A comparison of Long Short-Term Memory Algorithms with Sentimental Tweets Recommendation. Lack of Efficiency in Envisioning Human Sensibility Using Novel Naive Bayes Algorithms

J Venkata Kalyan[1], Shri Vindhya.A[2*]
[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.
[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha School of Engineering, Pincode:602105.

ABSTRACT
Aim: The main purpose of this study is to use machine learning classifiers to improve the efficiency of recognising relevant people with varied emotions in order to promote emotive tweets by comparing the Novel Naive Bayes algorithm and the LSTM (Long Short-Term Memory) algorithm. **Materials and Methods:** Novel Naive Bayes object detection algorithm with sample size = 132 with G power (value=0.6), the efficiency percentage was predicted using a 90 percent confidence interval and the LSTM algorithm with a sample size of 132. The prediction is mapped using the weights and configurations of Naive Bayes. **Results:** Naive Bayes algorithm has better efficiency (87.10%) when compared to the Long Short-Term Memory (LTSM) efficiency (81.60%). The results achieved with significance value p=0.904 (p>0.05) shows that two groups are statistically insignificant.**Conclusion:** The efficiency of the Naive Bayes algorithm outperformed that of the Long Short-Term Memory (LSTM) algorithm.
Keywords: Novel Naive Bayes, Long Short-Term Memory (LSTM) algorithm, Sentiment Tweets, Tweets Recommendation, Emotions, Positive and Negative Tweets.

INTRODUCTION

The main premise behind this study is that people with similar interests share similar interests via social media tweets utilising the Novel Naive Bayes algorithm, have distinct perspectives, and long-term proposed methods. When compared to the Long Short-Term Memory (LTSM) method, the experiment's goal in this scenario is to increase efficiency and provide output values that show how positive or negative the emotions portrayed in the positive and negative tweets or neutral tweets (Salim et al. 2021) . This required a tweet-id collection containing both good and negative tweets(Sharma et al. 2020). To increase the efficiency of suggestions, content-based recommendations based on good and negative tweets, emotional states, and views retrieved from user microposts are used (McFedries 2009) . Opinion mining may be done on a variety of topics. This information is beneficial to a variety of organisations as well as political parties. That is how to understand the importance of emotions and sentiments in our existence of positive and negative tweets.

Inferring relationships between users' sentiment tweets, the novel naive bayes method is more efficient and quicker, according to the same study (Jilka et al. 2022). Researchers frequently fail to assess individual phrases in tweets, and thus sentiment analysis struggles to adequately explain context; businesses, politicians, and marketing organisations all need to be aware of current trends and themes. These are emotional statements made by tweets in a big dataset of tweets, both good and negative (Oscar et al. 2017). The usage of this method aids in the definition of a new weighting function that may be used to enrich content-based user profiles of positive and negative tweets. This approach's applications include customized tweets based on traits and sentiment tweets using emotion tweet analysis (Mogaji, Balakrishnan, and Kieu 2021).

The novel compiles data from a variety of sources in order to determine characteristics that may impact the measure of similarity. On IEEE Xplore, there were roughly 40 publications regarding emotional good and negative tweets, and 140 articles were published on Google Scholar. To select the proper profile on twitter with positive and negative tweets, the process for finding emotive sentiment tweets employs a range of content-based and cooperative tactics . The emphasis of this research is sentiment analysis. Because this is a challenging process, several assumptions must be made before the algorithm is created. Details of sentiment analysis may be found at numerous levels. The following are the results of a comparison of our technique to two established approaches that do not address emotions: (i) Novel Naive Bayes and (ii) Long-term and short-term memory. Within collaborative social network filtering utilising the nature of the user, implicit sentiment analysis of positive and negative tweets is performed . Topic-based sentiment analysis may assist users in receiving broad public opinion regarding emotional items of interest, as well as positive and negative tweets emotions, a comprehensive user profile, and individualised recommendations(Vadivukarassi, Puviarasan, and Aruna 2017). Sentiment analysis reveals people's feelings on a variety of positive and negative tweet subjects, allowing for the creation of a more detailed user profile and individualised recommendations.Previously our team has a rich experience in working on various research projects across multiple disciplines(Balusamy et al. 2020; Arvind and Jain 2021; Zhao et al. 2020; Hani et al. 2020)

Long Short-Term Memory methods have several disadvantages, one of which is a lack of choice. Its application is a decent user recommendation classification technique, ignores user comments, and examines different positive and negative tweets recommendation tactics that take longer to execute than suggestions from other emotional tweet filters that contain the user's emotional sentiment tweets (Ghaly, Elabd, and Mostafa 2016). In the future, it will improve this model to make it more efficient, have a shorter run time, and deliver more accurate results. The goal of this study is to look at the sentiments that come with collaborative emotion filtering in social networks.

MATERIALS AND METHODS

The Saveetha School of Engineering Cyber Forensic Laboratory and the Saveetha Institute of Medical and

Technical Sciences coordinated this study (formerly known as Saveetha University). There are two groups in the planned work. Novel Naive Bayes is the first category, while Long Short-Term Memory is the second (Lanyi et al. 2021). With a sample size of 130, a confidence interval of 90%, G power of 60%, and a set maximum error bars of 0.05, the Novel Naive Bayes and Long Short-Term Memory algorithms were examined a different number of times.

Following dataset collection, preprocessing and data cleaning methods were employed to eliminate the datasets not-used and unimportant material of positive and negative tweets selection. After cleaning and prepping the data, access the data sets and apply the sentiment tweets with a tweet id using the opencv library and the efficiency of both the Novel Naive Bayes and Long Short-Term Memory algorithms(Kumar et al. 2022). The efficiency for Long Short-Term Memory is calculated. The Novel Naive Bayes and Long Short Term Memory algorithms clustering processes are shown below.

The following are the minimum hardware requirements to implement this model processor: Hardware configuration refers to the specifics and system resource settings allotted for specific devices; the following are the minimum hardware requirements to implement this model processor: Intel Core i3, 4GB RAM, 500GB HDD storage

Software specifications address the resources that must be installed in the target system in order for a programme to operate. The Windows operating system versions 7/8/10, the Python programming language version 3 or higher, the IDE PyCharm, and Jupyter are the minimum software requirements for this model to operate.

Novel Naive Bayes Algorithm

A supervised learning method is the Novel Naive Bayes algorithm. The Novel Naive Bayes classifier makes the assumption that the existence of one characteristic in a class is unrelated to the presence of another. Tweets contribute to sentiment tweets analysis of different data sets by users, which may be computed as the percentage of positive and negative tweets occurrences in tweets (de Groot 2012). It's mostly used in text classification with high-dimensional training datasets. Figure 1 depicts the Novel Naive Bayes method from dataset processing to output production.

Long Short-Term MemoryAlgorithm

A Long Short-Term Memory algorithm is a supervised learning is an artificial recurrent neural network (RNN) model deployed in deep learning LSTM networks are good for categorising and processing, and there may be unexpected gaps between significant occurrences in a time series, it can be difficult to make predictions using time series data (Kayıkçı et al. 2022). The collected data set shows the complete procedure of the proposed model. Figure 2 shows the algorithm for Long Short-Term Memory Algorithm from dataset processing to output generation.

The dataset has processed around one million sentiment tweets from various categories for testing and verification of categorization and sentiment analysis. The input dataset is processed using the OpenCV package, and each tweet is read separately. To increase the accuracy of the classifier and reduce noise, feature selection is employed. Simple phrases have less semantic value and are more difficult to understand than datasets.

By using complex semantic annotators positive and negative tweets, a tweet's hashtags, which are metadata tags used on twitter to indicate the context or flow of a tweet, are used to uniquely identify an idea. Novel Naive Bayes are compared using the Long Short-Term Memory method. Next I need to open a Python notebook and install the required modules to complete this work with Jupyter.

Statistical Analysis

IBM's SPSS Statistics is a collection of statistical tools for data management, advanced analytics, multivariate analytics, business intelligence, and criminal investigation. SPSS Statistics can scan and analyse data, other statistical programmes, spreadsheets, and databases, among other sources (McCormick and Salcedo 2017). For in-depth data analysis, SPSS is better. This tool is very useful for data analysis and visualisation. Tweets-ID is the independent variable, while emotional tweets by user's nature are the dependent variable. The T-Test analysis is carried out independently.

RESULTS

The dataset is shown in Table 1 for a variety of backgrounds and locations. The simulated efficiency study of innovative Naive Bayes and LSTM algorithms is shown in Table 2. Table 3 shows group statistical analysis for Naive Bayes and Long Short-Term Memory algorithms, with mean values of 87.10% and 81.60%, standard deviations of 4.383 and 4.115, respectively. Table 4 represents the independent T-test analysis of both the groups with significance value p= 0.904 ($p>0.05$) states that both groups are statistically insignificant.

Figure 4. depicts a bar graph study based on two algorithms efficiency. Novel Naive Bayes and Long Short-Term Memory have a mean efficiency of 87.10% and 81.60%, respectively. The results show that the Naive Bayes sentiment tweets recommendation system is better than the conventional method.

DISCUSSION

Novel Naive Bayes and Long Short-Term Memory algorithms are to estimate the efficacy of emotional tweets based on the user's nature. After evaluating the two models on the same dataset (Ahmed et al. 2014), the naive Bayes method beats the long-short term memory algorithms in positive and negative tweets. Naive Bayes emotional tweets and tweet recommendations are approximated by the suggested model (Ben Abdessalem Karaa et al. 2021). Emotional user recommendations on social media sites that include both good and negative tweets are not necessary. Various emotional datasets, Tweet IDs, and customised user tweets all contributed to a higher efficiency. From record processing to emotional tweet output, Figure 3 depicts the proposed architecture for sentiment tweets using the novel naive bayes algorithm. Research has shown that establishing more thorough user profiles is more successful than traditional content-based techniques. Because the difference between positive and negative tweets was not taken into account in the previous study, it was not possible to successfully implement the recommendation (Jia Wang et al. 2019). To calculate mood and recommend tweets based on the nature and feelings of users, the Novel Naive Bayes

algorithm was used. After all iterations on each dataset, the results revealed a continuous efficiency of 87.10 % (Banda et al. 2021).
When compared to the existing Long Short-Term Memory model, the new approach resulted in a 12% increase in efficiency (Banik et al. 2021). A similar study was undertaken with the goal of proposing sentiment tweets to future scholars who are interested in emotional tweets (Callahan et al. 2021). The recommendation algorithm additionally employs term synonyms to improve the efficiency of the analysis of tweets by. Existing tweet ideas to forecast sentimental tweeting techniques in the nature of users produced no such contradicting outcomes .
Long Short-Term Memory algorithms have several drawbacks, including a decision boundary between two classes that has a good ability to classify user recommendation approaches that take longer to execute than other sentiments, while ignoring user opinions while exploring different recommendation strategies. Here are some tips for using emotional tweets with tailored filtered emotional tweets from this Twitter user (de Groot 2017). This model might be improved in the future to get more efficient outcomes by boosting efficiency and reducing execution time.

CONCLUSION

When compared to the Long Short-Term Memory algorithm, the prediction of efficiency percentage for user nature sentiment tweets estimated by Novel Naive Bayes looks to have improved efficiency by 87.10% and 81.60%. By examining various recommendation algorithms, user recommendation systems that ignore user opinions have been presented. To make Twitter analysis more effective, our recommendation system shows tweets of various moods and emotions using contradictory emotional tweet recommendation keyword synonyms. The total number of favourable and negative tweets across all user observations is shown in the results.

DECLARATIONS

Conflict of Interest

The author states there are no conflicting interests.

Authors Contribution

The author, JVK was involved in data collection, data analysis, and manuscript writing. Author SVA contributed to the conceptualization, data validation, and critical review of the manuscript.

Acknowledgement

The author would like to thank the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the infrastructure needed to successfully complete this task.

Funding

We are thankful to the following organization for providing financial support in assisting us completing the research.

1. United Software Solutions.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

1. Ahmed, Soha GalalaldinKhider. 2014. *Sentiment Mining of Arabic Twitter Data*.
2. Banik, Sagar, Aniket Ghosh, Sumit Banik, and Anupam Mukherjee. 2021.

"Classification of COVID19 Tweets Based on Sentimental Analysis." *2021 International Conference on Computer Communication and Informatics (ICCCI)*. https://doi.org/10.1109/iccci50826.2021.9402540.

3. Ben Abdessalem Karaa, Wahiba, Eman Alkhammash, Thabet Slimani, and Myriam Hadjouni. 2021. "Intelligent Recommendations of Startup Projects in Smart Cities and Smart Health Using Social Media Mining." *Journal of Healthcare Engineering* 2021 (September): 3400943.
4. Friedrich, Natalie. 2015. *Applying Sentiment Analysis for Tweets Linking to Scientific Papers*. GRIN Verlag.
5. Ghaly, Rehab S., Emad Elabd, and Mostafa Abdelazim Mostafa. 2016. "Tweets Classification, Hashtags Suggestion and Tweets Linking in Social Semantic Web." *2016 SAI Computing Conference (SAI)*. https://doi.org/10.1109/sai.2016.7556121.
6. Groot, Roy de. 2012. *Data Mining for Tweet Sentiment Classification: Twitter Sentiment Analysis*. LAP Lambert Academic Publishing.
7. Jia Wang The University of Melbourne, Melbourne, Australia, Yungang Feng The University of Melbourne, Melbourne, Australia, Elham Naghizade The University of Melbourne, Melbourne, Australia, Lida Rashidi The University of Melbourne, Melbourne, Australia, Kwan Hui Lim The University of Melbourne, Melbourne, Australia, and Kate Lee The University of Melbourne, Melbourne, Australia. "Happiness Is a Choice." ACM Other Conferences. Accessed February 28, 2022. https://dl.acm.org/doi/abs/10.1145/3184558.3191583.
8. "Sentimental Analysis of Twitter Data with Respect to General Elections in India." 2020. *Procedia Computer Science* 173 (January): 325–34.
9. Sharma, Anjana Estelle, Ziva Mann, Roy Cherian, Jan Bing Del Rosario, Janine Yang, and Urmimala Sarkar. 2020. "Recommendations From the Twitter Hashtag #DoctorsAreDickheads: Qualitative Analysis." *Journal of Medical Internet Research* 22 (10): e17595.
10. Vadivukarassi, M., N. Puviarasan, and P. Aruna. 2017. "Sentimental Analysis of Tweets Using Naive Bayes Algorithm." *World Applied Sciences Journal* 35 (1): 54–59.

TABLES AND FIGURES

Table 1. Dataset name, Extension, Dataset Sources.

S.NO	DATASET NAME	DATASET EXTENSION	DATASET SOURCE
1	Sentiment Tweets	CSV	Kaggle
2	Tweets Recommendation	CSV	Kaggle

Table 2. Efficiency of Novel Naive Bayes and Long Short-Term Memory. The Naive Bayes algorithm is 6% more efficient than the Long Short-Term Memory algorithm.

ITERATION NO.	Naive Bayes (%)	LSTM (%)

1	94.2	86.7
2	92.5	87.3
3	86.1	84.1
4	87.8	79.4
5	82.3	76.7
6	89.5	83.2
7	86.9	81.9
8	81.5	75.4
9	83.1	85.1
10	91.2	80.4

Table 3. Group Statistics of novel naive bayes and long short-term memory algorithm with the mean value of 87.10% and 81.60%.

GROUP	**N**	**Mean(%)**	**Std.Deviation**	**Std.Error Mean**
Naive Bayes	10	87.10	4.383	1.386
LSTM	10	81.60	4.115	1.301

Table 4. Independent sample T-test is performed for the two groups for significance and standard error determination. The significance value p=0.904 (p>0.05) shows that two groups are statistically insignificant.

		Levene's Test for Equality of Variance		**T-test for Equality of Means**						
									95% Confidence Interval of the Difference	
	Equal Variance	**F**	**Sig**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std. Error Differ-ence**	**Lower**	**Upper**

Efficiency	**Assumed**	.015	.904	2.893	18	.005	5.500	1.901	1.508	9.494
	Not Assumed			2.893	17.929	.005	5.500	1.901	1.505	9.494

1. Start Program
2. Import the data from the required library
3. Give directory of the dataset in the csv extension file
4. Data as indexes with sentiment tweets validation
   i. Tweets ID of the data
   ii.Data analysis as emotions per the dataset
   iii.Plot the graph by using Matplot
5. Import GaussianNB
6. Library Classification
7. Now the use model selection for importing use train and test split
8. Use sklearn.naive_bayes for the importing and sklearn cluster
9. Give the sample size
10. Give the test size and train size then fit the train and test
11. Then print the efficiency score
12. End the program

Fig. 1. Pseudocode for Novel Naive Bayes algorithm.

1. Begin the programme
2. Import the required library's data.
3. In the csv extension file, specify the dataset's path.
4. Validation of sentiment tweets using data as an index
5. i. The data's Tweets ID
   ii.Analysis of data based on the dataset
   iii.Plot the graph with Matplotlib.
6. Library Classification, Import Kera Models
7. Now choose between the train and test split when importing models.
8. Importing and sequential models should both be done with LSTM.
9. Please specify the sample size.
10. Give the test and train sizes, then put the train together and test it.
11. The efficiency score will then be printed.
12. End the program.

Fig. 2. Pseudocode for LSTM (Long Short-Term Memory**)** algorithm.

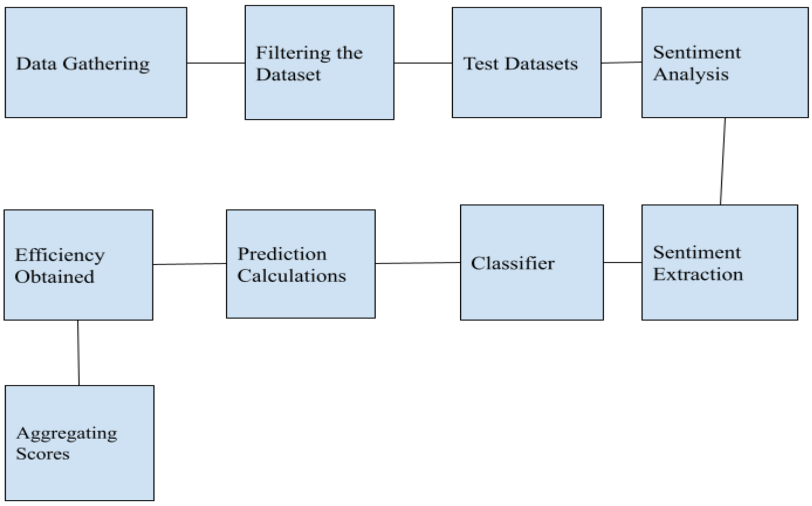

Fig. 3. Architecture for sentimental tweets recommendations using Novel Naive Bayes algorithm, from dataset processing to output of sentiment tweets.

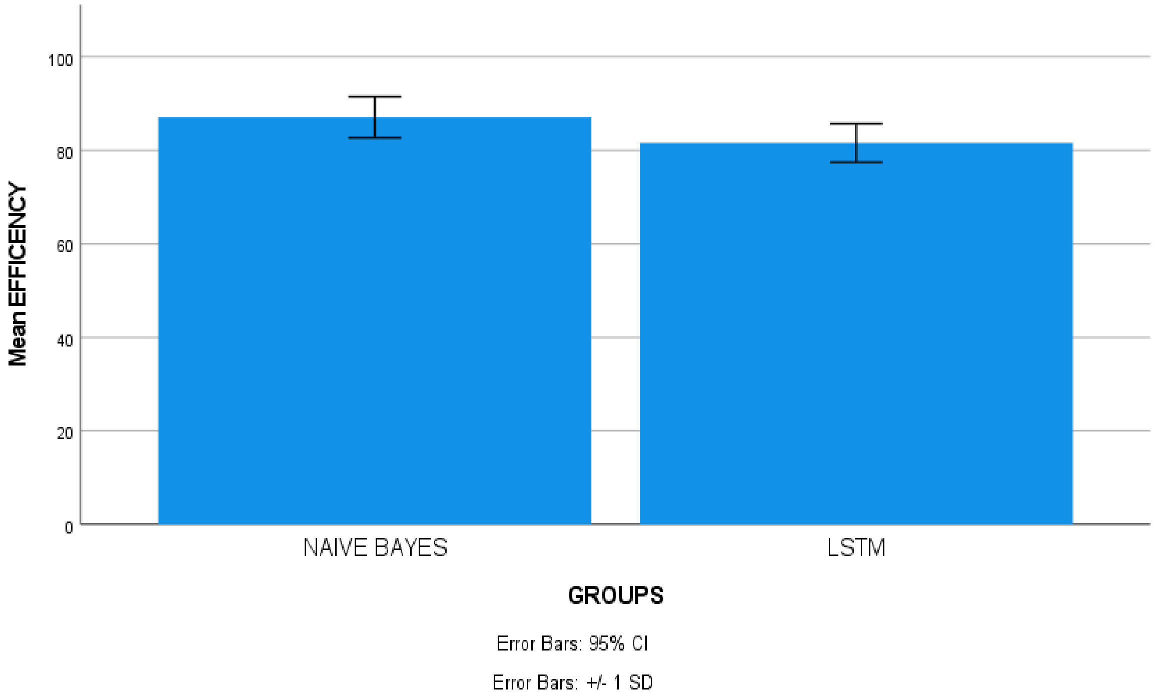

Fig. 4. Bar graph analysis of Novel Naive Bayes algorithm and Long Short-Term Memory algorithm. Graphical representation shows the mean efficiency of 87.10% and 81.60% for the proposed algorithm Novel Naive Bayes and Long Short-Term Memory respectively. X-axis : Novel Naive Bayes vs Long Short-Term Memory, Y-axis : Mean precision ± -1 SD.