**Twitter Sentiment Analysis Using Machine Learning And Deep Learning**

Bedadala Akhileswar Reddy, Gopireddy Someswara Reddy, Kollepara Lokesh, Dr. Murali K

*Department of Computer Science and Engineering,  
Amrita School of Computing, Bengaluru,  
Amrita Vishwa Vidyapeetham, India.*

***Abstract***— ***In recent times social media platforms serve as the main source for communication, especially on public relations or on any*** ***economical crisis. During such situations, many organizations depend on tweet conversations on X platform (earlier Twitter), to know the public sentiment and reactions, and provide responsive strategies. This research mainly focuses on using machine learning and deep learning techniques for analyzing the tweet conversations on PR. The novelty in this research is to use the power of natural language processing (NLP) techniques, to analyze every word and its occurrence to know the semantic meaning and understand the sentiment of that conversation. The proposed methodology begins with preprocessing the conversation data, building deep learning models namely LSTM, BiLSTM and machine learning models like logistic regression, Naive Bayes, SVM and XGBoost, which is followed by evaluation through certain metrics. This study helps in providing automated tools for improving the organizations to know the public sentiment during crisis and respond as fast as possible with effective strategies to address the public needs.***

Keywords— Crisis management, Natural Language Processing (NLP), Sentiment analysis, Machine learning and deep learning.

# Introduction

Social networks bring the world closer and easier to be connected with the others by offering people the platform for communication, discussions, and mutual reflections. But this greater connection has also revealed a darker side of human contact: people construct types of racism and hate speech in internet dialogues. The role of the identification and deactivation of these hazardous elements will be determined by the rise the rate of digital communication.  
Social media is the platform where the war is occurring and there is the presence of every kind of obstacle. As the subject takes on the ground on the site X (Twitter) at a very high speed, the traditional content management strategy is no longer able to see it. The racism and hate speech detection from various multilingual, culturally different user-generated content can become fully complicated to handle due to the ambiguity involved in language, context, and cultural norms shifting [1]. Nevertheless, in addition to arising-from complexity, to be precise, the massive size of data being generated day by day by social platforms is another major bottleneck to manual moderation operations.  
  
The technology of X messaging can be found during discussions on public communication, the most controversial or popular social and political topics in its framework. Interactions during the debates can cover the area as if there are many participants leading to disagreements or the memes of the racist and hateful content from the platform. Companies prop up Twitter chats on the regular, such as when they are in the limelight or trying to recover from bad PR, to know what the public thinks and how the episode is affecting them.

Organisations can identify and remove hateful language as part of their robust crisis communication strategy by including hate speech detection procedures in their crisis communications plans [2]. Possible models can, for example, pick out certain offensive language over the span of tweets to make companies react faster by analysing tweet content in real-time, thus leading to proactive resolution of mentioned issues.

These tactics place the emphasis on community standards for courteous discussion, limit the pollution by the early identification of and ultimate resolutions of the issues of hate speech, thus protecting the integrity of the medium. Moreover, such tools, however, serve for raising the awareness and encouraging cheerful communicative culture based on caritative and humane discourse by letting the users know in time that their language is not appropriate.

A measurable solution to these problems is to utilize machine learning (ML) and deep learning methods to detect racially prejudiced and hateful words by means of an automated system [3]. Particularly those that make use of recurrent neural networks (RNNs), such as Bidirectional LSTM (BiLSTM) or Long Short-Term Memory (LSTM), show the way to view text and to find its peculiarities. Also few machine learning models like naive bayes, decision tree and random forest can work on text data [4]. These models can be trained to distinguish between harmful and innocuous messages, among others, and bot themselves detect and classify them based on the correction datasets including examples of racism and hate speech.

This elaborate job of recognition of hate speech signals in the X messages is decent for LSTM and BiLSTM networks due to their inherent capability to consider long-term dependencies and context [5]. These algorithms are very good at comprehending language subtleties, historical facts, and various changes of language usage patterns which could curse racists thoughts and acts. By bidirectional they mean the ability to interpret message context which is again raising the accuracy of classification. It is enabling them to consider both the past and future sentence together in the understanding of the context.

Altogether, deep learning and machine learning models for hate speech identification present businesses with robust tools to cope with crisis during brand expressions on Twitter [6]. With such tools, businesses can guard their stakeholders, maintain their image, and receive recognition for what they do in difficult moments like these.

The key contributions of the paper are:

1) The article highlights the utility of LSTM and BiLSTM networks for hate speech detection as they enable the networks to recognize the long sequences and the words in the contexts which have not been seen before. These models are known for capturing grammar and context clues, thus, made classification rather more exact.

2) The result which is obtained, is compared with other traditional machine learning algorithms like Logistic Regression, Naive Bayes, SVM and XGBoost.

The article is organized into several subsequent parts as is explained below. There is a section 2 that attempts descriptions of the research. In section 3, the data is described. In section 4, the exploratory data analysis and the methodology used are explained. Section 5 gives the analysis and results. This article finishes with section 6 conclusion and future works.

# Related works

The investigation by Bhavani M et al. [7] focuses on perception of participants through Twitter data mining. The paper proposes a deep-learning system of multi-layer structure, for instance Embedding, CNN, and LSTM, which can precisely categorize emotions. Moreover, systems like SVM, ANN, and text stemming are among the tools used for cleaning and filtering the data. The model has turned out to be very accurate, having an accuracy rate of 86.33% for the training, 79.61% for the validation, and 79.73% for the testing sets.

Another innovative approach by Jazib A et.al [8] utilizes a group classifier consisting of Naive Bayes, SVM and LSTM to sentimentalize Twitter data. Utilizing technique of feature extraction as Count Vectorizer and TF-IDF it exceeds the baseline classifiers and outperforms with a high F1-Score ratio of 0.77 for both positive and negative emotions. The method gives real time decision-making solutions for entrepreneurs due to Twitter analytics, which examines the popular company ideas and technology, producing an overall accuracy rate of 77%. The (system) efficiency emphasizes its ability to enhance social media-based analytics and decision-making processes.

Another study by M Sindhuja et al. [9] on evaluating Naive Bayes classifiers and text preprocessing techniques applied to Twitter data aiming at perception of the public opinion and the polarization of thoughts. The system utilizes text preparation techniques and Naive Bayes classifiers that allow it to classify attitudes in the given Twitter dataset. Results are in form of 77% accuracy using the multinomial Naive Bayes classifier and the count vectorization approach.

Siswanto B et al. [10] proposed methodology studies the city sentiment on Java Island ‘s gastronomic data through Twitter by applying Sastra Wi text mining library for tokenisation, indexation and other Indonesian language text mining functions. Python and Twitter API are used for the collecting of data with displayed sentiment that was mostly positive across Jakarta, Bandung, Yogyakarta, and Surabaya posting compared values of 54%.

L Sandra et al. [11] proposed model surveys the sentiments within Twitter satisfactorily through the option of the use of the support vector machine (SVM) algorithm. The framework differentiates between attitudes as positive, negative, and neutral and can reach a 62% accuracy rate. Outcome accords with the common denotation of the mudik keyword, as evidenced by the text engagement found in the tweets with neutral emotion. Yet it appeared that some tweets which were labelled as mean remarks were not considered in this way. Maybe this finding is because the researcher has included more information in Malay language.

Another approach by Varun Pininty et al. [12] a study of a new classifier modelled by machine learning algorithms overlapped with Twitter dataset that are aimed to detect sadness, with Random Forest being the one that achieved the highest accuracy of 96.52%. The study advocates a scoring approach like confusion matrix to evaluate outcomes of machine learning by detecting false positives. Beyond the astounding results, rests the advancement of using multilingual tweets and looking for the models beyond the Naive-Bayes and LSTM, leading to the superior findings.

Bac Le et al. [13] proposed methodology deals with Twitter Sentiment Analysis, with a special focus on Machine Learning approaches to data pre-processing, feature selection, and classification. SVM, Logistic Regression, and Naive Bayes belong to the classifier category, with TF-IDF used for feature extraction. Logistic Regression is an effective classifier that we found using Sentiment140 dataset and basic language pre-processing procedures. The Voting Classifier works well enough, and there are some suggestions on how to improve the trigram feature. Generally, the article covers important basics of sentimental analysis and real-world implementations of social media data processing.

M Tetteh et al. [14] proposed model deals with sentiment analysis on Internet movie reviews is the main subject of this research. It deals with the detection techniques and tools applied for this purpose. Datasets from IMDB and Rotten Tomatoes are expected to be used to classify the attitudes as Positive or Negative using Polarity scores. Text Blob's Naïve Bayes analyzer would probably TextBlob's Naïve Bayes analyzer would be the best of the lot, followed by lexicon-based tools such as VADER Sentiment Intensity Analyzer and SpaCy. The project proceeds with data pre-processing which involves cleaning and preparing the data, and further, it involves research in both supervised and unsupervised machine learning algorithms.

The investigation by Meghana Bl et al. [15] deals with the sentiment analysis done during the COVID-19 second wave in the Indian sub-continent by utilizing Twitter data. By using a hand annotation of two-weeks data and applying advanced machine learning models such as Support Vector Machine and Linear Regression the study emphasizes the importance of methods like data preprocessing and feature extraction using Count vectorizer, TF-IDF, and BBERT. The findings attest the AWS library competence to deliver annotations closely in line with hand annotations, and provide large datasets, technique details and insights that narrow down the literature gaps while building the ground for sentiment analysis research in COVID-19 epidemic.

B Venkatesh et al. [16] proposed model points out the necessity of sarcasm detection in sentiment analysis and natural language processing, mainly in the context of sarcasm recognition in real time on Twitter. The implementation utilizes ensemble approaches like Stacked Generalization and Boosting to employ machine learning algorithms SVM, Random Forest, KNN, and Logistic Regression. Notably, Stacked Generalization comes up with the best results getting an unexpectedly high 97% accuracy and detection rate.

Another approach by Divya Udayan J et al. [17] examines utilization of social media comments for predicting cybercrimes and affective use of both supervised and unsupervised learning algorithms for comment classification. The study employs preprocessing, feature extraction, and classification with LSTM being applied in the input process, all leading to the repetitive correlation between the model's accuracy and the potency of crime predictions. The project will intertwine ML techniques and NLP methodologies to maximize law enforcement agencies' ability to leverage social media data for crime prevention activities.

G Devanathan et al. [18] proposed methodology appreciates the importance of the strong infrastructural framework capable of handling numerous languages while focusing on the multilingual sentiment analysis in Indian Twitter. Embrace several approaches, including machine learning and lexicon-based systems, and delve into topics such as code-mixed texts and linguistic peculiarities. The methodologies consist of developing a dataset, preprocessing, and modification of the models indicating a great efficiency of multilingual models in sentiment recognition and in public opinion analysis. Research breakthroughs were made in different areas like text mining algorithms, social media monitoring, crisis management, and political analysis.

S Saha et al. [19] proposed model imparts the capability to differentiate between the former and fake news by encoding tweets with time through the time convolution and pooling operations. This becomes particularly valuable during pandemics, when the distinction between fact and fiction becomes blurry due to abundance of information in the social media platforms. The authors recommend the use of both techniques massive tweet volume and ambiguous information identifiers, and they suggest enhancing model precision by using data from global health agencies.

The investigation by Amrutha B et al. [20] is confined to the sentiment analysis of music reviews and lyrics by using SVM, logistic regression, Naive Bayes, and Random Forest models as machine learning methods. It demonstrates vital achievements in sentiment analysis of lyrics and reviews, particularly, the quickening of recommendation system accuracy and the detection of consumer perception. The paper proposes the potential by using deep learning architectures such as LSTM and GRU models for future development.

Michael Cai et al. [21] proposed model is focused on sentiment analysis on tweets using VD-CNNs and BERT model by Google, succeeding on Sentiment140 dataset. Two topologies are explored: A shallower design incorporating two convolutional blocks or a deeper design with four blocks that use max pooling and fully connected layers. K. Hulliyah et al.[22] studies on the strength of deep convolutional neural networks in binary sentiment categorization of Twitter data is proven, with the 2 Convolutional Block model being the most reliable and effective model. Moreover, two algorithms, K-NN and Lexicon-based, were employed to do the data analysis, resulting into good results with great precision for different k-values. This shows how sentiment analysis can be done to measure the quality of TV shows.

Sentiment analysis using Twitter data reveals the evidence of standardization and methodology comparison tactic which is a prominent area as far as research is concerned. Each study development makes use a numerous models: examples are mentioned – deep learning algorithms, ensemble classifiers, Naive Bayes classifiers, and SVM algorithms – and different methods of text processing are used such as tokenization, stemming and a few are explicitly mentioned like Count Vectorizer and TF-IDF. Most studies incite that the model performances are reaching the ceiling of 85% correct prediction rate, whereas it is noted that deep learning-based models can achieve a maximum of 89% correctness. To address the research gap we have added the data preprocessing steps along with models LSTM and BiLSTM models integrated with additional layers resulting in an increase accuracy to 96%.

# Dataset Description

Sentiment140 dataset has 16 million tweets gathered from X API with annotation by two polarity labels i.e. negative (0) and positive (1) sentiment. Structured in CSV format, the dataset includes six fields: tweet polarity, unique tweet ID, time of tweet posting, associated query, user handle, and tweet body without emoticon deleted. Through an equal distribution of 80,000 negative and 80,000 positive tweets, the research data is a crucial tool that provides a way for doing sentiment analysis work, allowing for the trend analysis of different companies, products or subject areas on Twitter.

# Methodology

The steps involved in the proposed methodology are represented in Fig 1. and detailed explanation is given below.

Data set

Data Preprocessing

Data Visualization

Data Splitting

Model building

Model Training

Model Evaluation

Pie chart of class composition

Word cloud

Hate & non-hate words Bar graph

x

LSTM  
BiLSTM  
Logistic Regression

Naive Bayes

SVM

XGB Classifier

x

x

Target and Features

Sampling Data

x

x

Fig1: Flow diagram representing the proposed methodology

1. Reading the Dataset:

Install libraries using pip to install all necessary libraries. Functions from these libraries are used in data pre-processing, model development and training phases. Read the dataset into memory via Pandas; you may need to specify the col-names too. This way, it can be analyzed or further processed smoothly.

1. Selecting Target and Features:

Select the columns you feel are most appropriate from the data set to use as the target variables (labels) and feature (input data) during sentiment analysis. The target variable here is the sentiment label ('target') while the feature is the tweet texts ('text').

1. Sampling Data:

In cases where data sets are skewed such as in most emotion analysis tasks, equalize class distribution by randomly choosing the same number of examples from any class thus avoiding directional systematics in the model toward towards the majority class.

1. Data Visualization:

Plotting the relevant plots such as pie charts, bar graphs or histograms in order to view the distribution of classes within your dataset which enables to know how important each class is relative to other classes . There is equal count of positive and negative text with target labels as 0 and 4 respectively which is shown in Fig2.

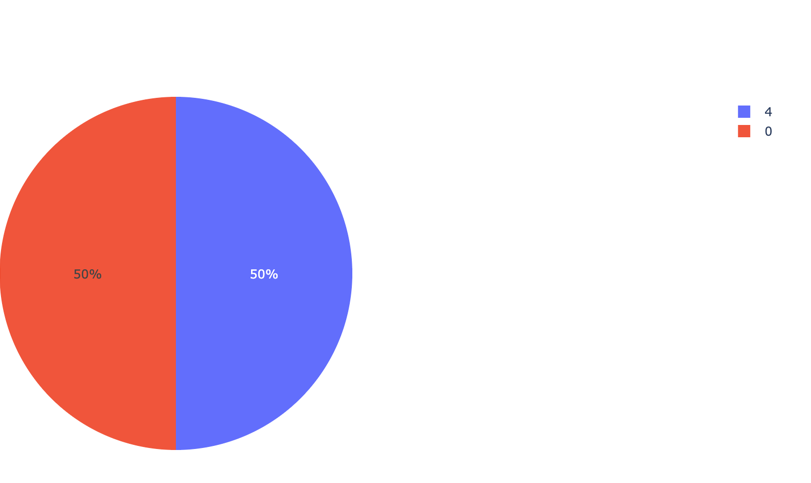


Fig2: Pie chart representing the count of target labels, where the target labels are 0 and 4 which represent positive and negative text respectively

Word clouds help to quickly pick the prominent words and identify the trends and patterns in the test data. The following Fig3 and Fig4 shows the word cloud representation of positive and negative words from the chosen data set which are used to differentiate the keywords for sentiment analysis.

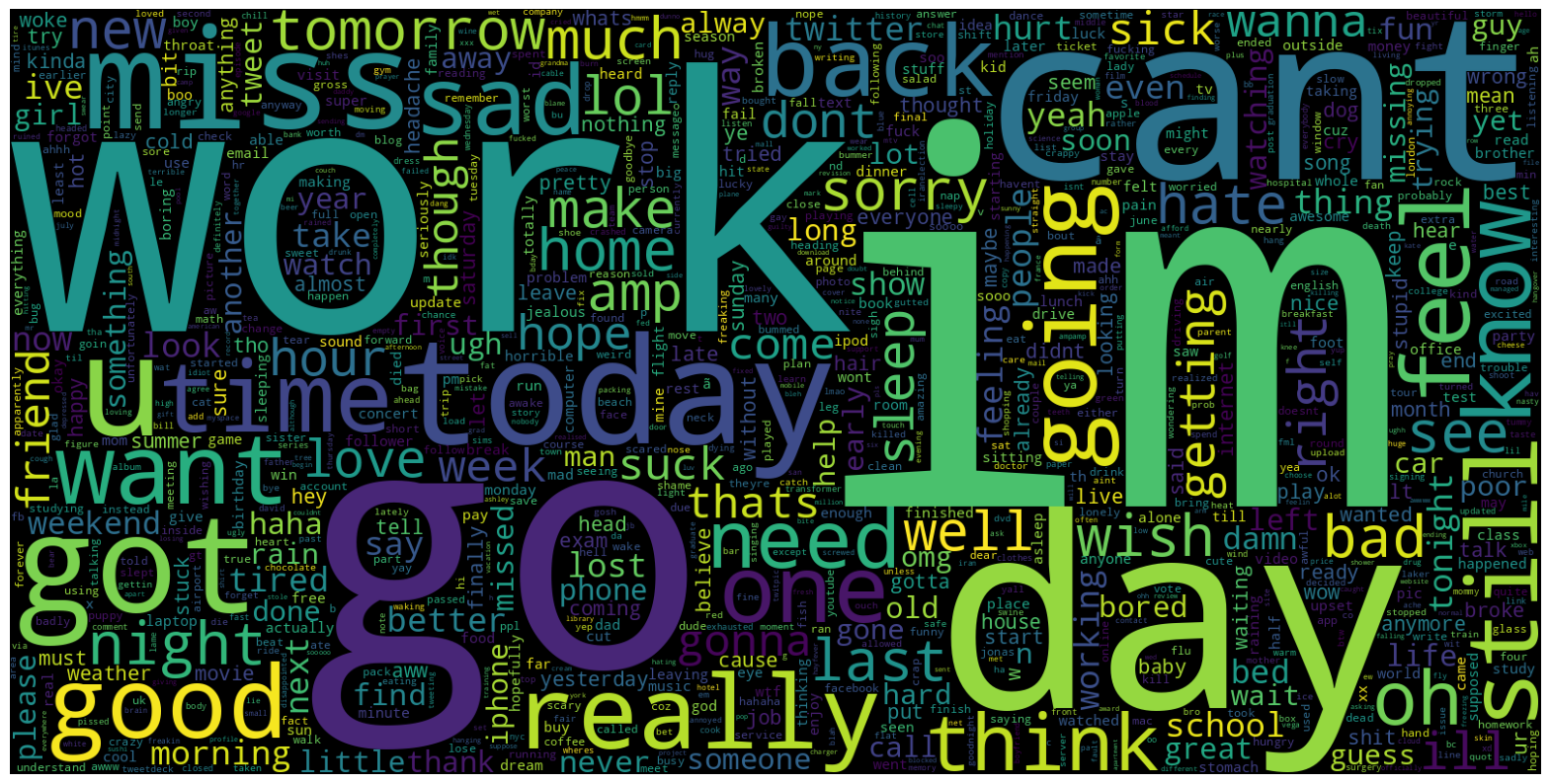


Fig3: Word cloud representation of negative words

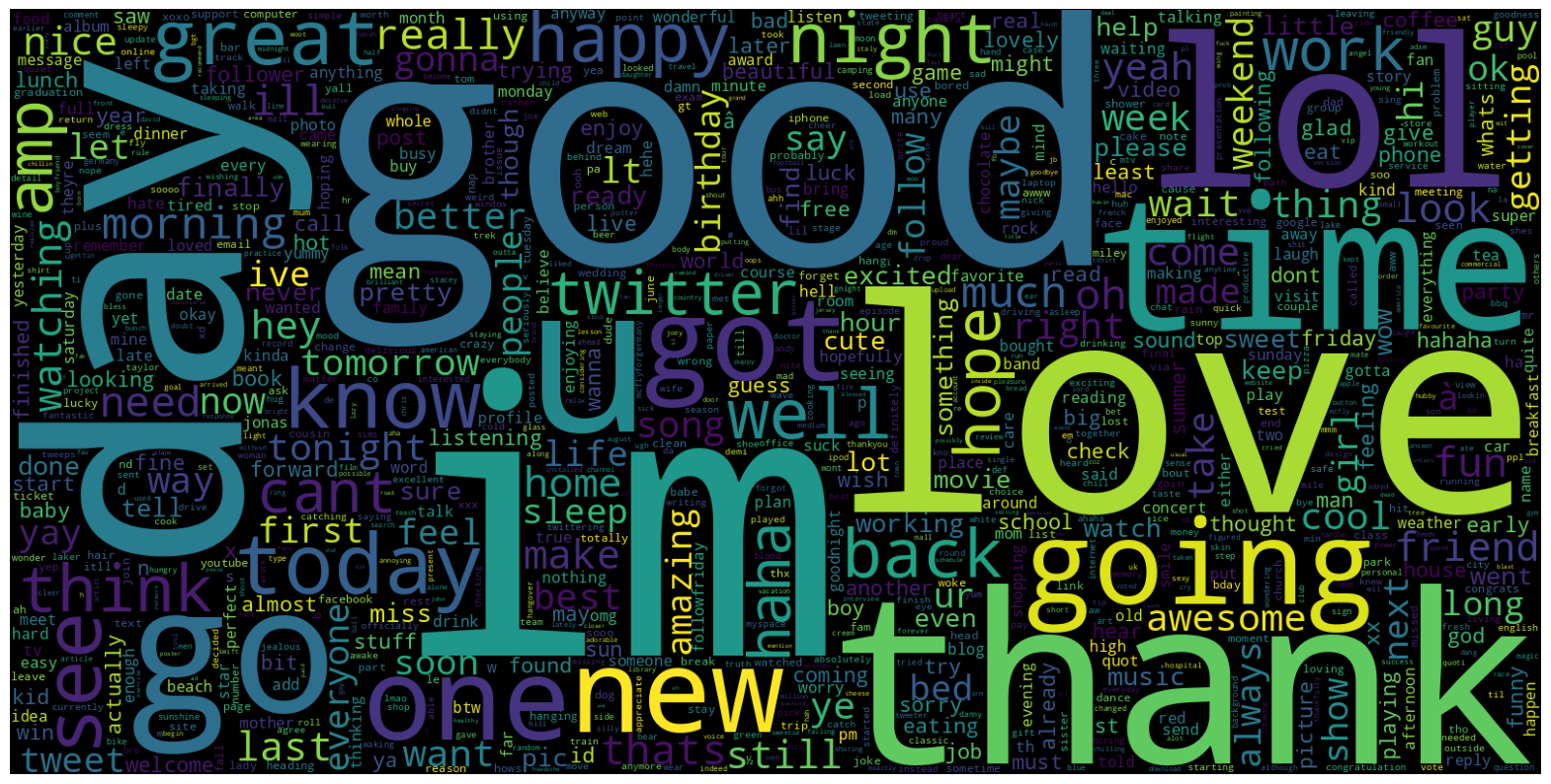


Fig4: Word cloud representation of positive words

Hashtags are very common on any social media platforms, where they are the key identifiers of any set of trending topic/social debate. So the common hashtags in regular tweets and hate tweets are shown in Fig5 and Fig6. People search for hashtags to find the relevant tweets, which is used in the research to categorize the tweets and isolate them form further analysis, so that the focus lies on specific aspects of tweets only.

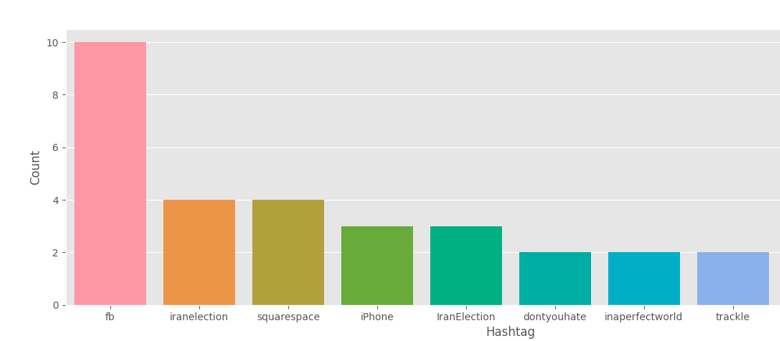


Fig5: Bar graph representing the most frequent regular hashtags

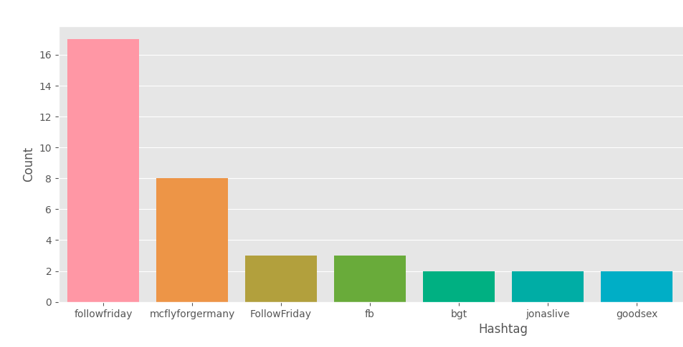


Fig6: Bar graph representing the most frequent hate tweets hashtags

1. Data Pre-processing:

The preprocessing steps mentioned in Fig 7 are explained below.

Removing

Tweet

Lower

Case

Stop word removal

Tokenization

Lemmatization

URLs

Punctuations

Numbers

Emails

Data Preprocessing

Fig7: Flow diagram representing of the data pre-processing steps

1. Lowercasing: Change all text data to lowercase so that words represent themselves consistently.
2. Stop words Removal: Remove stop words for sentiment analysis purposes since they do not contain any meaningful content such as 'and', 'the', 'is'.
3. Removing Emails and URLs: Eliminate email addresses and URLs from the text data, as they are not relevant for sentiment analysis.
4. Removing Punctuations: Strip off punctuation marks from the text to simplify text processing.
5. Removing Numbers: Omit numerical symbols from the given data as they may not contribute to sentiment analysis.
6. Tokenization: Splitt words and phrases within sentences apartfor further processing.
7. Lemmatization: Modify the words to maintain one standardized form for more accurate presentation in paper written materials or speeches.
8. Data Splitting:

The dataset should be divided into two sets - training set and test set. An ordinary 70-to-30 ratio is usually taken as an example for most cases with most of the data set aside for training purposes while a small portion allocated for testing purposes.

1. Model Building:

LSTM Model: A sequential LSTM model is built using deep learning frameworks such as TensorFlow or Keras for binary classification tasks. The architecture consists of input layer followed by embedded layer where the input is converted into dense vectors, the next comes LSTM layer, where the output of embedding layer is fed to capture dependencies and relationships between the words. The fully connected layers perform non linear transformations where activation functions like ReLU is applied on the output of fully connected layers. Finally in the output layer a single output neuron is produced which is passed through sigmoid activation function to give the probability class as either 0 or 1.

BiLSTM Model: The LSTM model is extended to a bidirectional LSTM model by adding two bidirectional LSTM layers with dropout regularization, followed by dense layers with ReLU activation function. The input to this model is separately tokenized and converted into sequence of indices. To ensure uniform length, the tokenized words are padded so that the neural network model performs well on fixed dimensions. There is a sigmoid activation function at the final dense layer for binary classification. Then the model is compiled for training.

An LSTM Based Fine-tuned Model: Different hyperparameters and model architecture experimentation are done to find the best model that suits the sentiment analysis requirements. Here the architecture has a combination of sequential model containing input layers, embedded layers, a bidirectional LSTM layer by modifying the number of units and dropout rate, followed by dense layers and output layer.

Examine different ML models: logistic regression, Naive Bayes, SVM and XGBoost were all evaluated in terms of their performance on the text data. The comparison with the deep learning models aims at understanding the relative strengths and weaknesses that these models have for sentiment analysis tasks.

h. Model Training:

The models are fed with training data, which are then optimized by relevant testing algorithms like RMSprop and hyperparameters. Ensuring that the learning process is ongoing by monitoring the training epochs while optimizing model parameters for better results.

i. Model Evaluation:

The LSTM model is compiled using binary cross entropy loss function and its performance verification is measured by using validation sets that have evaluation criterions such as loss, accuracy, precision and F1-score. Similarly the BiLSTM model is compiled, but by using Adam optimizer to ensure adaptive learning rate. Followed by compiling the fine tuned LSTM model and the machine learning models.

# Results and Analysis

The evaluation metrics and results for three different types of models: The LSTM, BiLSTM as well as standard machine learning model including Logistic Regression, Naive Bayes, SVM and XGBoost schemes are provided below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithms** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| LSTM | 96.30 | 92.67 | 90.73 | 91.68 |
| BI-LSTM | 95.94 | 93.44 | 91.49 | 92.45 |
| CNN | 95.59 | 91.84 | 94.76 | 92.00 |
| RNN | 95.59 | 92.38 | 98.31 | 96.54 |

Table 1: Performance Metrics of DEEP LEARNING Classifiers

For LSTM and BiLSTM:

* With the LSTM training yielding an accuracy of 96.30%, the model is ready for deployment. The training and testing loss of LSTM model are plotted in Fig 8. The results were 92.67%,90.73%, and 91.68% for precision, recall, and f1-score.

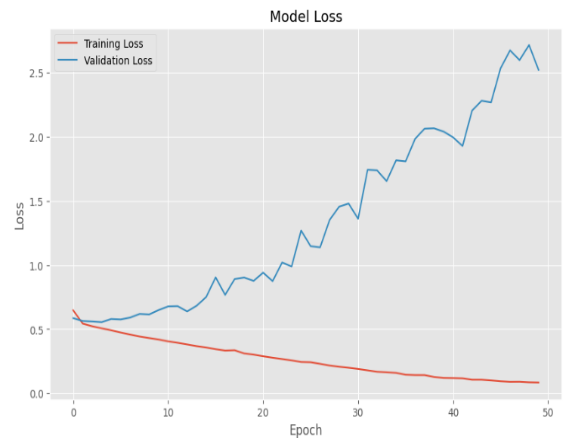


Fig8: Plotting training loss and testing loss

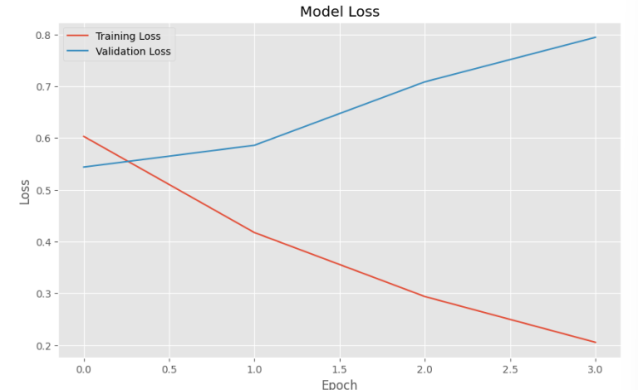
of LSTM Model

Fig9: Plotting training loss and testing loss of BiLSTM Model

* BiLSTM got a training accuracy of 95.94%.The training and testing loss of LSTM model are plotted in Fig9. BiLSTM accuracy average is 95.94%; Precision is 93.44%; Recall is 91.49%; F1-score is 92.45%.

For traditional machine learning models:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithms** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| Logistic Regression | 72.18 | 70.33 | 74.76 | 72.48 |
| SVM | 71.81 | 68.91 | 77.38 | 72.90 |
| Bernoulli Naïve Bayes | 72.41 | 71.08 | 73.67 | 72.35 |
| XG Boost | 70.46 | 66.15 | 81.32 | 72.96 |

Table 2: Performance Metrics of MACHINE LEARNING Classifiers

* Logistic regression obtained an accuracy of 72.18% as well as precision and recall values of 70.34%, 74.76% and 72.48% respectively.
* The results of the Bernoulli Naive Bayes Classifier are 72.42% in terms of accuracy, 71.09% for precision, 73.67% for recall and 72.36% for F1-scores respectively.
* SVM resulted in an accuracy score of 71.82% where it moved to higher values obtained as 68.92% for precision, 77.38% for recall, and 72.90% for F1-score.
* XGBoost achieved an accuracy 70.47% along with the values of the other parameters precision 66.16%, recall 81.33%, and F1-score 72.96%.

The accuracies of all models are plotted and compared to one another in Fig 10. Results from this experiment suggested multiple alternatives for sentiment analysis on the X dataset which can be solved automatically. LSTM and BiLSTM have been proved good for the task. The models both obtained the best accuracy of 96.3% and 95.94% respectively on the validation data.

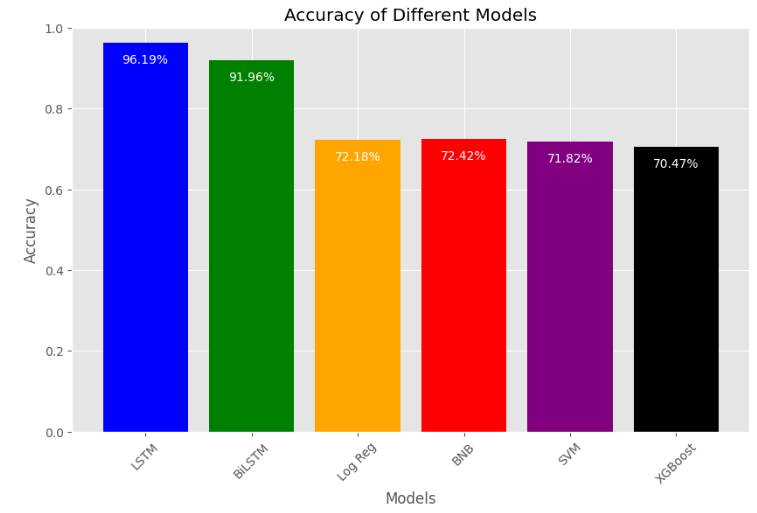


Fig10: Bar graph representing the accuracies of all the ml and deep learning models implemented

The performance of traditional machine learning models too was at par with that of deep learning models, reaching accuracies between 70.47% and 72.42% for Logistic Regression, Naive Bayes, SVM, XGBoost. Although LSTM models demonstrated better results, LSTM technology remains a relevant tool bringing affordable functionality with very simple implementation.

# Conclusions and Future Scope

Finally, our research demonstrated the applicability of a broad range of models in solving a wide range of sentiment analysis tasks of the X dataset. Both LSTM and BiLSTM networks managed to obtain amazing results but using classic machine learning methods turned out to be equally good. This pretty much stresses that a combination of different research methods is a perfect approach to get the sentiment patterns on social media platforms like X.

Moving on, further research should explore hybrid models including a deep learning and classical machine learning algorithm. Therefore, the hybrid models have these potentials to better lead to the increase in the sentiment analysis by combining the qualities of each approach. Along with it, there could be a need to explore more advanced preprocessing approaches and utilize data of a great size to create the state-of-the-art sentiment analysis algorithms which in turn would help us study the users' sentiments and their opinions on online platforms such as X.

References

[1] Y. Malhan and S. Singal, "Sentiment Analysis of Ayodhya Verdict using Twitter," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 504-509.

[2] R. Patil, N. Gada and K. Gala, "Twitter Data Visualization and Sentiment Analysis of Article 370," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 2019, pp. 1-4.

[3] T. G. Sougandh, S. S. K, N. S. Reddy and M. Belwal, "Automated Resume Parsing: A Natural Language Processing Approach," 2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2023, pp. 1-6.

[4] H. Parveen and S. Pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm," 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, India, 2016.

[5] S. I. . Khan, S. V. . Athawale, M. P. . Borawake, and M. Y. . Naniwadekar, “Sentiment Analysis of Customer Reviews using Pre-trained Language Models”, Int J Intell Syst Appl Eng, vol. 11, no. 7s, pp. 614–620, Jul. 2023.

[6] Satish, P. V., & Belwal, M. (2022). Comparative Analysis of Machine Learning and Deep Learning Algorithms for Real-Time Posture Detection to Prevent Sciatica, Kyphosis, Lordosis. In IoT Based Control Networks and Intelligent Systems: Proceedings of 3rd ICICNIS 2022 (pp. 331-345).

[7] B. M, S. S, R. M, S. K. R and S. R, "A detailed study on sentimental analysis using Twitter data with an Improved deep learning model," 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2021.

[8] A. Jazib, W. Tariq and M. Mahmood, "Sentiment Analysis using Ensemble Classifier for Entrepreneurs based on Twitter Analytics," 2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 2022, pp. 207-212.

[9] M. Sindhuja, K. S. Nitin and K. S. Devi, "Twitter Sentiment Analysis using Enhanced TF-DIF Naive Bayes Classifier Approach," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 547-551.

[10] B. Siswanto, F. L. Gaol, B. Soewito and H. L. H. S. Warnars, "Sentiment Analysis of Big Cities on The Island of Java in Indonesia from Twitter Data as A Recommender System," 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, Jakarta, Indonesia, 2021, pp. 124-128.

[11] L. Sandra and F. Lumbangaol, "When Homecoming is not Coming: 2021 Homecoming Ban Sentiment Analysis on Twitter Data Using Support Vector Machine Algorithm," 2021 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2021, pp. 1-5.

[12] Varun Pininty . Detecting Depression in Social Media Posts using Machine Learning Models. TechRxiv. February 01, 2023.

[13] Le, B., Nguyen, H. (2015). Twitter Sentiment Analysis Using Machine Learning Techniques. In: Le Thi, H., Nguyen, N., Do, T. (eds) Advanced Computational Methods for Knowledge Engineering. Advances in Intelligent Systems and Computing, vol 358. Springer, Cham.

[14] M. Tetteh and M. Thushara, "Sentiment Analysis Tools for Movie Review Evaluation - A Survey," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 816-823.

[15] M. Bl, S. Midha and V. R. Murthy Oruganti, "Sentiment Analysis in Indian Sub-continent During COVID-19 Second Wave using Twitter Data," 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 2021.

[16] B. Venkatesh and H. N. Vishwas, "Real Time Sarcasm Detection on Twitter using Ensemble Methods," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021.

[17] D. U. J, R. M. R. S, Sravan, S. Ganesh and Vidyacharan, "Social Media Comment Classification and Crime Forecasting," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-6.

[18] G, Devananthan & Nair, Lekshmi. (2023). Exploring Multilingual Indian Twitter Sentiment Analysis: A Comparative Study.

[19] S. Saha, S. Jayan and S. R, "Tweet2Vec model to detect misinformation about the COVID-19 Pandemic on Twitter," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1624-1630.

[20] A. B. and S. M., "Recommendation of Independent Music based on Sentiment Analysis," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6.

[21] Cai, Michael. "Sentiment Analysis of Tweets using Deep Neural Architectures." Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada, 2018.

[22] K. Hulliyah, A. M. Almaisah, F. Mintarsih, S. U. Masrurah, D. Khairani and S. Aripiyanto, "Analysis of Public Sentiment Using The K-Nearest Neighbor (k-NN) Algorithm and Lexicon Based on Indonesian Television Shows on Social Media Twitter," 2022 10th International Conference on Cyber and IT Service Management

(CITSM), Yogyakarta, Indonesia, 2022, pp. 1-5