

Advanced Natural Language Processing (NLP) Techniques for Text-Data Based Sentiment Analysis on Social Media

Srinivas Chippagiri *, Savan Kumar, Olivia R Liu Sheng

¹ Department of Operations and Information Systems, David Eccles School of Business, University of Utah, Salt Lake City, UT, 84112, USA

*Correspondence: Srinivas Chippagiri (u1014717@utah.edu)

Abstract: The field of sentiment analysis is a crucial aspect of natural language processing (NLP) and is essential in discovering the emotional undertones within the text data and, hence, capturing public sentiments over a variety of issues. In this regard, this study suggests a deep learning technique for sentiment categorization on a Twitter dataset that is based on Long Short-Term Memory (LSTM) networks. Preprocessing is done comprehensively, feature extraction is done through a bag of words method, and 80-20 data is split using training and testing. The experimental findings demonstrate that the LSTM model outperforms the conventional models, such as SVM and Naïve Bayes, with an F1-score of 99.46%, accuracy of 99.13%, precision of 99.45%, and recall of 99.25%. Additionally, AUC-ROC and PR curves validate the model's effectiveness. Although, it performs well the model consumes heavy computational resources and longer training time. In summary, the results show that deep learning performs well in sentiment analysis and can be used to social media monitoring, customer feedback evaluation, market sentiment analysis, etc.

Keywords: Social Media, Sentiment Analysis, Natural Language Processing (NLP), Twitter Data, Machine Learning, Text Classification

How to cite this paper:

Chippagiri, S., Kumar, S., & Sheng, O. R. L. (2016). Advanced Natural Language Processing (NLP) Techniques for Text-Data Based Sentiment Analysis on Social Media. *Journal of Artificial Intelligence and Big Data*, 1(1).
DOI: [10.31586/jaibd.2016.1293](https://doi.org/10.31586/jaibd.2016.1293)

Received: July 26, 2016

Revised: October 19, 2016

Accepted: November 12, 2016

Published: December 21, 2016



Copyright: © 2016 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this digital age social media has become a global communication system as it provides avenues for one to express emotions, share opinions, and interact while alive. Emotions have a huge role in defining online discussions, consumer behavior, brand perception and public sentiment. Businesses, policymakers and researchers interested in analyzing and interpreting these emotions are important to be able to understand the public sentiment and emerging trends. As social media grows in an exponential way, lots of text-based content is generated every day [1]. User reviews, comments, and discussions are all part of this data that consists of rich sources of information to determine what is the preferred brand, what is the perception, and what are the societal attitudes. However, due to the nature of unstructured and the large scale of such text data, it is highly impractical to analyze such data manually, and hence automated techniques for extracting meaningful patterns and sentiments from such data become necessary [2].

As a major domain of NLP, sentiment analysis refers to classifying and interpreting the sentiments conveyed in text [3]. Current sentiment analysis techniques are based on basic word-level techniques to decide whether. The emotion of a writing might be neutral, negative, or positive [4]. Despite these, such methods often fall short of context, sarcasm and Variations in language, which hinder its effectiveness in the actual world. In order to mitigate these challenges, the sentimental analysis of social media text has been enhanced through advanced NLP techniques [5].

Sentiment analysis made by DL models, transformer architectures and contextual embeddings boosts the accuracy of classification. Modern methods that combine ML with

sentiment analysis allow for the processing of massive piles of social media data for more effective identification of sentiment polarity and for further insights into public opinion [6]. On the basis of these techniques, we offer contextual embedding, sentiment lexicon adaptation and domain-specific sentiment analysis, that help businesses and researchers to extract sensible results, optimize decision making and improve user engagement strategies.

1.1. Aim and Contribution

Sentiment analysis is a very important tool in understanding public perception of a topic, and it has become a very influential channel through which people express their opinions on many topics through social media platforms, especially Twitter. Tweet is unstructured, and so it is a challenge for sentiment classification, and thus advanced NLP and DL are needed. On the one hand, traditional ML is very sensitive to context, while sequential data is manageable on the other. The contribution of this study is to boost classification accuracy, facilitating decision-making for businesses, decision-making for researchers, or decision-making for policymakers. The main contributions are:

- Create sentiment classification models using natural language processing and the Twitter dataset.
- Implements advanced text preprocessing techniques, including filtering, tokenization, and stop-word removal, to enhance data quality for sentiment analysis.
- Utilizes the bag-of-words method to identify key attributes, improving sentiment classification accuracy.
- Employs LSTM networks to effectively capture contextual dependencies in textual data.
- Use an evaluation based on a confusion matrix that takes into account F1-score, recall, accuracy, and precision.

1.2. Structure of the paper

The study is structured as follows: Relevant work for text-data-based sentiment analysis on social media is presented in Section II. The approach, including data collection, preprocessing, and feature extraction techniques, is described in depth in Section III. The experimental findings and performance assessment are shown in Section IV. Section V concludes the study and summarizes key findings.

1.3. Literature Review

This section reviews research articles on for sentiment analysis in social media with advanced ML algorithms and natural language processing.

Kanakaraj and Guddeti (2015) examine social sentiment toward a specific news story from Twitter postings. The mood of the mined text data is ascertained by applying ensemble classification. Ensemble classification combines the capabilities of several individual classifiers to address a particular classification problem. Ensemble classifiers outperform standard ML classifiers by 3–5%, according to experiments [5].

Chirawichitchai (2014) suggested Thai Text-Based Emotion Classification comparing many popular word weighting systems utilizing ML techniques and term weighting. I found that Boolean weighting with an SVM performs well in our experiments. With an accuracy of 77.86%, our experiments revealed that the SVM approach with the Information Gain feature selection worked best. Furthermore, our experimental results show that the Thai Emotion Classification Framework is enhanced by feature weighting strategies [7].

Hogenboom et al. (2014) transmit the target sentiment lexicon by analyzing the sentiment of seed words in a semantic lexicon for the target language. When sentiment analysis is expanded from English to Dutch, it yields a significant performance boost of

around 29% over the baseline in terms of accuracy and macro-level F1 on our data. This is achieved by mapping sentiment across languages by using relationships across semantic lexicons. Sentiment propagation in language-specific semantic lexicons can exceed the baseline by up to 47%, depending on the seed set of sentiment-carrying words [8].

Anjaria and Guddeti. (2014) used supervised machine learning methods to classify Twitter data using a feature extraction model that combined unigram and bigram, as well as ANN. The case study included the US presidential election of 2012 as well as the Indian Karnataka state assembly election of 2013. Results from experiments show that SVM are the best classifiers, achieving up to 88% accuracy in the 2012 US elections and 68% accuracy in the 2013 Indian state assembly elections [9].

Volkova, Wilson and Yarowsky At el. (2013) focus on finding gender differences in subjective language use in Twitter data related to English, Spanish, and Russian. Additionally, investigates cross-cultural variations in the usage of hashtags and emoticons by male and female users. According to our findings, the statistical significance of the relative F-measure improvement over the gender-independent baseline is established. 2.5% and 5% for English, 1% and 1.5% for Russian, and 2% and 0.5% for Spanish, according to the polarity and subjectivity study [10].

Table 1 provides a comparative analysis of different previous reviews on sentiment analysis based on the datasets, key findings, limitations, and future work.

Table 1. Summary of Sentiment Classification Techniques in Social Media Using Machine Learning

Paper	Method	Dataset	Key Findings	Limitations &Future Work
Kanakaraj and Guddeti (2015)	NLP techniques, Word Sense Disambiguation, Ensemble classification	Twitter posts on news events	Ensemble classification improves accuracy by 3-5% over traditional ML classifiers	Future work could explore deep learning models for further accuracy enhancement
Chirawichitchai (2014)	Term weighting, SVM, Information Gain feature selection	Thai text dataset	Boolean weighting with SVM achieves the highest accuracy (77.86%)	Future work can focus on expanding emotion classification for multilingual settings
Hogenboom et al. (2014)	Spreading sentiment lexicon and cross-linguistic sentiment mapping	English and Dutch language datasets	Sentiment propagation improves accuracy by up to 47%	Further research can investigate additional languages and domain-specific sentiment lexicons
Anjaria and Guddeti (2014)	Supervised ML (SVM, Naïve Bayes, ANN), Unigram & Bigram features, Influence Factor	Twitter statistics (Karnataka State Assembly Elections 2013, US Presidential Elections 2012)	SVM achieved highest accuracy (88% for US Elections, 68% for Indian Elections)	Future work can incorporate deep learning models and social influence factors for better prediction
Volkova, Wilson, and Yarowsky (2013)	Understanding how gender differs in the classification of sentiment, polarity, and subjectivity	English, Spanish, and Russian Twitter data	Gender-based language differences improve polarity classification (2.5-5% improvement in F-measure)	Future studies can explore additional cultural and linguistic variations for sentiment analysis

2. Methodology

The methodology for sentiment analysis using NLP and deep learning involves multiple stages, beginning with the Twitter dataset, which undergoes comprehensive

preprocessing, including filtering, tokenization, and stop-word removal. Feature extraction is then applied using the bag-of-words method to identify the most relevant attributes for sentiment classification. The dataset is preprocessed and then split into training and testing groups in an 80-20 ratio. For text categorization, use DL models such as LSTM networks. The effectiveness of the trained models on the testing subset is then evaluated using performance metrics like as accuracy, precision, recall, and F1-score, producing the final findings. The overall workflow of the methodology is displayed in Figure 1.

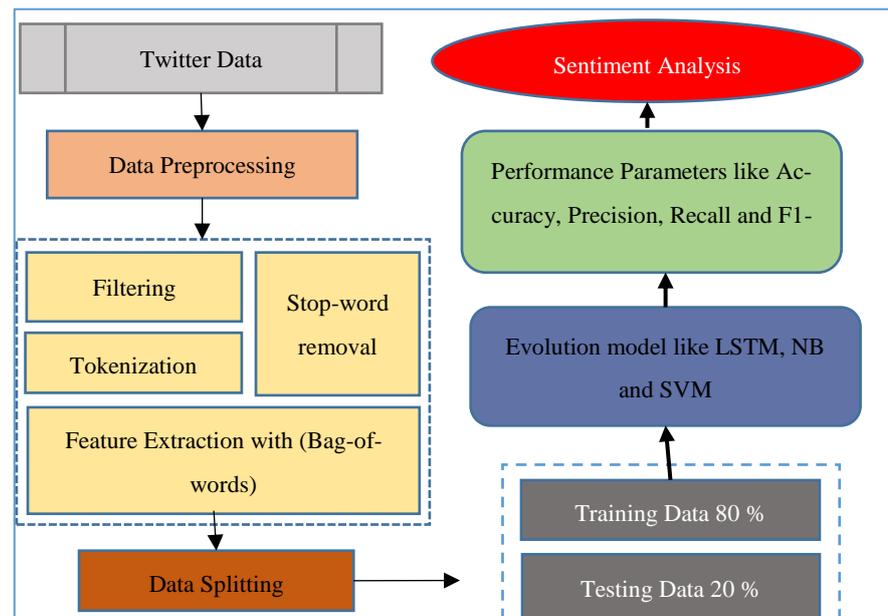


Figure 1. Flowchart for sentiment analysis

The flowchart's subsequent phases are briefly described below:

2.1. Data collection

The Twitter dataset consists of 73,000 tweets, out of which 12,000 are labeled as “irrelevant.” These include tweets in foreign languages, containing only URLs, or with unreadable Unicode characters, which are excluded from analysis. This leaves 61,000 tweets with sentiments categorized as positive, negative, or neutral, covering topics related to brands, public opinion, and social discussions. The visualization of data insight is given in below:

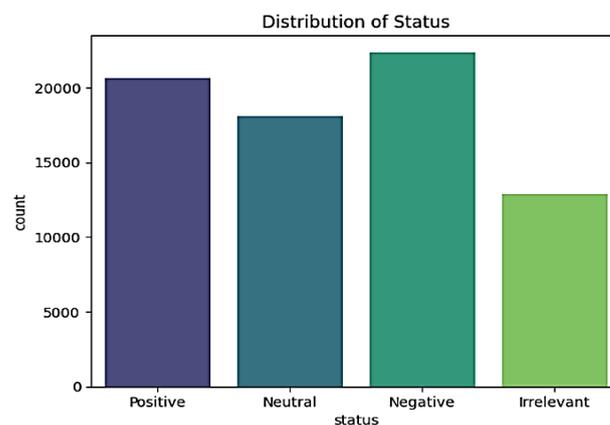


Figure 2. Count plot for Data distribution

The bar chart in [Figure 2](#) displays the distribution of 73,000 tweets into four sentiment classes: Negative (22,000), Positive (21,000), Neutral (18,000), and Irrelevant (12,000). The Negative class is the most frequent, while Irrelevant tweets are excluded from analysis. The dataset's slight imbalance may require data preprocessing and balancing techniques for optimal sentiment classification.

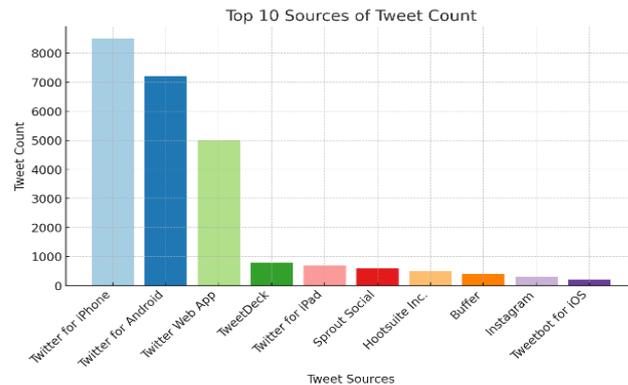


Figure 3. Top 10 sources of tweet count

The bar graph in [Figure 3](#), shows the top 10 sources of tweets, with "Twitter for iPhone" leading, followed by "Twitter for Android" and "Twitter Web App." Mobile devices dominate tweet generation, while platforms like Tweet Deck, Hootsuite, and Instagram contribute minimally. Third-party tools play a minor role, emphasizing users' preference for official Twitter applications.

2.2. Data preprocessing

Pre-processing the data lowers the computational complexity and produces text classifications of greater quality. The following stages are typical of a pre-processing procedure:

- **Filtering:** This stage involves removing the URL link, special terms on Twitter (like "RT," which stands for "ReTweet"), user names on Twitter (like "@Ron" with the @ sign next to a user name), and emoticons[11].
- **Tokenization:** Tokenize or segment text by dividing it into word containers using punctuation and spaces.

2.3. Stop-words removal

A group of terms known as "stop words"—such as a, the, I, am, and so on—are commonly employed in everyday speech. These words have no bearing on the text's sentiment or meaning, hence they are not important for the study [12]. Because stop words eliminate low-level information, our text may concentrate more on the key information.

2.4. Feature extraction with Bag-of-words

The BoW technique is the source of the feature extraction procedure. (In this instance, the text is shown as a bag of words.) The frequency with which each word occurs acts as a feature for training the classifier. Additionally, redundant and sparse data are eliminated from the original raw data to minimize overfitting of the training set and speed up algorithm execution of the reduced set of features.

2.5. Data splitting

The dataset was divided into a test set and a training set. Eighty percent of the data was in the training set and twenty percent was in the testing set to guarantee successful model validation.

2.6. Classification with Long short-term memory (LSTM)

The LSTM model is one type of recurrent neural network [13]. In order to provide a typical RNN more precise control over memory, LSTMs include more factors. These variables determine the importance of the present input in forming the new memory, the significance of the prior memories in forming the new memory, and the memory's key components in producing the output. The LSTM mathematical equations are shown below (1 to 6).

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$g_t = \tanh(W_g x_t + U_g h_{t-1} + b_g) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

The logistic sigmoid function is represented by σ in the equations above, whereas element-wise multiplication is represented by \odot . The LSTM unit has a memory cell c_t at each time step t , a hidden unit h_t , an input gate i_t , a forget gate f_t , and an output gate o_t . B stands for the additional bias, while W and U are the learnt parameters. It makes sense that the output gate regulates the amount of internal memory state that is exposed, the forget gate regulates the amount of memory cell erasure, and the input gate regulates the amount of each unit's updating.

2.7. Performance metrics

Utilize the four standard information retrieval assessment criteria listed below for the subsequent analysis stage. The confusion matrix is used in this study to gauge the model's effectiveness. It takes into account the following factors. The performance metrics contain the various parameters, which are shown below:

- **True Positive (TP):** demonstrates the amount of correctly recognized data as well as the presence of any undesirable event.
- **False Positive (FP):** indicates the number of data points that have been incorrectly labeled as the presence of any undesirable event.
- **True Negative (TN):** shows the number of records that are correctly classified as usual.
- **False Negative (FN):** shows the number of records that are incorrectly classified as usual.
- **Accuracy:** The ratio of accurately predicted values to all test cases is known as accuracy. It is computed according to Equation (7).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (7)$$

Precision: Equation (8) provides precision, which is the number of true positives among all positively assigned documents:

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Recall: Recall is determined by equation (9) and is the number of true positives among the real positive documents:

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

F1-score: The F-measure, a weighted approach to recall and accuracy, is calculated using equation (10).

$$F1 - Score = \frac{2(Precision*Recall)}{Precision+Recall} \quad (10)$$

ROC (Receiver Operator Characteristic): The performance of a binary classification model may be assessed graphically using a ROC graph. It compares the True Positive Rate (TPR), often referred to as Sensitivity, with the False Positive Rate (FPR) at various classification levels.

3. Result Analysis and Discussion

This study was conducted and assessed in an experimental environment on a PC running Windows 10 Professional (64-bit) with a Core TM i5-8250U CPU with 12 GB of RAM running at 1.8 GHz. The LSTM model for text-based sentiment analysis on social media data was implemented using Python 3 and deep natural language processing methods for classification and processing. For the same, the Twitter dataset performance is shown in below and the proposed models were trained on it.

Table 2. LSTM Model Performance for Text-based Sentiment Analysis on a Twitter Dataset

Evaluation measures	Long Short-Term Memory (LSTM)
Accuracy	99.13
Precision	99.45
Recall	99.25
F1-score	99.46

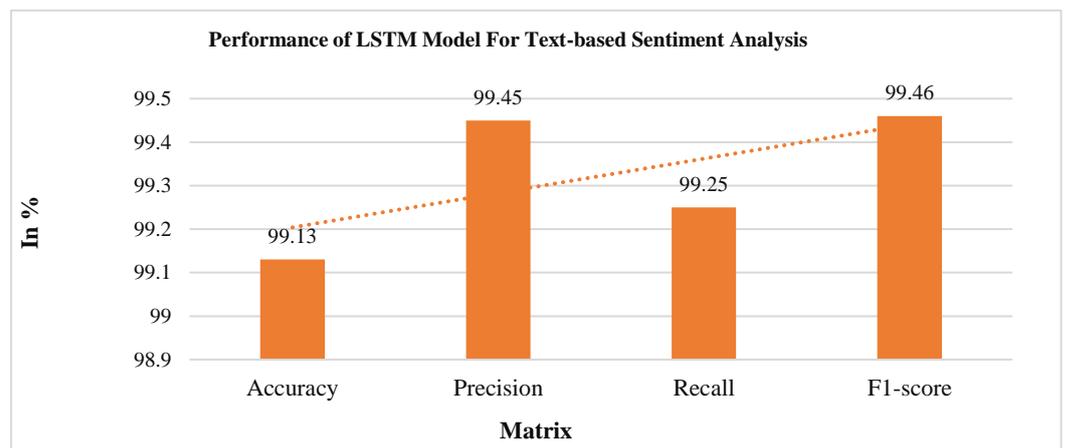


Figure 4. Bar Graph for Performance of LSTM Model

The results of the LSTM model's performance in text sentiment analysis using data from social media are displayed in Table II and Figure 4. With 99.13% accuracy, the model achieves good success since it can correctly classify sentiments. It achieves a precision of 99.45% with regard to positive sentiment prediction or a recall of 99.25% as a signal of actual sentiment detection. The robustness of the F1 score of 99.46% is further supported by the good balance between accuracy and recall. This result reaffirms the usefulness of advanced NLP techniques in processing complex social media texts irrespective of linguistic variances and context.

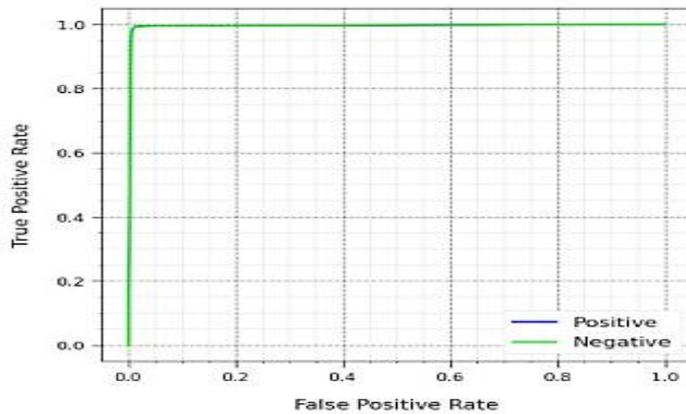


Figure 5. AUC-ROC Curve for LSTM model

The LSTM model's AUC-ROC curve provides a strong classification, as seen in [Figure 5](#). The curve is almost entirely in the upper-left corner, indicating an excellent TPR and FPR. This steep rise indicates that the classification capability is excellent, i.e., the misclassification is negligible. This indicates the high AUC score of the model, which implies it is effective in sentiment analysis.

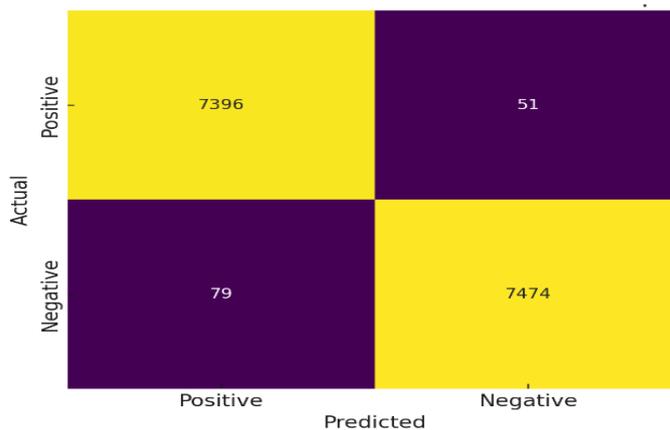


Figure 6. Confusion matrix for LSTM model

The LSTM model's high predictive potential is seen in [Figure 6](#). The model correctly classifies 7,396 positives and 7,474 negatives with very few misclassifications, as it has 51 false negatives and 79 false positives. The results show high accuracy and balanced performance, which means that the model is suitable at classifying sentiment classes with few errors.

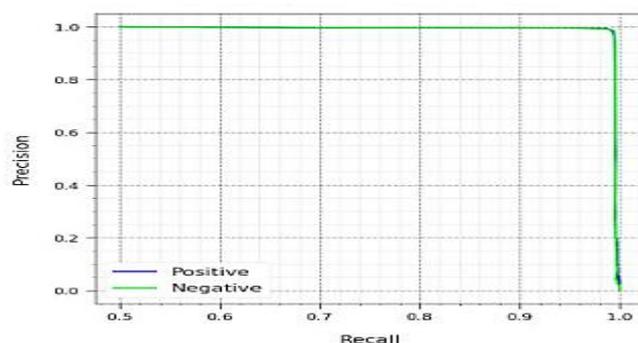


Figure 7. PR Curve of LSTM model on Twitter data

The PR curve in Figure 7 demonstrates how well the LSTM model classifies the Twitter dataset. The curve has high precision (0.98+) and recall (0.98+), meaning there are few falsely positive and negative. This validates sentiment analysis's model resilience as the steep drop close to recall of 1.0 shows that it can balance recall and accuracy.

3.1. Comparative analysis

In this section, the ML and DL models like LSTM, NB and SVM are compared for sentiment analysis and deep learning's suitability in handling contextual relationships is emphasized over traditional ones. Table III analyzes several sentiment analysis models using the Twitter dataset and demonstrates that LSTM attains a 99.13% accuracy rate. The lower accuracies of 81.30% achieved by NB and 70.33% by SVM are obvious and traditional ML models. These results demonstrate that deep learning techniques are highly efficient in performing sentiment analysis tasks through LSTM which can capture the contextual information of the text.

Table 3. Various models Performance comparison on the Twitter dataset for sentiment analysis

Models	Accuracy
LSTM (Long Short-Term Memory)	99.13
NB (Naïve Bayes) [14]	81.30
SVM (Support Vector Machine) [15]	70.33

The proposed LSTM model detects complex patterns effectively because it reaches 99.13% accuracy while outperforming classic machine learning approaches. The arrangement of words throughout the text remains untouched to deliver a high accuracy rate together with minimal erroneous classifications. The model is useful for social media research and gathering client input in real-world applications since it performs well with large datasets.

4. Conclusion and Future Work

Sentiment analysis is an automated method for identifying and comprehending the emotions expressed in a text. In the last ten years, SA's overall prevalence among NLP users has increased. SA is currently essential for businesses to obtain customer information and mold their marketing strategy because to the unavoidable usage of social media and online platforms. The study demonstrates the effectiveness of the LSTM model for sentiment analysis, achieving an accuracy of 99.13%, surpassing traditional ML models like Naïve Bayes 81.30% and SVM 70.33%. The model excels in capturing contextual dependencies in text, ensuring minimal misclassification, as evident from high precision 99.45% and recall 99.25% scores. The AUC-ROC and PR curves further confirm their robustness in sentiment classification. However, despite its superior performance, the LSTM model has limitations, including high computational costs, long training times, and potential inefficiency for real-time sentiment analysis. Additionally, it may struggle with highly imbalanced datasets or sarcasm detection. Future work should focus on optimizing computational efficiency, integrating hybrid deep learning models, and exploring transformer-based architectures like BERT for improved contextual understanding.

References

- [1] Y. Zhao, K. Niu, Z. He, J. Lin, and X. Wang, "Text sentiment analysis algorithm optimization and platform development in social network," in *Proceedings - 6th International Symposium on Computational Intelligence and Design, ISCID 2013*, 2013. doi: 10.1109/ISCID.2013.108.
- [2] M. C. Ganiz, M. Tutkan, and S. Akyokus, "A novel classifier based on meaning for text classification," in *INISTA 2015 - 2015 International Symposium on Innovations in Intelligent Systems and Applications, Proceedings*, 2015. doi: 10.1109/INISTA.2015.7276788.
- [3] L. Keri and R. T. Watson, "The impact of natural language processing based textual analysis of social media interactions on decision making," *ECIS 2013 - Proc. 21st Eur. Conf. Inf. Syst.*, 2013.

-
- [4] G. Paltoglou, "Sentiment analysis in social media," in *Online collective action: Dynamics of the crowd in social media*, Springer, 2014, pp. 3–17.
 - [5] M. Kanakaraj and R. M. R. Guddeti, "Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques," in *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing, IEEE ICSC 2015*, 2015. doi: 10.1109/ICOSC.2015.7050801.
 - [6] M. Moh, A. Gajjala, S. C. R. Gangireddy, and T.-S. Moh, "On Multi-tier Sentiment Analysis Using Supervised Machine Learning," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, Dec. 2015, pp. 341–344. doi: 10.1109/WI-IAT.2015.154.
 - [7] N. Chirawichitchai, "Emotion classification of Thai text-based using term weighting and machine learning techniques," in *2014 11th Int. Joint Conf. on Computer Science and Software Engineering: "Human Factors in Computer Science and Software Engineering" - e-Science and High Performance Computing: eHPC, JCSSE 2014*, 2014. doi: 10.1109/JCSSE.2014.6841848.
 - [8] A. Hogenboom, B. Heerschop, F. Frasincar, U. Kaymak, and F. De Jong, "Multi-lingual support for lexicon-based sentiment analysis guided by semantics," *Decis. Support Syst.*, 2014, doi: 10.1016/j.dss.2014.03.004.
 - [9] M. Anjaria and R. M. R. Guddeti, "A novel sentiment analysis of social networks using supervised learning," *Soc. Netw. Anal. Min.*, vol. 4, no. 1, p. 181, 2014, doi: 10.1007/s13278-014-0181-9.
 - [10] S. Volkova, T. Wilson, and D. Yarowsky, "Exploring demographic language variations to improve multilingual sentiment analysis in social media," *EMNLP 2013 - 2013 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. October, pp. 1815–1827, 2013.
 - [11] M. Venugopalan and D. Gupta, "Exploring sentiment analysis on Twitter data," in *2015 Eighth International Conference on Contemporary Computing (IC3)*, 2015, pp. 241–247. doi: 10.1109/IC3.2015.7346686.
 - [12] B. Alhadidi and M. Wedyan, "Hybrid Stop-Word Removal Technique for Arabic Language.," *Egypt. Comput. Sci. J.*, 2008.
 - [13] J. N. Schrading, *Analyzing domestic abuse using natural language processing on social media data*. Rochester Institute of Technology, 2015.
 - [14] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *Processing*, 2009.
 - [15] R. Soni and K. J. Mathai, "Improved Twitter Sentiment Prediction through Cluster-then-Predict Model," vol. 4, no. 4, pp. 559–563, 2015.