV), lemmatization as well as stemming. Embedding based on characters in Convolutional Neural Network (CNN) is effective for performing SA which employs reduced number of factors in representing features. The proposed scheme normalises and classifies sentiments in unstructured sentences. Patel &Passiet al [2020]have analysed Twitter data of World Cup Soccer at Brazil in 2014 for identifying sentiments using Machine Learning (ML) schemes. By filtering and analysing data using NLP schemes, sentiment polarity is determined based on emotion-based words identified in tweets. Dataset is standardized using ML algorithms and prepared using NLP involving word tokenization, Name Entity Recognition (NER), Part-of-Speech (PoS) tagger along with parser for extracting emotions from text. The algorithm extracts words representing emotions by using WordNet with PoS for a word with meaning in present context, and allocates polarity by using SentiWordNet dictionary or lexicon-based scheme. Polarity is analysed by using K-Nearest Neighbour (KNN), Random Forest (RF), Naïve Bayes (NB)and Support Vector Machine (SVM)algorithms. NB offers enhanced accuracy and RF gives best Area under Curve (AUC).Ajitha et al [2021]have used ML and NLP schemes for performing mining sentimentsin

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

tweets. Text is classified using ML-based fusion. Feelings are graded based on lexicons. Bag-of-Words (BOWs) is used for modelling text in SA. Data overload is highly handled.Garg & Sharma et al [2022]have focussed on pre-processing of text and their influence on dataset. The proposed method may be used in analysing market and customer behaviour, polling as well as monitoring brands. Data is pre-processed by tokenization, PoS, lemmatization, normalization and word stemming, where network of words is formed. Relevant events as well as sub-events are mapped to associated words. Stemming algorithm is implemented on clean text.In the scheme proposed by Ningsih et al [2023], while pre-processing data, the labelled data is cleaned. Data gathered from twitter includesinessential words or characters like punctuation marks, numbers, symbols, whitespace etc., which interfere with data analysis. Once data is cleaned, sentimentsare processed in stages including case folding, removal of stop words and stemming. Duplicate data as well as needless links are to be removed. Sentiments are included for processing. Dataset with sentiments is transformed into dataframe includingdifferent features. The dataset is labelled with Valence Aware Dictionary for Social Reasoning (VADER) scheme, following which ensemble learning including Logistic Regression (LR), RF, Decision Tree (DT) and SVM algorithms is included.Soesanto et al [2023]has dealt with pre-processing of text for training and testing dataset taken from Kaggle to understand the variations between outcomes before and after applying pre-processing. Text is pre-processed by removing punctuations, numbers, stop words as well as white text. Further, words are tokenised and lemmatized.Dataset may include variables which hinder processing of retweets, links as well as mentions. Such noises are removed using Regular Expression (Regex). This has an impact on the outcome of SA.

2.2 *Feature Extraction*

Proposing accurate, robust and efficient feature extraction is challenging. Designingeffective as well as reliable feature set which yields increased categorization accuracy is significant in SA.Ahuja et al [2019]have examined the influence of TF-IDF at word level along with N-Gram features in SS-Tweet dataset. It is seen that TF-IDF offers better performance when compared to N-gram features. DT, SVM, KNN, RF, LR, NB based on Accuracy, Precision, F-Score and Recall.Habib et al [2021]have propounded ML method for classifying data into positive and negative sentiments. Diverse pre-processing schemes are applied for cleaning tweets, and numerous feature extraction schemes are used for extracting and reducing dimensions of tweets' feature vector. Supervised ML models including LR, NB, and SVM are used. It is seen that LR offers better results in contrast to other models.Chiny et al [2021]have propounded hybrid SA model based on LSTM, SA lexicon that is rule-based and TF-IDF. The models are combined to form a binary model. The algorithms including LR, K-NN, RF, SVM and NB are implemented. The model is trained using limited data from IMDB dataset. The proposed scheme offers improved Accuracy as well as F1-score in contrast to scores stored by input models.The proposed model transfers knowledge to deal withsentiments in Twitter dataset of US Airlines.Singh et al [2022] have proposed a ML-based scheme to perform SA on text available insocial media. The proposed scheme is split into separate phases. Initially, pre-processing is carried out to filter as well as refine text. Secondly, TF-IDF is used for feature extraction. Further, extracted features are involved in making predictions. Numerous ML methods are used in analysis as well as classification. Performance is analysed based on accuracy, precision, F1-score and recall. SVM

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

offers improved results.Kaur & Sharma et al [2023]have performed feature extraction by using Review related features (RRF) along with Aspect related features (ARF). Numerical features of RRF as well as ARF are included in Hybrid Feature Vector (HFV). A robust and efficientSAframework employing integrated feature extraction method is proposed. RRF features are extracted by employing diverse schemes to get polarity of each term within pre-processed text which includes emoticons as well asnegations. Aspect terms with polarities are extracted using ARF. Sarcastic as well asunclear reviews are expressed using ARF. The outcomes of both ARF and RRF are integrated and pre-processed review is represented as HFV.Parveen et al [2023]has proposed a scheme in which features are extracted using Term Weighting-based Feature Extraction (TW-FE) method called Log Term Frequency-based Modified Inverse Class Frequency (LTF-MICF) based on term weights. The proposed scheme includes 2TW schemes namely,LTF and MICF. The terms which occur frequently are calledTerm Frequency (TF) which is not adequateas these terms possess more weight in document. To handle this, hybrid feature extraction scheme is proposed. Hence, TF is combined with MICF to get an efficient scheme.

2.3 *Classification*

PresentNN models are inefficient in capturing sentiment messages from comparativelyextended time-steps. To handle this, Rao et al [2018]have propounded SR-LSTM which is a NN model that involves 2 hidden layers. The initial layer takes sentence vectors to showsentence semantics usingLSTM network. In second layer, sentence relations are encoded in document representation. The proposed method cleans the datasets and removes sentences that have least emotional polarity to have improved input.SA is a sequence model that decodes input sequences into particular length vector. In case vector length is less, input will be lost, and there are chances for misjudging the text. To deal with this, Li et al [2020]have proposed Self-Attention Mechanism and multi-channel Features based Bi-LSTM (SAMF-BiLSTM) model. It models linguistic knowledge and resources forSA tasks to establishdiverse feature channels, and employs self-attention scheme forimproving sentiment information. The model shows the associationamid target as well as sentiment polarity words, and does not trust on sentiment lexicons that are organized manually. SAMF-BiLSTM-D model is proposed for classifying text at documentlevel.

Soni&Mathur et al [2022]have propounded a hybrid model using LSTM and Encoder for SA based on diverse factors, contexts as well as aspects when analysing word or phrase in tweet. Encoder attention method is used for determiningsignificanceof aspects. To get related context, Paragraph2vec is used for facilitating the determination of contextual meaning. Features of Paragraph2vec and encoder output are jointly fed into LSTM and classified.

Iparraguirre-Villanueva et al [2023]have analysed the people's feelings who have published their opinions on Monkeypox. Hybrid-based model is built using CNN as well as LSTM,and prediction accuracy is determined. The polarity of feelings is shown using confusion matrix of CNN-LSTM, wherein people expressed positive, negative and fearful feelings about the disease.

3. PRE-PROCESSING OF TWEETS

The tweets are pre-processed by tokenising them through stemming and lemmatization.

3.1 *Tokenisation*

The customer reviews are pre-processed by performing tokenisation before performing feature extraction. Text-processing is useful for handling sparsity and normalising vocabulary. It reduces

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

redundancy, as word stem and modified words have similarmeaning; it permits NLP models to acquire links amid modified words as well as word stems that helps the model to comprehend the usage in comparable contexts. Tokenisation includes Stemming and Lemmatization.

Stemming

Stemming is used for reducing a transformed word to word stem. Itdeals witheliminating prefixes as well as suffixes from words. Diverse algorithms are involved in stemming. BoWsinclude words which may beidentical or similar but with morphological variants. By implementing these algorithms, words get reduced to root, permitting documents to be shownas stems rather than original words. Stemming aids incircumventing mismatches whichreduces recall while retrieving information. Porter's stemmer is commonly used as itapplies a collection of rules and iteratively removes suffixes. It has a well-documented collection of constraints. The algorithm is split into different steps which are linearly implementedtill final word is obtained. Stemming algorithms consider a set of common prefixes and suffixes seen in modified words and removesthe word end or beginning. This leads to word stems which are not real words.It offers improved model performance, grouping of comparable words and is easier to analyse as well as understand.

LemmatizationLemmatization is a significant pre-processing step in text mining and used in NLP. It resembles stemming as both decrease a word to 'stem' in case of stemming and to 'lemma' in case of lemmatization. It aims at removing inflectional ends and givingwords' dictionary forms called lemma as output. It is a scheme used for reducingmodified words to root. It shows the process of identifying a transformed word's 'lemma' based on proposed meaning. It involves morphological and vocabulary analysis for offering words to dictionary form. Words are converted to lemma. In contrast to stemming, it depends on precisely finding intended Po Sand word meaning based on the context. The modified word falls in a sentence andhuger context adjoining the sentence likenearby sentences or whole document. It is more accurate and involves less time.

4. EXTRACTION OF FEATURES FROM PRE-PROCESSED TWEETS

Features are determined from pre-processed tweets using TF-IDF,Word2Vec and BERT.

4.1 *Term Frequency-Inverse Document Frequency (TF-IDF)*

TF-IDF is employed for finding word's significancewithin a document. The term's frequency (t) is given by ratio of amount of times it is seen in a text to total words in document. IDF aids in findingthe relevance of a term. Words like 'is', 'an' and 'and' are extensivelyused but haveless meaning.TF-IDF shows suitability of a word found in a corpus in contrast to the text. Meaning increases proportionatelybased on amountof times a word is seen in text but is balanced by its frequency in corpus.'TF' shows the amount of instances of particular word (t) in document (d).It isappropriate when a word is seen in rational text. As ordering of items seem to be unsubstantial, vectorscan be used to define text in Bag-of-term models. For every term, an entry with TF exists.The term weight is proportional to TF.

$$\mathrm{TF}_{t,d} = \frac{\text{Number of times } 't' \text{ appears in } 'd'}{\text{Number of words in } 'd'} \quad (1)$$

'Document Frequency (DF)' determines text meaning similar to TF in complete corpus collection. The difference is that, TF represents frequency of 't', whereas 'DF'shows the amount of occurrences in Document Set (DS) of 't'.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

$$DF_t = \text{Occurrence of } 't' \text{ in } 'd' \quad (2)$$

'IDF' checksthe appropriateness ofa word. The appropriate records which fit the demand are sought. As 'TF' considers the terms to be equally important, they can be used for measuring term weight. 'DF_t' is determined by counting the amount of documents which include the term.

$$DF_t = DS_t \quad (3)$$

where

DF_t - DFof 't'

DS_t- Number of 'd's containing 't'

'TF' represents the amount of occurrences of a term in a document. DFshows thequantity of documents with term depending on whole corpus. 'IDF' of a word represents quantity of documents separated by text frequency.

$$IDF_t = \frac{DS}{DF_t} = \frac{DS}{DS_t} \quad (4)$$

Common words arefound to be less significant. By taking logarithm (base 2) of IF,

$$IDF_t = \log\left(\frac{DS}{DF_t}\right) \quad (5)$$

TF-IDF assigns weight to every word based on 'TF' and 'IDF'. Words with greaterweight scores are seen to be significant.

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t \quad (6)$$

4.2 *Word to Vector (Word2Vec)*

Word Embedding (WE) facilitates word representation that deals with converting text into numeric form. It is a vector involving words which capture the context related to other words. WE schemes play a dominant role in mining text as MLschemes cannot be applied on text. Technically, WE schemeconverts a word into a numberby using vocabulary. It can be trained on hugequantity of text usingNN. Embedding methods are based on frequency and prediction. Frequency-based schemes produce text vectors by determining frequencyof words which occur often. Prediction-based methods vectorise words based onformer knowledge as well asNN. Text is cleaned and tokenizer splits words, letters and symbols in every sentence. Word2Vec aids in training word vectors for classifying subjectivity, objectivityand sentiment from subjectivity corpus. Features from text are producedthrough Word2Vec for supervised MLapproaches.The proposed model is based on prediction(Tomas Mikolovet al 2013). This model involves a shallow NN to include word vectors of high quality. The words in comparable context are associated. The algorithm includesContinuous BoW(CBoW) and Skip-Gram (SG) models. CBoWforecasts target word for a context of words, while SG model flips CBoWframework by envisaging context of a particular word. The SG model combined with negative sampling outdoes CBoW making it suitable for this work. It enables NNs to learn words along with the context. CBoW comprehends the ensuing word depending on context, while SG model finds the word's context.To classify texts after vector training, LSTM, Bi-LSTM and SBi-LSTM are used.

4.3 *Word Embedding using BERT*

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

Transfer Learning (TL) is a ML paradigmwhich emphases on using information obtained for a task to handle other comparable tasks. Transformer, arelated BERT model is developed by Devlin et al (2018). BERT is a consecutive language model which considers input as a sequence say,$I = (I_0, \dots, I_n)$ and produces contextualized vector $H = (h_0, \dots, h_n)$ asinput.It is anextremely generalisedframework that is a language representation. It performs task using encoder.Encoders are NN frameworkswhich are used for creating encoded illustrations of text. BERT-Mini includes4 encoder layers. Every encoder block includes 2 sub-layers that are multi-headed and are Feed Forward (FF).Every encoder layer includes 2 sub-processes namely, multi-headed self-attention layer that comprisesof a collection of metric handlingfunctions. Input is given to encoder through self-attention layer which helps in miningsignificant features. Once features are extracted, they are normalized by using residual connection and input ofFF layer. FF layer's outputis given as input to encoder's 2nd layer and process is repeated for ensuing encoder layers.Multi-headed attention includesnumerous heads whichexecute in parallel and every head is denoted by self-attention whichfinds the associationamid words present in a phrase (Vaswani et al 2017).

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{Dim_K}}\right).V \quad (6)$$

Where,

Q, K, V - Query, Key, Valuevectors

Dim_k - Dimension of 'K'

- For determining similarity scores, query's dot product of 'Q' and 'K' (QK^T).
- 'QK^T' is divided by '$\sqrt{Dim_k}$'
- Softmax function is employed in normalization and in determining score matrix.

Attention matrix (A) is got by finding the product of score matrix and 'V'.Likewise, attention facilitates model to diverse representations as well as subspaces at diverse locations, whilean attention-head averaging avoids this. Multi-head attention (M_A) is given by,

$$M_A(Q, K, V) = \text{con}(z_0, z_1, \dots z_h) \quad (7)$$

FFincludes 2 linear transformations divided by ReLU activation. It is identically included in each spot. It is given by,

$$F_n = \max(0, xW_1 + b_1)\, W_2 + b_2 \quad (8)$$

Word2Vec focuses on creating word embeddings which occurs before BERT. It produces embeddings which are independent of context, so every word is expressed as one vector. It offers trained word embeddings that can be usedregularly. Embeddings are key-value pairs, basically 1-1 mappings amid words and corresponding vectors. It takes word as input and generatesa word vector.BERT produces context-based embeddings which permitnumerous representations of every word based on particular word's context. It is bidirectional encoding approach whichpermits it to ingest location of every word and includeit into word's embedding, whereas Word2Vec embeddings do not offer word location.Word2Vec knows aboutword level embeddings, it producesword embeddings that exist in the training set. It doesn't support words out of the vocabulary. Instead, BERT learns sub-word level representations, a model with lesser vocabulary space than amount of unique words in training corpus. BERT is capable

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

ofproducingword embeddings outside its vocabulary space offering a closeunlimited vocabulary.

5. CLASSIFICATION OF TWEETS

Tweets are classified using LSTM and Bi-LSTM.

5.1 *Long Short-Term Memory (LSTM)*

LSTM is a type of sporadic neuronal structure. In Recurrent Neural Network (RNN), output of former stage is treated as a contribution to present stage. Hochreiter and Schmidhuberet al (1997)have proposed LSTM. It addressesissue of long-term conditions in RNN that is incapable of anticipating long-term memory words but providing accurate estimates based on new information. RNN does notoffer convincing execution with increase in total length. LSTM can store data for long periods. This is utilized for time series data-based planning, analysis and forecasting. LSTM involves a chain structure with 4 neuronal groupings and distinguishable memory units known as cells. Such cells are responsible for data management, while gates are in-charge of memory management. LSTM gates control information flow into and out of memory cell. A unit with these gates along with a memory cell is considered as a neuron layer in traditional FF-NN with every neuron having a hidden layer and present state.The Forget gate confirms that the data that is no longer helpful is deleted, while Inputgate is incharge of providing useful data to cells. The output gate is used for extracting useful information from cells.

- ***Input Gate:*** To change memory, data to be taken should be considered. The sigmoid method determinesthe values that give the output '0' and '1'. The 'tanh' function is liable for allocating weights to the values that are passed, deciding their significance levels from - 1 to 1.

$$i_t = \sigma(W_i.[h_{t-1}x_t] + b_i) \quad (9)$$

$$\tilde{C}_t = \tanh(W_c.[h_{t-1}x_t] + b_c \quad (10)$$

- ***Forget Gate:***It decides information whichmust be chosen from a block. The sigmoid functionis used to determine it. For every number in cell state (C_{t-1}), it generates a number amid 0 (omit) and 1 (keep) using former state (h_{t-1}) and content input (X_t).

$$f_t = \sigma(W_f.[h_{t-1}X_t] + b_f) \quad (11)$$

- ***Output Gate:***It determines the outcome dependent on its memory and input. The sigmoid method determines the values that give the output '0' and '1'. The tanh function is responsible for allocating weights to the values that are passed, deciding their significance levels from - 1 to 1. The result of this function is multiplied with sigmoid result.

$$O_t = \sigma(W_o[h_{t-1}x_t] + b_o) \quad (12)$$

$$h_t = o_t \tanh(C_t) \quad (13)$$

5.2 *Bi-directional LSTM (Bi-LSTM)*

Bi-LSTM is a kind of RNNwhich processes input in forward as well as backward directions. In conventional LSTM, there is a flow of information from past to future, and predictions are made based on previous context. Nevertheless, in Bi-LSTMs, future context is considered, enabling it to determine dependencies in eitherdirection.Bi-LSTMhas 2 layers to process input in forward and backward directions. It enables NN to have sequence information in both directions. This flow of input in either direction enables preserving both future as well as past information and

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

simultaneously accessing them.Bi-LSTMs are mainly useful for tasks which demand acomplete understanding of input sequence like NLP tasks including SA, machine translation as well as recognition of named entity.By includinginformationin both directions, Bi-LSTMs improve model's capability to determine long-term dependencies and make preciseforecastsinvolving sequential data.

5.3 *Proposed Stacked Bi-directionalLSTM (SBi-LSTM)*

Media includingimages, audios and videos are on the increase in the recent past. Advanced tools for digital manipulation along withrecent techniques make it simple for generating content as well as post them on social media. Tweet polarity is substantialfor determiningpeople's sentiment. People's sentiments are to be analysed. The proposed method includes pre-processing and several ML and DL models are verified (Figure 1). In this paper, a Deep Learning (DL) model is propounded to envisage polarity of tweets. LSTM layer can be used for determining forward dependency, and hence employed as topmost layer of model. Deep framework called Stacked Bi-directional LSTM (SBi-LSTM) is proposed to forecast values. SBi-LSTM network is designed to categorize sentiments of false tweets. ML classifiers like LSTM, Bi-LSTM and SBi-LSTM are applied. These classifiers use TF-IDF, Word2Vec and word embedding using BERT for feature extraction. Deep LSTM frameworks with numerous hidden layers can build gradually increased levels of representation of sequence data. The networks include stacked hidden layers, wherein output of hidden layer is given as input to following hidden layerBi-LSTMs use forward as well as backward dependences. They are appropriate for being initial layer for learning useful information. When predicting values, top layer of framework should use learned features called outputs from lower layers for iteratively computing in forward direction and produce anticipated values. If input includes missing values, masking layer must be used by SBi-LSTM. Every SBi-LSTM includes Bi-LSTM layer as initial layer for learning features, and LSTM layer as final layer. To support efficient use of input data as well as learning complex wide-ranging features, proposed SBi-LSTM may include more number of LSTM and Bi-LSTM layers

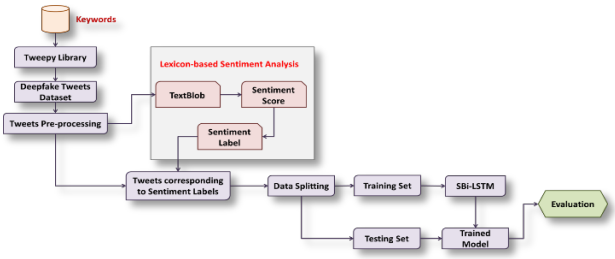

Figure 1: Architecture of Proposed Method

. The proposed model is capable of envisaging values for numerous future steps depending on historical data.

SBi-LSTM offers better performance due toits modest structure with stacked LSTM layers. It includes6 layers comprising of 1 embedding, 2 drop-out, 2 Bi-LSTM and 1 dense layer. Initially, pre-processed data includes word sequences that are passed to embedding layer. The embedding layer's output is passed to dropout layer with which has a dropout rate of 0.5that reduces complexity of input data. Output is passed through a layered stack. Bi-LSTM facilitatesadded training by traversing input twice in both directions. Added training of data offers

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

improvedoutcomes. First Bi-LSTM's output is consideredas input for next Bi-LSTM for supportingprecise prediction. The dropout layer is employed before and after Bi-LSTMs. Lastly, dense layer is employed with 3-units along withSoftmax function (Figure 2). Asintegrated models show improved performance, SBi-LSTM includes 2 Bi-LSTMs to build a stack. In stacked structure, initial layer determines significant features regarding target class. It helps next layer to offerprecise results. Stacking aids in incorporating the abilities of the models and make improved predictions when compared to one model. By employing 2 Bi-LSTMs in stacked structure, better outcomesare obtained for classifying sentiments. Moreover, it simplifiesthe model, thus supporting extensive use of propounded method.The model is compiled using 'Adam' optimizer, 'categorical_crossentropy' loss function with 100 epochs. The proposed ensemble model outdoes other models due to ensemble framework. Performance of DL models like LSTM, Bi-LSTM and SBi-LSTM is analysed. From the results, it is seen that propounded SBi-LSTM outdoes ML and DL models. The study leverages use of diverse ML methods. Initially, data are obtained from Twitter using 'tweepy' library.Classifiers are used for training as well as testing pre-processed data.

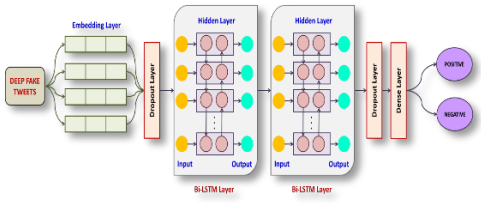

Figure 2: Architecture of Proposed Ensemble Model

The proposed SBi-LSTM classifies the tweets based on sentiments into positive and negative. It offers improved classification accuracy. Performance is analysed based on accuracy, recall, precision as well as F1-Score and compared with other standard techniques. SBi-LSTM supports deep classification of sentiments and a framework of ensemble model is designed. It even performs better for small datasets.

6. RESULTS AND DISCUSSION

Word2Vec in Gensim python library is trained on processed data. For improved WEs, particular hyper parameters like training algorithm, context window, dimensionality and sub-sampling are taken. Negative sampling is used for training as it proves to be computationally effective in contrast to Softmax. Hidden layer of NN is assigned a dimension of 300 as it offers improved WEs. A window size of 10 is used for skip-gram models. Sub-sampling rate of le-3 is used for handling the imbalance among rare as well as frequent words.***Dataset Details***

Sentiment140 dataset with 16×10^5tweets extracted using Twitter API is taken for study. Following columns are present in Twitter data.

- **target:** Refers toTweet polarity (+ or -)
- **ids:** SignifiesTweet'sUnique ID
- **date:** Represents Tweet date
- **flag:** Refers to query. Returns NO QUERY in case such query does not exist
- **user:** Refers to name of the person who tweeted
- **text:** Refers to tweet text

The following figures (Figure 3-5) show the number of Positive and Negative tweets.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

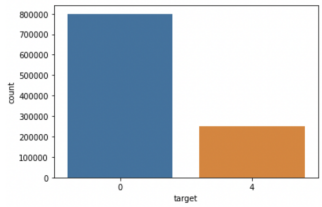

Figure 3:Number of Positive and Negative Tweets

Figure 4: Some Negative Tweets from Dataset

Figure 5: Some Positive Tweets from Dataset

Model Architecture

Embedding Layer isaccountable for transforming tokens into vector representation which is produced by Word2Vec model. Pre-defined layer from Tensorflowis used.The following arguments are involved.

- **input_dim**: Vocabulary size
- **output_dim**: Dense embedding dimension
- **weights**: Embedding matrix initialisation
- **trainable**: Mentions whether layer is trainable or not

Bidirectional wrapper for RNNsmeans the context carried in both directions in wrapped RNN layer.LSTM, a RNN variant comprises of memory state cell for gaining understanding ofword context for including contextual meaning rather thanneighbouring words as RNN.The following arguments are used.

- **units:** Output space dimension (Positive)
- **dropout:** Portion of units for dropping for linear conversion of inputs
- **return_sequence:** Returnsfinal output in output or complete sequence

The Conv1D layer forms a convolution kernel that is convolved with input of layer over single dimension to produce an output tensor.The following arguments are used.

- **filters:** Output space dimension
- **kernel_size:** Length of 1D convolution window
- **activation:** Activation function

GlobalMaxPool1D deals with down sampling of input representation by consideringmaximum value over diverse dimensions.Dense layer includes a Fully Connected (FC) layer into the model.

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

The argument denotes quantity of nodes in the layer.Final dense layer includes'Sigmoid', an activation function that is used for transforming input to a number within the range [0, 1]. These activations are usually used when 2 classesare there for output.

Confusion Matrix

Confusion matrix is used for deciding whether the model has mademore number of false negative predictions when compared to positive. This means that model is biased forenvisaging negative sentiments.From classification report, it is evident that an accuracy of 97% is gotafter training for 12 epochs which is far better in contrast to other models.

Evaluation parameters

Performance is analysed in terms of Accuracy, Recall, Precision and F1-Score which are most common. There possible outcomes of the models include the following:

- ***True Positive (TP)***: Represents number of positive predictions of a class which areappropriately predicted.
- ***True Negative (TN)***: Representsnumber of negative predictions of a class which are appropriately labelled.
- ***False Positive (FP)***: Representsnumber of negative predictions of a class which are inaccurately labelled as positive.
- ***False Negative (FN)***: Representsnumber of positive predictions of a class which areinaccurately labelled as negative.
- ***Accuracy*:** It is widely used forevaluatingperformance of models. It is ratio ofappropriatelyanticipated instances to amount of predicted instances.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (14)$$

- ***Precision*:** It representsclassifier's correctness. It is ratio of positive instances to total number of instances that areanticipatedto be positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (15)$$

- ***Recall*:**Itrepresents classifier's completeness. It gives percentage of TP instances which are correctly labelled.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (16)$$

- ***F1-score:***It is a combination of precision as well as recall signifying their harmonic mean, and is considered as a balanced and well-expressed performance metric.

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (17)$$

Performance of CNN, LSTM, Bi-LSTM and SBi-LSTM are shown below without and with BERT. The proposed SBi-LSTM offers improved results when compared to CNN, LSTM and Bi-LSTM.

Table-1 shows CNN, LSTM and Bi-LSTM schemes without BERT

Algorithm	Accuracy	Precision	Recall	F –Measure
CNN	87	82	81	84

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

LSTM	90	85	86	87
Bi-LSTM	93	90	89	91
Stacked Bi-LSTM	97	94	94	94

Initially, the results are shown without BERT. SBi-LSTM without BERT offers better Performance Analysis in contrast to CNN, LSTM and Bi-LSTM schemes without BERT (Figure 6).

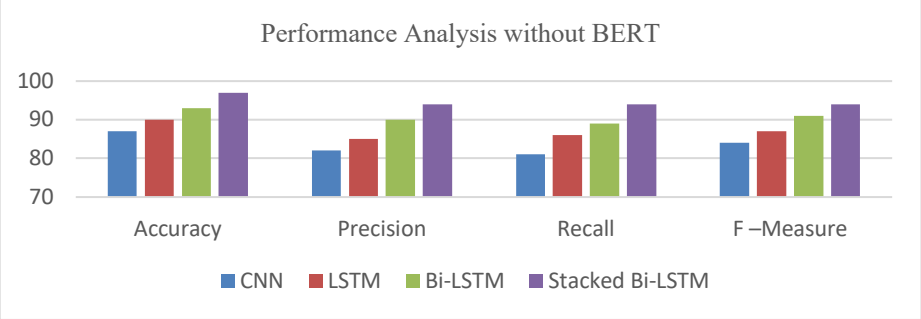

Figure 6: Performance Analysis without BERT

Table-2 shows CNN, LSTM and Bi-LSTM schemes with BERT

Algorithm	Accuracy	Precision	Recall	F – Measure
CNN	85	86	81	85
LSTM	90	88	85	89
Bi-LSTM	93	91	88	92
Stacked Bi-LSTM	97	96	94	95

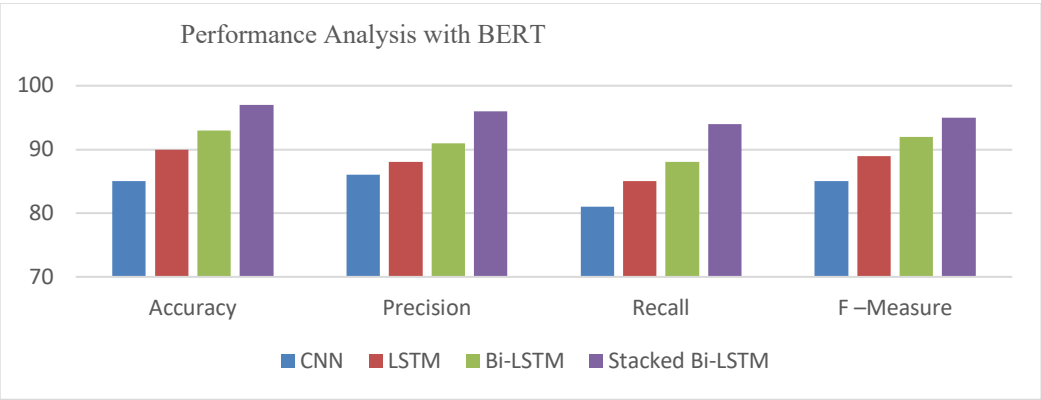

Figure 7: Performance Analysis with BERT

7. CONCLUSION

In this paper, tweets extracted using Twitter API. The tweets are pre-processed using tokenisation

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

by stemming and lemmatization. Features are extracted using Word2Vec, TF-IDF and word embedding using BERT. BERT is a bi-directional encoding approach which permits it to ingest location of every word and include it into word's embedding, whereas Word2Vec embedding does not give word location.Inclusion of BERT offers improved performance. Tweets are classified into positive and negative using LSTM, Bi-LSTM and proposed Stacked Bi-LSTM (SBi-LSTM)model. Proposed SBi-LSTM includes more number of LSTM and Bi-LSTM layers to support efficient use of input data as well as learning complex wide-ranging features. It is seen that SBi-LSTM offers better results in terms of Accuracy, Recall, Precision and F1-Score. In the future, it is planned to apply the proposed model on a larger dataset and focus on involving other enhanced DL-based models for analysing sentiments in the dataset.

REFERENCES

[1] LeCompte T., & Chen J. 2017. Sentiment analysis of tweets including emoji data. *In IEEE International Conference on Computational Science and Computational Intelligence (CSCI), pp. 793-798.*

[2] Chen Y., Yuan J., You Q., & Luo J. 2018, October. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. *In Proceedings of 26thACM International Conference on Multimedia, pp. 117-125.*

[3] Alrumaih A., Al-Sabbagh A., Alsabah R., Kharrufa H., & Baldwin J. 2020. Sentiment analysis of comments in social media. *International Journal of Electrical & Computer Engineering (2088-8708), 10(6).*

[4] Ullah M. A., Marium S. M., Begum S. A., &Dipa N. S. 2020. An algorithm and method for sentiment analysis using the text and emoticon. *ICT Express, 6(4), pp. 357-360.*

[5] Fernández-Gavilanes M., Costa-Montenegro E., García-Méndez S., González-CastañoF. J., &Juncal-Martínez J. 2021. Evaluation of online emoji description resources for sentiment analysis purposes. *Expert Systems with Applications, 184, 115279.*

[6] Kumar T. P., &Vardhan B. V. 2022. A Pragmatic Approach to Emoji based Multimodal Sentiment Analysis using Deep Neural Networks. *Journal of Algebraic Statistics, 13(1), pp. 473-482.*

[7] Zhu M. 2022. Sentiment Analysis of International and Foreign Chinese-Language Texts with Multilevel Features. *Discrete Dynamics in Nature and Society, 2022.*

[8] Amrullah M. S., Budi I., Santoso A. B., & Putra P. K. 2023. The effect of using Emoji and Hashtag in sentiment analysis on Twitter case study: Indonesian online travel agent. *In AIP Conference Proceedings, 2654(1), AIP Publishing.*

[9] AroraM., &Kansal V. 2019. Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis. *Social Network Analysis and Mining, 9, pp. 1-14.*

[10] Patel R., &PassiK. 2020. Sentiment analysis on twitter data of world cup soccer tournament using machine learning. *IoT, 1(2), pp. 14.*

[11] Ajitha P., SivasangariA., Immanuel Rajkumar R., &Poonguzhali S. 2021. Design of text sentiment analysis tool using feature extraction based on fusing machine learning algorithms. *Journal of Intelligent & Fuzzy Systems, 40(4), pp. 6375-6383.*

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

[12] Garg N., & Sharma K. 2022. Text pre-processing of multilingual for sentiment analysis based on social network data. *International Journal of Electrical & Computer Engineering (2088-8708), 12(1).*

[13] NingsihM. R., Wibowo K. A. H., Dullah A. U., &Jumanto J. 2023. Global recession sentiment analysis utilizing VADER and ensemble learning method with word embedding. *Journal of Soft Computing Exploration, 4(3), pp. 142-151.*

[14] Soesanto A. M., Chandra V. C., &Suhartono D. 2023. Sentiments comparison on twitter about lgbt. *Procedia Computer Science, 216, pp. 765-773.*

[15] Ahuja R., Chug A., Kohli S., Gupta S., & Ahuja P. 2019. The impact of features extraction on the sentiment analysis. *Procedia Computer Science, 152, pp. 341-348.*

[16] Habib M. W., &SultaniZ. N. 2021. Twitter sentiment analysis using different machine learning and feature extraction techniques. *Al-Nahrain Journal of Science, 24(3), pp. 50-54.*

[17] Chiny M., Chihab M., BencharefO., &ChihabY. 2021. LSTM, VADER and TF-IDF based hybrid sentiment analysis model. *International Journal of Advanced Computer Science and Applications, 12(7).*

[18] Singh S., Kumar K., & Kumar B. 2022. Sentiment Analysis of Twitter Data using TF-IDF and Machine Learning Techniques. *In IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Vol. 1, pp. 252-255.*

[19] Kaur G., & Sharma A. 2023. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data, 10(1), pp. 5.*

[20] Parveen N., ChakrabartiP., Hung B. T., & Shaik A. 2023. Twitter sentiment analysis using hybrid gated attention recurrent network. *Journal of Big Data, 10(1), pp. 1-29.*

[21] Rao G., Huang W., Feng Z., & Cong Q. 2018. LSTM with sentence representations for document-level sentiment classification. *Neurocomputing, 308, pp. 49-57.*

[22] Li W., Qi F., Tang M., & Yu Z. 2020. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing, 387, pp. 63-77.*

[23] SoniJ., &MathurK. 2022. Sentiment analysis based on aspect and context fusion using attention encoder with LSTM. *International Journal of Information Technology, 14(7), pp. 3611-3618.*

[24] Iparraguirre-Villanueva O., Alvarez-Risco A., Herrera Salazar J. L., Beltozar-Clemente S., Zapata-Paulini J., Yáñez J. A., &Cabanillas-Carbonell M. 2023. The public health contribution of sentiment analysis of Monkeypox tweets to detect polarities using the CNN-LSTM model. *Vaccines, 11(2), pp. 312.*

[25] Devlin J., Chang M. W., Lee K., &ToutanovaK. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

[26] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., ... &Polosukhin I. 2017. Attention is all you need. *Advances in neural information processing systems, 30.*

[27] Hochreiter S., &Schmidhuber J. 1997. Long short-term memory. *Neural computation, 9(8), pp. 1735-1780.*

[28] RoessleinJ. 2009. tweepy Documentation. Online] *http://tweepy. readthedocs. io/en/v3, 5, pp. 724.*

ISSN:2731-538X | E-ISSN:2731-5398
Vol. 18 No. 01 (2024)

[29] Rustam F., Ashraf I., Mehmood A., Ullah S., & Choi, G. S. 2019. Tweets classification on the base of sentiments for US airline companies. *Entropy, 21(11), pp. 1078.*

[30] Rustam F., Khalid M., Aslam W., Rupapara V., Mehmood A., & Choi G. S. 2021. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *Plos one, 16(2), e0245909.*