The Scientific Temper (2024) Vol. 15 (3): 2700-2708
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.38
https://scientifictemper.com/
RESEARCH ARTICLE

An improved social media behavioral analysis using deep learning techniques

Muhammed Jouhar K. K.*, K. Aravinthan

Abstract

Most online users share their opinions and comments or give their valuable feedback on a variety of subjects. Public opinions and comments in social media have had a great impact on social and political systems. This vast information can be reviewed and analyzed. As this online information grows in numbers, it requires efficient processing. Thus, this information can be mined or analyzed effectively, making it a suitable candidate for data mining. Twitter's micro-blogging service has more than 250 million active users who post short messages about any topic. This vast information is a meaningful source of information regarding different aspects of. This paper proposes to mine and extract information from tweets called IBADL (Improved Behavioral Analysis using Deep Learning). The goal of the proposed technique is to mine information through the study of the tweets posted and conduct an analysis for drawing meaningful conclusions about the behavior of Twitter users.

Keywords: Deep Learning, Behaviour Analysis, ConvNet, Twitter, Positive tweets.

Introduction

Social networking platforms like Twitter and Facebook have grown rapidly in terms of users and usage, where people express their views about things, places and personalities. The current generation buys products, shares profiles and opinions online and also express reviews, discussions and opinions about a subject. Moreover, the Internet has been an easy platform for end-users to provide feedback on many topics and products or services used by them. Organizations have been reducing their expenditure on surveys or opinion polls and have started to focus on more easily available public opinions. Analyzing these opinions or studying user behavior on tweets can be categorized predominantly (Medhat W *et al.*, 2002), Lexicon-based (Taboada M *et al.*, 2011) and hybrid (Prabowo R *et al.*, 2009) (Dang Y *et al.*, 2010). It can also be presented as statistical, knowledge-based and hybrid approaches (Cambria E *et al.*, 2016). There are gaps in performing research in broad areas by computationally analyzing opinions and Behaviour (Jagdale O *et al.*, 2017). Therefore, a gradual practice has grown to extract the information from data available on social networks for the prediction of an election, to use for educational purposes, or for the fields of business, communication and marketing. The accuracy of analysis and predictions can be obtained by behavioral analysis based on social networks (Anjaria M *et al.*, 2014). Studies (Miranda Filho R *et al.*, 2015; Castro R *et al.*, 2017) have proved that it is possible to get an insight in contrast to traditional ways of obtaining information about perceptions. Their opinions can be exploited from tweets while considering a large scale of data for analysis. However, the volume of such reviews is growing rapidly, necessitating data mining approaches to extract meaningful information from this huge volume of data.

Behavioral analysis (BA) is the computational study of opinions, emotions and attitudes toward entities of interest. It is applied very importantly in academia and commerce to analyze the attitudes or behavior of a writer or speaker on subjects in relation to some topic. Computationally, it refers to the use of natural language processing (NLP), text analysis, computational semantics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. The inception and rapid growth of BA coincide with social media activities on the Web like reviews, forum discussions, blogs, micro-blogs, Twitter, and

Department of Computer Science, Adaikalamatha College, affiliated to Bharathidasan University, Vallam, Thanjavur, Tamilnadu, India.

***Corresponding Author:** Muhammed Jouhar K. K., Department of Computer Science, Adaikalamatha College, affiliated to Bharathidasan University, Vallam, Thanjavur, Tamilnadu, India., E-Mail: muhammedjouhar87@gmail.com

How to cite this article: Jouhar, M. K. K., Aravinthan, K. (2024). An improved social media behavioral analysis using deep learning techniques. The Scientific Temper, **15**(3):2700-2708.

Doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.38

Source of support: Nil

Conflict of interest: None.

Received: 09/07/2024 **Accepted:** 10/08/2024 **Published** : 16/09/2024

social networks. BA has grown to be one of the most active research areas in NLP and is widely studied in data mining, Web mining, text mining, and information retrieval. BA is also not devoid of challenges. Professional writer's writings about science or facts have different structures. Semi-structured sentiments lie on the range between formal structured sentiments and unstructured sentiments. These require an understanding of several issues and depend on Pros and Cons, which are usually short phrases. Free text forms like chats and comments are sometimes unstructured. Thus, recognizing user behavior from tweets is an uphill task. This paper automates BA with its proposed novel technique IBADL and attempts to overcome certain challenges in BA. The proposed technique analyzes people's thoughts on various subjects and categorizes them by subject while demarcating them into positive, neutral and negative behaviors.

Literature Review

A comparison of politicians was done on tweets extracted using a streaming application programming interface (API) in (Ibrahim M *et al.*, 2015), where scores of the candidates were based on positives and negatives. Negation handling and word sequence disambiguation (WSD) were used for accuracy (Navigli R *et al.*, 2009). Twitter data was also used to forecast a Swedish election outcome using sentiment analysis and link structures from' tweet conversations (Dokoohaki N *et al.*, 2015). A link-prediction algorithm equated and showed popularity from structural links with vote outcomes in the general election. As the sentiment analysis of tweets has gained popularity in recent years, the sentiments of queries generated by users has been calculated (Damaschin M *et al.*, 2012) by page-rank algorithms and the Naïve Bayes classifier. For the prediction of the 2016 US elections, manually annotated corpus-based hashtags (Rezapour R *et al.*, 2017) along with negation detection were tested and a claim was made that a 7% accuracy level had increased. The model adopted to rank the candidates of political parties included the lexicon-based approach (Ding X *et al.*, 2008) and the linguistic inquiry word count (LIWC) (Pennebaker *et al.*, 2018).

Although a lot of work has been done to analyze tweet sentiments by applying techniques of machine learning (Pang B *et al.*, 2002), the work of (Gautam G *et al.*, 2014) was based on a set of techniques of machine learning that were Naïve Bayesian, SVM (Joachims T *et al.*, 1998) and entropy-based (Berger *et al.*, 1996) along with semantic analyses to classify product reviews or sentences. Additionally, a hybrid approach was adopted (Anjaria M *et al.*, 2014) that involved supervised classifiers which were the artificial neural networks, feed-forward support vector machines (SVM), maximum entropy and Naïve Bayes. To predict election results, they proposed an approach of user influence factor exploitation. Understanding influential factors and analyzing data with SVM resulted in approximately 88% accuracy in both the 2013 Karnataka assembly elections and the 2012 US presidential elections. To show that Twitter trends play an important role in electoral sentiments, the authors of (Khatua *et al.*, 2015) collected hashtag-based tweets covering the candidates of the Indian elections 2014. They did not include neutral tweets for the analysis because they found these kinds of tweets problematic for sentiment analyses that are in favor of more than one party. Two lexicons (Hansen *et al.*, 2011) were combined for the sentiment analysis of tweets. This bipolar lexicon was best in the case of the analysis of two parties, but for the classification of multiple parties, they created variables and their approach was not sufficiently state-of-the-art to calculate sentiment score when more parties were involved in the analysis.

Authors of (Saha *et al.*, 2017) used Textblob for pre-processing, polarity, and the polarity confidence calculation, and they validated the obtained results by SVM and Naïve Bayes using Weka; they reported the highest accuracy of Naïve Bayes with a 65.2% rate, which was 5.1% more than the SVM accuracy rate. An unsupervised machine-learning algorithm was introduced to rate the reviews as thumbs up and thumbs down; 410 reviews were gathered from Epinions and tested; the observed accuracy of the algorithm was 74%. Moreover, for election prediction, authors of (Sharma *et al.*, 2016) tested sentiments related to political parties by SVM and Naïve Bayes. Consequently, they predicted the possibilities of the BJP winning the elections in 2016 on the basis of the SVM result with 78.4% accuracy, which was 16.3% greater than the Naïve Bayesian results.

Research Gap

The Table 1 provides a comprehensive overview of the research gaps identified in the existing literature on social media behavior analysis, sentiment analysis, and election forecasting. Across various studies, there is a consistent need for integrating advanced deep learning techniques with sentiment analysis methods to enhance accuracy and prediction capabilities. While previous research has employed methods such as negation handling, word sequence disambiguation, and lexicon-based sentiment analysis, there remains a significant gap in leveraging deep learning models for improved sentiment classification and behavior understanding. Studies have reported varying levels of accuracy, with some achieving notable results using supervised classifiers and traditional machine-learning algorithms. However, the exploration of more advanced techniques, including the integration of deep learning models like artificial neural networks and SVMs, is crucial for achieving higher accuracy in predicting election outcomes and understanding social media sentiments effectively. Additionally, there's a need to address challenges such as handling multiple parties in sentiment analysis and

Table 1: Research Gap

Author	*Proposed methods*	*Results*	*Research gap*
Ibrahim M *et al.*, 2015	Negation handling, Word Sequence Disambiguation (WSD)	Scores of candidates based on positives and negatives	Lack of integration of deep learning techniques for improved sentiment analysis and behavior understanding in social media
Dokoohaki N *et al.*, 2015	Sentiment analysis, Link structures from tweet conversations	Forecasting Swedish election outcome, Link-prediction algorithm	Need for deeper integration of deep learning methods and sentiment analysis techniques for more accurate predictions
Damaschin M *et al.*, 2012	Page-rank algorithms, Naïve Bayes classifier	Sentiments of queries generated by users	Absence of comprehensive approach integrating deep learning for sentiment analysis and behavior understanding
Rezapour R *et al.*, 2017	Manually annotated corpus-based hash tags, Negation detection	7% increase in accuracy level for 2016 US elections prediction	Lack of exploration on the integration of advanced deep learning models for sentiment analysis and prediction
Gautam G *et al.*, 2014	Naïve Bayesian, SVM, Entropy-based, Semantic analyses	Classification of product reviews or sentences	Need for exploring the integration of deep learning models and semantic analyses for improved sentiment classification
Anjaria M *et al.*, 2014	Supervised classifiers (Artificial Neural Networks, SVM, Maximum Entropy, Naïve Bayes)	Approximately 88% accuracy in predicting election results	Exploration required for integrating advanced deep learning techniques for improved prediction accuracy
Khatua *et al.*, 2015	Combination of lexicons for sentiment analysis	Analysis of Twitter trends in electoral sentiments	Need for more advanced sentiment analysis techniques integrating deep learning for improved accuracy
Saha *et al.*, 2017	Textblob for pre-processing, Polarity calculation, SVM, Naïve Bayes	Highest accuracy with Naïve Bayes (65.2%)	Exploration needed on integrating deep learning for sentiment analysis and prediction
Sharma *et al.*, 2016	Sentiments related to political parties, SVM, Naïve Bayes	SVM predicted BJP winning 2016 elections with 78.4% accuracy	Further exploration is required for integrating deep learning models for more accurate prediction

integrating deep learning algorithms for better sentiment analysis and prediction. Overall, the literature review highlights the pressing need for further research and development in integrating deep learning techniques for enhanced sentiment analysis and behavior understanding in social media platforms like Twitter.

Methodology

Figure 1 depicts the data flow and processing stages within the improved behavioral analysis using deep learning (IBADL) system, which is designed for analyzing Twitter data. The process begins with the extraction of tweets using the Twitter Streaming API, followed by data normalization to filter out retweets and remove special symbols, and punctuation. Next, the system performs polarity analysis to categorize tweets as positive, negative, or neutral based on sentiment. Feature extraction is then conducted to identify frequently occurring words and train the entropy classifier. Finally, behavioral analysis is performed to analyze the sentiment and behavior of Twitter users. The system utilizes various techniques such as convolutional layers, k-max pooling, and softmax output layers to process and classify tweet data effectively. Additionally, the system incorporates data stores for storing Twitter data at different stages of processing, including raw data, cleaned data, and classified data. Overall, the diagram illustrates the comprehensive approach of the IBADL system in analyzing Twitter data to extract meaningful insights into user behavior and sentiment.

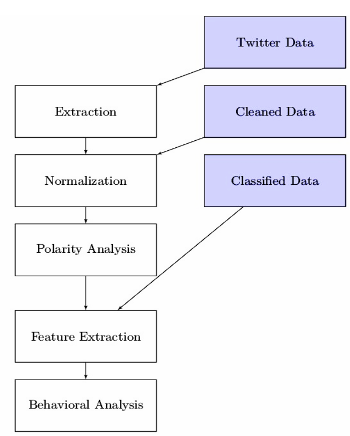

Figure 1: Data flow diagram

Description of Algorithm

Figure 2 illustrates the procedural flow of the algorithm implemented in the research methodology. It begins with the initialization of variables and the downloading of tweets from a source, followed by a series of sequential steps. The

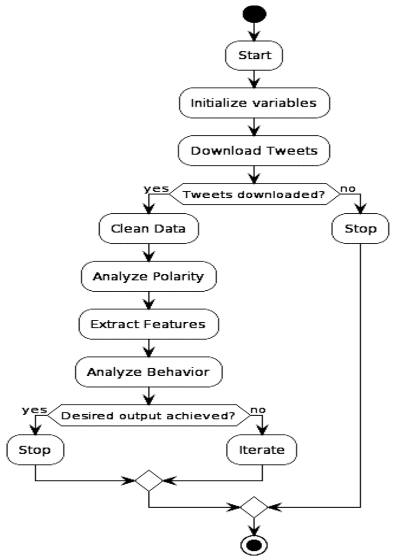

Figure 2: Algorithm flowchart

algorithm first checks if tweets have been successfully downloaded, proceeding to clean the data, analyze polarity, extract features, and analyze behavior. At each stage, the algorithm evaluates whether the desired output has been achieved. If the desired output is obtained, the algorithm stops; otherwise, it iterates through the process again until the desired outcome is reached or until a stopping condition is met. This flowchart provides a clear visualization of the algorithmic workflow, encompassing data acquisition, pre-processing, analysis, and iterative refinement, which are key components of the research methodology.

Deep Learning Technique

Convolutional neural network (CNN) also known as ConvNets is a deep learning tool that has gained expertise in Computer Vision (CV) applications (Srinivas *et al.*, 2016). The use of neural networks for NLP applications is attracting huge interest in the research community and they are systematically applied to all NLP tasks (Kim *et al.*, 2014). The fundamental idea of ConvNets is to consider feature extraction and classification as one jointly trained task. The scope of using this methodology in text analytics has proven to be advantageous in various ways (Severyn *et al.*, 2015). This idea has been improved over the years, in particular by using many layers of convolutions and pooling to sequentially extract a hierarchical representation of the input. By means of hyper-parameter tuning and a series of forward and back propagation we could end up with the desired output of our choice. Figure 3 depicts a ConvNet.

IBADL

This work proposes a ConvNet for the classification of tweets, which is divided into positive, negative and neutral tweets. The proposed technique studies the behavior of real-time Twitter data using Twitter API. The methodology followed

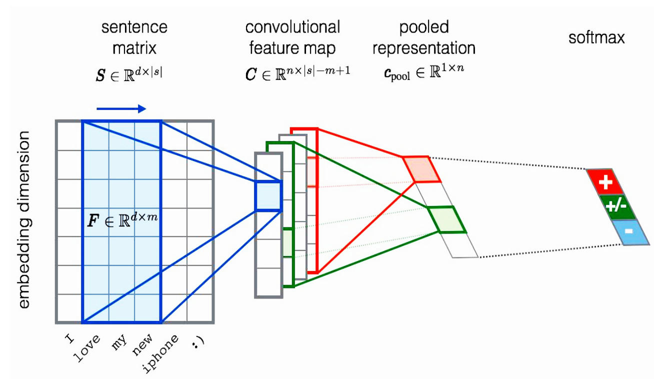

Figure 3: ConvNet structure

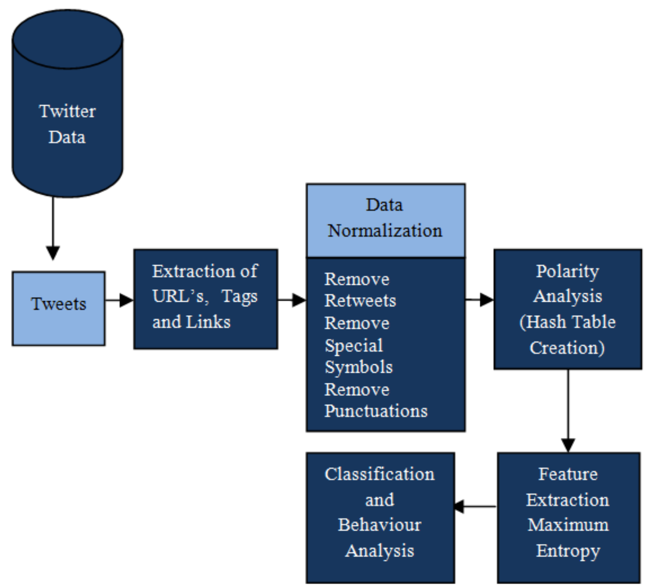

Figure 4: IBADL architecture

in this technique consists of six stages, namely extraction, normalization, polarity analysis, feature extraction, classification and BA. The architecture of IBADL is depicted in Figure 4.

IBADL Input

Twitter Streaming API is a tool that makes interaction with computer programs and web services easy. Many web services provide APIs to developers to interact with their services and to access data in programmatic way. IBADL uses Twitter Streaming API to download tweets.

IBADL Extraction of URL's, Tags and Links

Since Twitter only handles 280 starting in 2017, most URLs are shortened. These links are converted into the full links. The URLs are then classified as "good" or "bad" depending on the information's relevance to the topic. Each URL's is represented as a vector of values ranging between 0 and 1. NLP is then used to detect sentences, remove stop words, and lemmatize other words, thus creating a vector for each article. Data Normalization in the proposed technique

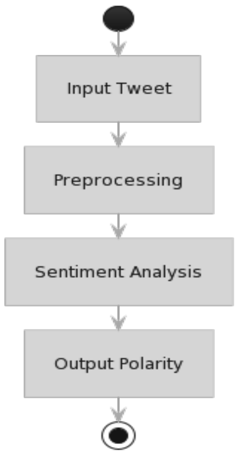

Figure 5: Polarity analysis

involves filtering out the tweets that are retweeted. Special symbols and punctuations are also removed to get clean data as this can improve the efficiency and effectiveness of IBADL.

IBADL Data Normalization

Data is normalized for classification in order for it to be rescaled to the unit interval. Normalization is important because without it the measure will be dominated by the largest scale variable.

BADL Polarity Analysis

Polarity Analysis is used as a next step to create categories of tweets. Social media is a major platform for sharing information and as Tweets related to products are extracted and cleaned, polarity analysis is used to identify positives and negative in tweets. To find the positive tweets using polarity, the normalized data of the previous step is used. Every hash table stores data in the form of (key, value) combination. Interestingly, every key is unique in a Hash Table 1, but values can repeat, which means values can be the same for different keys present in it. IBADL thus creates a hash Table 1 on the polarity analysis process.

Figure 5 illustrates the process of polarity analysis, which involves analyzing the sentiment of a given tweet. The process begins with the input of a tweet, which undergoes pre-processing to standardize and clean the text. Following pre-processing, the sentiment analysis stage employs various techniques to determine the polarity of the tweet, assessing whether it expresses positive, negative, or neutral sentiment. Finally, the output polarity is generated, indicating the overall sentiment of the tweet. This flowchart provides a clear and concise overview of the sequential steps involved in polarity analysis, highlighting key stages from input to output.

IBADL Feature Extraction

The proposed system is train on entropy for the maximum classifier. A weighted word frequency feature trains it. It finds frequently occurring words present in the tweets irrespective like ‘good’, ‘bad’, ‘ok’, etc. will be selected as features words. The entropy classifier is effectively used in a number of NLP applications. The estimate takes the exponential form as in Equation (1),

$$P_{ME}\left(c/d\right) = \frac{1}{z(d)} exp\left(\Sigma_i \lambda_{i,c} F_{i,c}\left(d,c\right)\right) \tag{1}$$

Where, Z (d) is a normalization function. Fi, c is a feature/class function for feature fi and class c, as in Equation (2),

$$\left(d,c'\right) = \ 1, if \ \left(n_i(d) > 0 \ and \ c' = c\right. \tag{2}$$

This classifier works by finding a probability distribution that maximizes the likelihood of testable data. The following are the basic steps on doing maximum entropy classifier:

- Collect a large number of training data which consists of samples represented on the following format: (xi,yi).
- Summarize the training sample in terms of its empirical probability distribution.
- For each word w and class c, define a joint feature f(w,c) = N, where N is the number of times that ‘w’ occurs in a document in class c.
- Iterated to get optimization results, assign a weight to each joint feature so as to maximize the log-likelihood of the training data [13].

The classified tweets are analyzed based on polarity of the words like good, bad, not etc. Based on the polarity, the number of positive tweets and negative tweets are identified.

Table 2 outlines the sequential steps involved in the feature extraction process. It begins with “Input Data,” representing raw data obtained from sources. The next step, “Pre-processing,” involves cleaning, normalization, and transformation of the data to make it suitable for analysis. Following pre-processing, “Tokenization” breaks the text into individual words or tokens, facilitating further analysis. “Feature Selection” involves identifying relevant features from the tokenized data, while “Feature Extraction” extracts numerical representations of these features for analysis. Finally, the process concludes with “Output Features,” where the processed features are ready for use in the analytical process. Each step in the Table 2 highlights a crucial stage in the feature extraction process, from initial data acquisition to the generation of processed features ready for analysis.

Table 2: Feature extraction process

Step	*Description*
Input Data	Raw data obtained from sources
Pre-processing	Cleaning, normalization, and transformation
Tokenization	Breaking text into individual words or tokens
Feature Selection	Identification of relevant features
Feature Extraction	Extracting numerical representations of features
Output Features	Processed features ready for analysis

IBADL Behavioral Analysis

Let us consider a tweet with m tokens where each token in a tweet is mapped onto the corresponding word vector by looking up the word vector table

$$L \epsilon R^{(nx|V|)} \quad (3)$$

where V is the word vocabulary and n being dimensions of the word vector. Each word.

$$\omega_i \epsilon R^n \quad (4)$$

After mapping, the tweet is expressed as a vector of word embeddings concatenation to which unigram, bigram, word sentiment polarity score feature vectors are applied as a feature vector v of tweet

$$v = \omega_1 \oplus \omega_2 \oplus \omega_3 \oplus \omega_4 \dots \oplus \omega_{n+1} \oplus \omega_{n+2} \oplus \omega_{n+3} \quad (5)$$

Where $\oplus$ is the concatenation operator of the vector. $\omega_{n+1} \epsilon R$ is the word sentimental polarity vector. $\omega_{n+2} \epsilon \{0,1\}$ is unigram and bigram feature and ω_{n+3} were the twitter specific features. To unify the matrix representation of tweets in different length, the maximum length of all tweets in the dataset is used as the fixed length for tweet matrices. For shorter tweets, zero vector was padded at the back of a tweet matrix. In the first convolution layer, convolution calculation are performed using employ multiple filters with variable window size h, and generate local sentiment feature vector χi for each possible word window size. We can use a bias $b \epsilon R$ and transition matrix $\omega \varepsilon R^{(hu*hn)}$ generated for each filter,where *hu*: amount of hidden units and *hn*: total units in the convolution layer where each convolution operation will generate a new contextual local feature

$$\chi i = f(\omega.vi : i + h - 1 + b) \quad (6)$$

where *f* is the non-linear active function and v is the local vector from ith position to *(i+h-1)th* position in vector *v*. The convolution filter generates a local feature mapping vector for each possible word window in the tweet, which is followed by the completion of the convolution operation to generate a new vector

$$x=\{x1,x2,x3,x4\dots xn-h+1\} \quad (7)$$

This is followed by k-max pooling operation that is employed over the new feature vector x generated by the convolution layer. It maps the vector x to a fixed length vector where the length is a hyper parameter determined by the user and corresponds to the number of hidden layers within the convolution network. The top k features are selected through the k-max pooling technique which corresponds to the multiple hidden layers so as to retain the important sentiment feature information. In order to obtain better feature information, we fed the fixed length vectors created by the k-max pooling to a convolution layer for obtaining a new vector again. In the model, we select the hidden layers to contain three convolution layers and three k-max pooling layers.

The convolution layer involves τ with *f* filters *3*d* resulting in feature maps $M \epsilon R^{(f*(n-2))}$ The max-pooling-over-time layer is responsible for selecting the most relevant features within the temporal dimension by using lters of size f*(n2).In twitter classification the resultant outcomes classes can have two polarities positive and negative which can be configured using a softmax output with two neurons. The output layer of the architecture is a softmax layer that generates probability value of positive or negative sentiment. The output layer uses a fully connected softmax layer to adjust the sentiment characteristics of the input layer, and gives a probability distribution of the sentiment classification labels $y_j = \omega_j y_{j-1} + b_j$.

y_j: output vector of softmax layer.
y_{j-1}: output vector of pooling layer.
w_j: transition matrix of softmax layer.
b_j: bias factor of softmax layer.

The probability distribution over the sentiment labels is:

$$P(i|t,\theta) = \left((\exp(yi \wedge j))/\Sigma(exp(yk^{j})\right) \quad (8)$$

Where k runs from 1 to n and so does the summation series. We apply dropout regularization to the fully connected layers to eliminate the problem of a lot of hidden units and the connections between them.

IBADL Results

Twitter Extraction

User can interface with twitter for fetching tweets by establishing a connection the twitter API called Tweepy. Figure 6 depicts a screen shot of Twitter Tweets Extraction.

IBADL data normalization is the removal of URLs and unwanted text using regex expressions and functions. Figure 7 depicts the data normalization function used in IBADL.

IBADL classifies tweets after combining them into a single sentence to identify positives in tweets. Figures 8,9 and 10 display the positive, negative and neutral tweets categorized by IBADL.

IBADL behavioral analysis takes the entire tweets in its processing and based on the classification of tweets, it

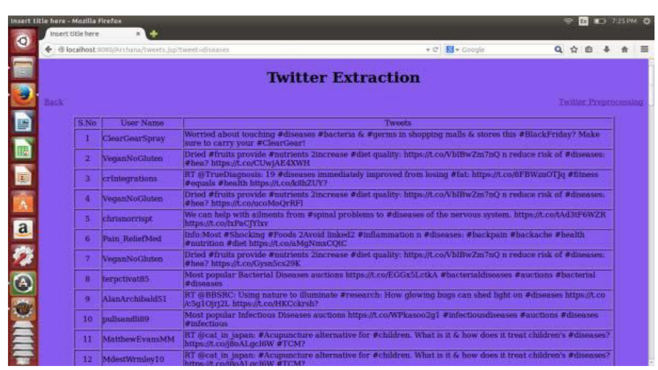

Figure 6: IBADL Twitter extraction

```
def clean_tweet(self, tweet):
        '''
Utility function to clean tweet text by removing links, special characters
using simple regex statements.
        '''
return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", " ", tweet).split())
```

Figure 7: IBADL data normalization function

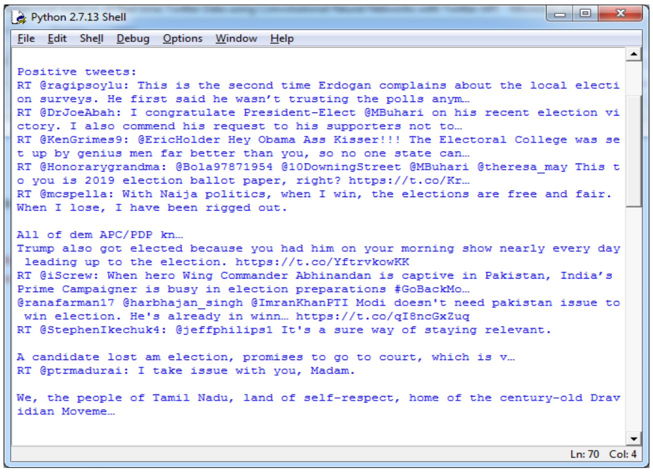

Figure 8: Positive tweets

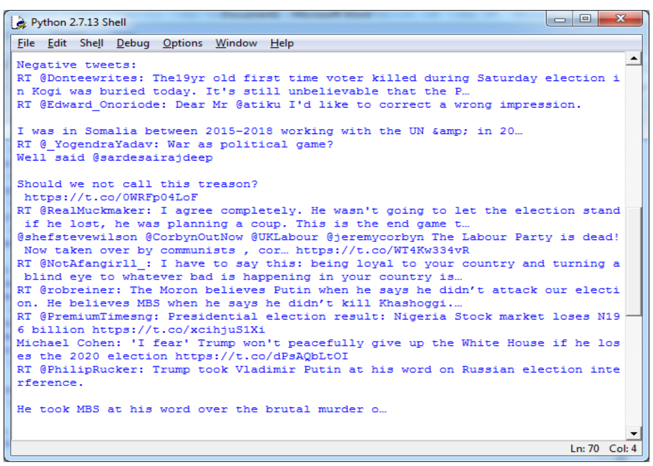

Figure 9: Negative tweets

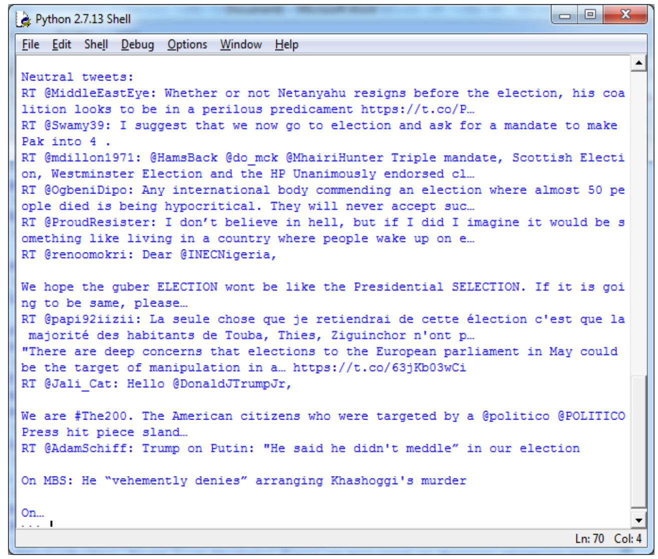

Figure 10: Neutral tweets

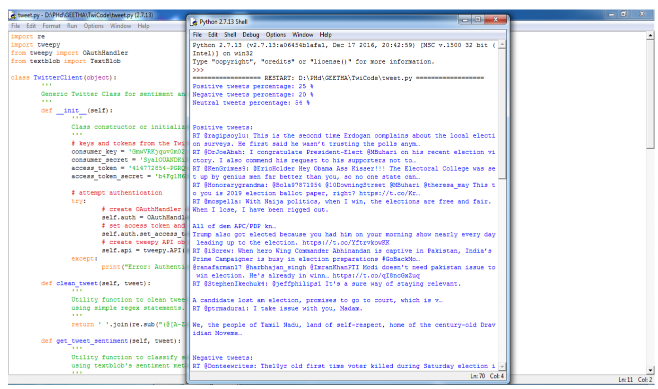

Figure 11: IBADL behavioral analysis

attempts to study the behavior of users in terms of positive, negative and non-committal behaviors. Figure 11 depicts IBADL behavioural analysis, while Table 3 lists IBADL analysis of tweets.

It can be seen from Table 4 that IBADL performed better than SVM classifiers on MATLAB in terms of behavioral analysis. The individual accuracies obtained by SVM were lesser than IBADL. In classifying individual subjects, IBADL also performed better, thus validating it. The performance

Table 3: IBADL analysis of tweets from Twitter data

Keywords	*IBADL*				*SVM classifier*			
	Positive tweets %	*Negative tweets %*	*Neutral tweets %*	*Prediction Accuracy*	*Positive tweets %*	*Negative tweets %*	*Neutral tweets %*	*Prediction Accuracy*
Election	25	20	54	99	21	25	50	96
Tamilnadu	24	5	70	99	13	10	74	97
Adani	28	12	59	99	41	11	45	97
India	39	16	43	98	31	16	51	98
Nature	30	12	57	99	17	11	69	97

Table 4: IBADL performance on analysis of tweet behaviour

Keywords	*IBADL*				*SVM classifier*			
	Positive Rate	*Negative Rate*	*F Measure*	*Accuracy*	*Positive Rate*	*Negative Rate*	*F Measure*	*Accuracy*
Election	31	8	14.85	98.88	30	9	13.85	97.88
Tamilnadu	35	4	12.35	96.45	32	12	17.45	93.45
Adani	38	6	8.66	93.72	23	7	10.73	90.11
India	33	9	7.33	89.66	19	19	19	87.41
Nature	39	12	12.21	96.33	37	11	16.96	95.21

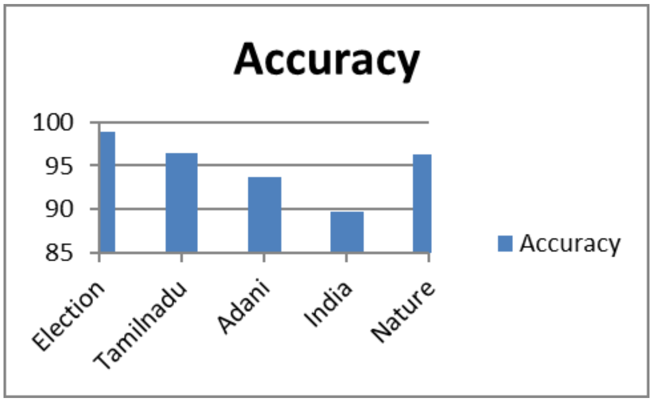

Figure 12: IBADL BA performance

of IBADL on BA, listed in Table 4, on various topics was automatically categorized. IBADL performance is depicted in Figure 9. The performance measures used are positive rate, negative rate and F measure are given below.

Positive Rate = TP/(TP + FP)

Negative Rate= TP/(TP + FN)

F-measure = 2*Precision*recall/(Precision + recall)

Accuracy = TP + TN /(TP + TN + FP + FN)

The trending topics at the time of downloads were upcoming elections in India and the party positioning in Tamil Nadu, Figure 12 has depicted the accuracy of IBADL BA analysis in classifying the downloaded tweets. Since the data (tweets) were downloaded and processed for positive, negative and neutral tweets, the accuracy was also cross-verified with the actual tweets generated before processing by IBADL.

Figure 13 illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) across various classification thresholds. A diagonal line from the bottom left to the top right (the line of no discrimination) indicates a random classifier with an area under the curve (AUC) of 0.5, implying no predictive power. A curve above the diagonal line suggests better-than-random classification performance, with a higher curve indicating superior performance. The AUC quantifies the overall performance of the classifier, with a value closer to 1 indicating better discrimination ability. In interpretation, a larger AUC suggests that the model is

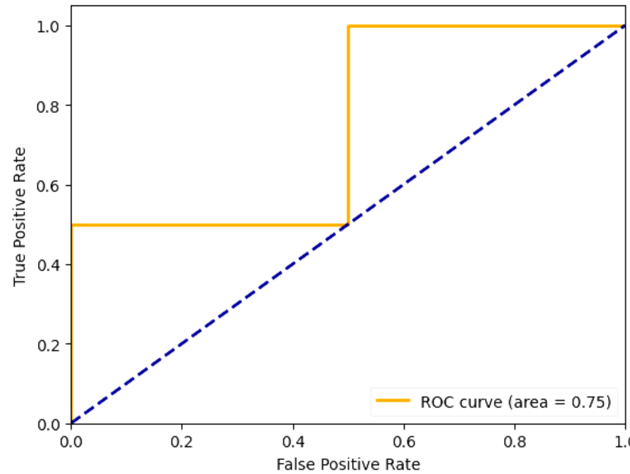

Figure 13: Receiver operating characteristic (ROC) curve

better at distinguishing between positive and negative classes. ROC curves are instrumental in comparing and selecting the most suitable classification models, aiding in the understanding of the model's predictive power and performance across different classification thresholds.

Conclusion

This paper has proposed and demonstrated with figures and tables the sentiment analysis of Twitter data using IBADL. The system establishes a relationship between users and ranks the most popular topics in social media and micro-blogging platforms with the number of tweets and retweets on the basis of the data provided and clusters them according to their intent. It can be concluded that IBADL is viable and implementable. There is, however, scope for further development regarding sentiment analysis and lexical analysis in areas of rhetoric tweets and sarcasm, which we shall further explore and develop on the basis of the current study conducted.

References

Anshari, M., & Almunawar, M. N. (2021). Adopting open innovation for SMEs and industrial revolution 4.0. *Journal of Science and Technology Policy Management*, *13*(2), 405-427.

Anspach, N. M., & Carlson, T. N. (2020). What to believe? Social media commentary and belief in misinformation. *Political*

Behavior, *42*(3), 697-718.

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, *3*(1), 43.

Ghanem, B., Rosso, P., & Rangel, F. (2020). An emotional analysis of false information in social media and news articles. *ACM Transactions on Internet Technology (TOIT)*, *20*(2), 1-18.

Ghermandi, A., Camacho-Valdez, V., & Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism Management*, *77*, 104002.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from# Brexit and# USElection. *European Economic Review*, *136*, 103772.

Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Annals of Operations Research*, *308*(1), 177-213.

Gurdgiev, C., & O'Loughlin, D. (2020). Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty. *Journal of Behavioral and Experimental Finance*, *25*, 100271.

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social science computer review*, *38*(1), 10-24.

Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*, 30-50.

Jena, R. (2020). An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. *Industrial Marketing Management*, *90*, 605-616.

Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia tools and applications*, *80*(8), 11765-11788.

Keles, B., McCrae, N., & Grealish, A. (2020). A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International journal of adolescence and youth*, *25*(1), 79-93.

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of marketing research*, *57*(1), 1-19.

Long, C., Lucey, B. M., & Yarovaya, L. (2021). I just like the stock" versus" fear and loathing on main street": The role of reddit sentiment in the GameStop short squeeze. *SSRN Electronic Journal*, *31*(04), 2021.

Mariani, M., & Baggio, R. (2022). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, *34*(1), 231-278.

Meier, A., & Reinecke, L. (2021). Computer-mediated communication, social media, and mental health: A conceptual and empirical meta-review. *Communication Research*, *48*(8), 1182-1209.

Moran, G., Muzellec, L., & Johnson, D. (2020). Message content features and social media engagement: evidence from the media industry. *Journal of Product & Brand Management*, *29*(5), 533-545.

Oh, S. H., Lee, S. Y., & Han, C. (2021). The effects of social media use on preventive behaviors during infectious disease outbreaks: The mediating role of self-relevant emotions and public risk perception. *Health communication*, *36*(8), 972-981.

Onan, A., & Toçoğlu, M. A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *Ieee Access*, *9*, 7701-7722.

Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: statistical mechanics and its applications*, *540*, 123174.

Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, *118*(26), e2024292118.

Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2021). Using data mining techniques to explore security issues in smart living environments in Twitter. Computer Communications, 179, 285-295.

Sharma, P., Berwal, Y. P. S., & Ghai, W. (2020). Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. *Information Processing in Agriculture*, *7*(4), 566-574.

Singh, S., Chauhan, A., & Dhir, S. (2020). Analyzing the startup ecosystem of India: a Twitter analytics perspective. *Journal of Advances in Management Research*, *17*(2), 262-281.

Sun, H., Burton, H. V., & Huang, H. (2021). Machine learning applications for building structural design and performance assessment: State-of-the-art review. *Journal of Building Engineering*, *33*, 101816.

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, *38*(1-2), 98-139.

Zhao, Y., Da, J., & Yan, J. (2021). Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. *Information Processing & Management*, *58*(1), 102390.