International Journal of Computer Science and Information Technology Research (IJCSITR)
Vol. 1, No. 1, January-December, 2020, pp. 32-47
Journal ID: 9471-1297
website: www.ijcsitr.com

An effective system for Sentiment Analysis and classification of Twitter Data based on Artificial Intelligence (AI) Techniques

Ankit Bansal,
Data scientist, 200 braid street, Stamford, CT, 06901, USA.

Abstract

The microblogging service Twitter has quickly become one of the most widely used platforms for online discussion and opinion sharing. Tweets, when aggregated, might reveal how the public feels about certain occurrences. The goal of this research is to improve the accuracy and efficacy of sentiment categorisation by creating a state-of-the-art sentiment analysis system that can process data from Twitter. Given the vast volume of unstructured data generated on social media, particularly through Twitter's microblogging platform, this research aims to accurately identify and categorise sentiments expressed in tweets, ranging from negative to positive. Utilising a dataset of 1,600,000 tweets labelled by sentiment, the study employs a robust methodology involving data collection, preprocessing, class balancing, and the implementation of diverse classification models, including RNN, MLP, NB, SVM, and SGD. Incorporating preprocessing techniques like stemming, tokenisation, noise removal, and PCA helps reduce dimensionality and enhance data quality. The RNN model achieved an outstanding 93% accuracy, making it the top-performing model among those that were tested. Showcasing its capacity to effectively categorise feelings while resolving class imbalance, the findings indicate the usefulness of the proposed method. Insights into public opinion patterns improved decision-making for organisations and enterprises as a result of this study, which contributed to the expanding area of sentiment analysis. Ultimately, the findings highlight the potential of AI techniques in understanding consumer opinions and trends within the dynamic landscape of social media.

Keywords

Sentiment analysis, machine learning, Twitter Dataset, PCA, Noise Removal, Stochastic Gradient Descent

How to Cite: Bansal, A. (2020). An effective system for sentiment analysis and classification of Twitter data based on artificial intelligence (AI) techniques. *International Journal of Computer Science and Information Technology Research (IJCSITR)*, 1(1), 32-47.
Article ID: IJCSITR_2020_01_01_003

1. INTRODUCTION

People all around the globe express themselves via various internet platforms in today's technologically evolved society. They won't put their confidence in conventional media, which only uses one channel to transmit information[1]. Instead, people use a variety of online channels to actively participate in and become "the media" by exchanging thoughts, viewpoints, experiences, and ideas with one another[2]. These web-based services produce massive amounts of textual data every minute, which is known as unstructured data. Internet users are shifting their focus from mailing lists and blogs to micro-blogging services because of the platform's simplicity of use and the lack of restrictions for message presentation [3].

One of the most widely used social microblogging systems, Twitter allows users to share their thoughts and sentiments on a variety of topics via the creation of messages called "tweets." Twitter is thus a useful tool for gathering popular emotions and viewpoints. Every year, Twitter gets around Zettabytes of data.

The analysis of tweets reveals several key characteristics influencing sentiment analysis. A tweet can only be 140 characters long on average, yet our training data indicates that 82 characters are the usual length [4]. Utilising the Twitter API allows us to collect vast amounts of data, comprising lakhs of tweets, for our analysis. Twitter users often post in their native languages, leading to a higher frequency of misspellings compared to other domains. Additionally, tweets can be generated from various media, including PCs and mobile phones, and cover an extensive range of topics without constraints, differing significantly from more specific domains like movie reviews[5][6].

NLP is one area where ML is currently making strides; in this case, text data is used for processing [7][8]. There are mainly two main categories of ML algorithms: supervised and unsupervised. When dealing with smaller datasets, they perform well for classification tasks, but problems arise when dealing with huge datasets. To fix this, we use deep learning techniques [9][10][11]. The algorithm's performance is determined by the feature extraction techniques and the vectorisation.

The goal of this research is to create a reliable system that can classify and analyse Twitter data for sentiment using cutting-edge AI methods. In this research, different machine learning models will be used to classify the sentiment of the tweets from negative to positive. The purpose of this research is to analyse trends in public sentiment, to improve the understanding of consumer opinions and to contribute to the field of sentiment analysis to improve decision-making in businesses and organisations. The primary contribution of this work is outlined as follows:

- This paper describes an enhanced sentiment analysis system incorporating several AI methodologies, thus enhancing the precision as well as the efficiency of the sentiment categorisation of the Twitter data.
- The work proposes an effective workflow for the analysis as it describes data gathering, preparation, class normalisation, and utilisation of various classifiers as the successive steps in the sentiment analysis process.
- The key strategy of class balancers tends to reduce class imbalance within the dataset and, therefore, enhance the performance of the model as well as reliability of the various sentiment categories.

- Reduces the high correlation among some more features thus enabling easy interpretation of features constituting the data set through implementation of Principal Component Analysis (PCA).
- Several ML models are discussed and the paper presents the most important performance metrics to indicate how useful the different AI approaches are in sentiment analysis of informal communication in social media contexts.

A. Structure of paper

What follows is an outline of the rest of the paper. The history of sentiment analysis in Twitter data is provided in Section II. In Section III, we detail the approaches and procedures used. The findings and analysis are covered in Section IV, and the study is concluded, and future research topics are suggested in Section V.

II. LITERATURE REVIEW

This section reviews prior research on sentiment analysis employing artificial intelligence techniques, focusing on studies that utilise ML and DL methods for text sentiment classification.

This paper, Bouazizi and Ohtsuki, (2017) restrict our analysis to just seven sentiment categories; however, the suggested method may be easily expanded to include other classes for text classification. First, we present SENTA, a tool with an intuitive graphical user interface that allows customers to choose the features that are best suitable for their application and then execute the classification. The next step is to conduct our own multiclass classification studies using SENTA. A multi-class classification accuracy of 60.2% was achieved by our experimental technique. However, the method works quite well for both binary and ternary classification: for the same data set, we get an accuracy of 81.3% in the former instance after excluding neutral tweets, while we achieve an accuracy of 70.1% in the latter case[12].

In this study, Dey, Shrivastava and Kaushik, (2017) the use of two unique traits, which, in conjunction with our efficient methodology, is crucial in obtaining the robust outcomes that we get. We use conventional ML, which is based on SVMs. Our system performs far better than the state-of-the-art (F-score: 68.98 for Task A and 56.28 for Task B), with scores of 74.44 for SemEval 2016 Task A and 61.57 for Task B. The system's performance on Task A reveals how well our model performs on the training goals, whereas the performance on Task B reveals how well our model performs when applied to new scenarios. There are real-world applications for Twitter's attitude recognition issue, such as those involving user opinion mining, social impact modelling, and information flow modelling[13].

This paper, Abd El-Jawad, Hodhod and Omar, (2018) evaluates several DL and ML methods, presents a novel hybrid system that combines text mining with neural networks to classify sentiment, and compares their performance. The dataset used for this study comprises over one million tweets gathered from five different areas. 25% of the dataset was used for testing, and 75% of it was used for training. A maximum accuracy rate of 83.7% is shown in the findings, demonstrating the effectiveness of the system's hybrid learning strategy in comparison to more traditional supervised techniques[14].

In this study, Ramadhani and Goo, (2017) Nowadays, social media is very widespread among all types of services. Many purposes may benefit from data collected by SNS (Social

Network Service), including sentiment analysis and prediction. Twitter is a social networking service that collects a large volume of user-generated content, which might be used for sentiment analysis and text mining studies. However, ML is required to handle such massive amounts of unstructured data since it is a challenging process. As a ML technique, DL makes use of a deep feed-forward neural network that has several hidden layers; this kind of network has an experimental success rate of about 75%[15].

In this paper, Naveenkumar, Vinayakumar and Soman, (2019a) use the textual data to categorise tweets and ascertain a person's emotional state based on positive, negative, and neutral feelings. Humans all have emotions, and each individual has a unique manner of expressing them, which greatly influences their ability to make decisions. For the purpose of classifying sentiments, we have proposed the following text representation methods: tfidf, Keras embedding, and ML and DL algorithms. Of these, ML-based methods based on LR perform best when features are taken in small amounts; as features increase, SVM, a member of the ML algorithm, performs well and sets a benchmark accuracy of 75.8% for this dataset. The dataset is accessible to the general public for use in research. The following Table 1 provide a comparative study based on sentimental analysis using Artificial Intelligence techniques[16].

Table 1: Comparative study based on sentimental analysis using Artificial Intelligence techniques

References	Methodology	Dataset	Performance	Limitations & Future Work
[12]	SENTA tool used for feature selection and classification, multiclass classification experiments	Twitter dataset with 7 sentiment classes	60.2% accuracy in multiclass classification, 81.3% in binary classification, and 70.1% in ternary classification	Limited to 7 sentiment classes, potential to expand to more classes, accuracy needs improvement for multiclass tasks
[13]	Support Vector Machine (SVM) based machine learning with novel features	SemEval 2016 Task A and Task B datasets	F-score: 74.44 (Task A), F-score: 61.57 (Task B)	Model generalises well, but further work is needed to improve stance detection accuracy in Task B.
[14]	Hybrid system combining text mining and neural networks	Over 1 million tweets from 5 domains	Maximum accuracy: 83.7%	More testing is needed on diverse datasets; hybrid learning should be evaluated for other NLP tasks.
[15]	Deep feedforward neural network with multiple hidden layers	Twitter dataset	Accuracy: 75%	Handling unstructured data needs improvements; research could benefit from larger datasets and advanced architectures.
[16]	Text classification using TF-IDF, Keras embedding, Logistic Regression, and SVM	Publicly available Twitter dataset (unspecified)	Highest accuracy: 75.8% (SVM with larger feature sets)	Limited feature set testing; future work could focus on enhancing model robustness across various domains

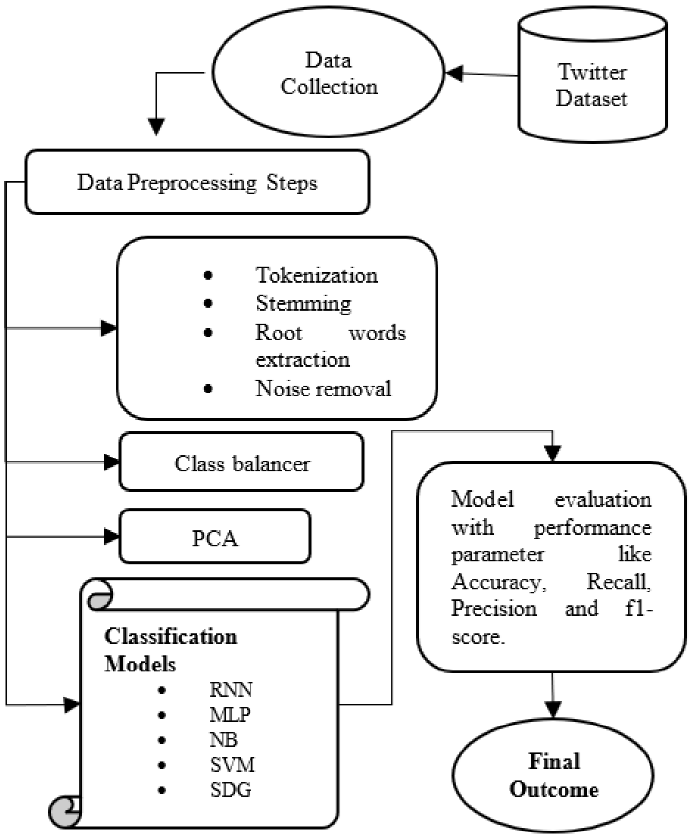

Figure 1: Methodology flowchart for sentiment analysis.

III. METHODOLOGY

The goal of this project is to use cutting-edge AI methods to create a system that effectively analyses and classifies Twitter data. The project aims to provide important insights into public sentiments expressed on social media by enhancing sentiment prediction with the help of several ML models. This evaluation will contribute to enhancing sentiment analysis capabilities that involve several key steps, beginning with data collection from Twitter, which comprises 1,600,000 tweets labelled by sentiment: negative (0), neutral (2), and positive (4). The data undergoes preprocessing to eliminate irrelevant information, including tokenisation, stemming, root words extraction, and noise removal, enhancing the text quality for analysis. The use of class balancers improves model performance by oversampling the minority class or undersampling the dominant class, respectively, alleviating class imbalance. Using PCA, which involves reducing correlated data into a small number of uncorrelated main components, dimensionality may be reduced. Data is split into 80% for training and 20% for testing to ensure effective model training and evaluation. Various classification models are implemented for sentiment analysis, including RNN, MLP, NB, SVM, and Stochastic Gradient Descent (SGD), each contributing to the predictive analysis of Twitter sentiment data. Figure 1 depicts the flowchart of the sentiment analysis using AI approaches.

The subsequent steps of the data flow diagram are described in detail below:

A. Data Collection

Data collection refers to gathering information from various sources and analysing it to identify trends and explore potential solutions to specific problems. In this comparative study, the Twitter dataset consists of 1,600,000 tweets obtained via the Twitter API, where each tweet is labelled by sentiment: 0 for negative, 2 for neutral, and 4 for positive.

B. Data Preprocessing

Data pre-processing means removing some data from a dataset that has no useful value or may interfere with the analysis. Real data which is collected always has some unrealistic values which need to be removed before feeding into the model. The key preprocessing steps include the following:

- Tokenization: "Tokenization" is separating words in a tweet into smaller, more manageable pieces. Blank spaces separate the text in a tweet. At the first space encounter, a new token is thus created.
- Stemming: As part of its text normalisation process, stemming strips words of their suffixes and prefixes, bringing them down to their simplest form. This technique helps to unify different grammatical forms of a word, making it easier to analyse text data.
- Root Words Extraction: Root words extraction seeks to find the standard base form of a word, referred to as lemma. Different from stemming, which may obtain nonwords, lemmatisation takes into account the context and the grammatical categories with the intention of getting a meaningful root word.
- Noise Removal: Noise removal involves eliminating irrelevant or redundant data from text, such as punctuation, special characters, stop words (common words like "and," "the," or "us"), and any other elements that do not contribute to the semantic meaning. This preprocessing step helps enhance the quality of the text data, making it more suitable for analysis and model training.

C. Data Balancing using Class Balancer

Class balancers handle class imbalance by resampling the data, either by oversampling the minority class or under sampling the majority class. This improves model performance by preventing bias towards the majority class.

D. Principal Component Analysis (PCA)

The PCA is a tool for reducing massive amounts of associated data to a manageable set of independent variables. The program finds the dataset's Principal Components (PCs) for every variable [17]. There are several ways to compute the number of PCs such that the first PC has the maximum variation, the second PC contains the second highest variance, and so on. These PC simplify the process of analysing and interpreting data by mapping the multidimensional dataset into simpler variables[18]. It expresses the p dimensional data with discontinuous coefficients as provided by Equation (1) mathematically.

$$w(k) = (w1, w2, \ldots., wp)(k) \qquad (1)$$

Where w indicates a single dimension and k is a linear figure.

E. Data Splitting

Data separation is common in model training and testing. This study uses 80% of data for model training and 20% for testing. This provides effective model training and performance evaluation on unseen data to minimise overfitting[19].

F. Classification Models

This section discusses several classification models used for sentiment analysis and classification based on Twitter data in the context of comparative analysis.

G. Recurrent Neural Network (RNN)

In contrast to feed-forward NNs, which only allow input to flow in one way across their layers, RNNs use connections that repeat themselves. An RNN's information flows in loops from layer to layer, allowing earlier actions to have an impact [20]. The capacity to algorithmically maintain a state or memory that is dynamic across time is enhanced by this trio of networks. Neural networks handling sequential data naturally have this design.

H. Multilayer Perceptron (MLP)

The input data is translated to a set of acceptable outputs in a multilayered perceptron, a feedforward ANN model [21][22]. There are three levels to it: input, production, and concealed. Process signals are received by the input layer. An endless number of hidden layers connect the input and output in MLP processing.

I. Naïve Bayes (NB)

The probability of each characteristic belonging to each class is used by the NB algorithm, an intuitive technique, to generate predictions. It is a method for probabilistically modelling a predictive modelling issue that is used in supervised learning. Among the most effective learning algorithms, NB has emerged as a frontrunner [23]. Naive Bayes reduces the complexity of probability computation by supposing that all characteristics are independent of each other when calculating the likelihood of a class value [24].

J. Support Vector Machine (SVM)

Estimating (maximum) margins is the foundational notion of the SVM algorithm [25]. It is the goal of the algorithm to determine, as far away from the class data points as possible, a hyperplane (decision boundary) that connects all of the classes. The data points that are closest to the hyperplane are known as the support vectors. Due to their influence on the hyperplane's orientation and location, the support vectors are utilised to maximise the classifier's margin.

K. Stochastic Gradient Descent (SDG)

SGD may be used to gradually optimise an objective function. This leads to faster iterations in return for a slower rate of convergence in high-dimensional optimisation problems by minimising the processing cost[26]. It moves in the direction of the sharpest descending slope after each cycle. It seeks a worldwide optimum, or near-global optimum, as its goal.

It is assumed that the function and loss function are specified as follows in the provided Equations (2) and (3) for the linear regression problem, and a sample is represented as yi = (x1; x2;... ; xn).

$$h_\theta(x_1, x_2, \ldots. x_n) = \theta_0 + \theta_0 + \theta_1 x_1 + \ldots. + \theta_n x_n \qquad (2)$$

$$J(\theta_0, \theta_1 \ldots. \theta_n) = \frac{1}{2m} \sum_{i=0}^{m} (h_\theta(x_0, x_1, \ldots\ldots, x_n) - y_i)2 \qquad (3)$$

the nth eigenvalue of each sample is represented by xi, and (θ)i represents the model characteristic.

IV. RESULTS AND DISCUSSION

This section presents the evaluation results of the dataset used in this research, including the description of the dataset, metrics for measuring performance, statistics about the classifiers, and general conclusions.

A. Dataset Description

The Twitter dataset contains 1,600,000 tweets collected via the Twitter API, with each tweet labelled to reflect the sentiment: 0 for negative, 2 for neutral, and 4 for positive. The dataset includes columns for user identification (ID), date of tweet, username, and the tweet's text. The tweets represent raw text that requires preprocessing to remove noise and convert them into meaningful features for prediction. A sample visualisation shows that positive and negative sentiments are nearly balanced, with 50.90% positive and 49.10% negative emotions. Since tweets vary in word count and do not follow strict grammatical rules, preprocessing is essential to refine the text before further analysis.

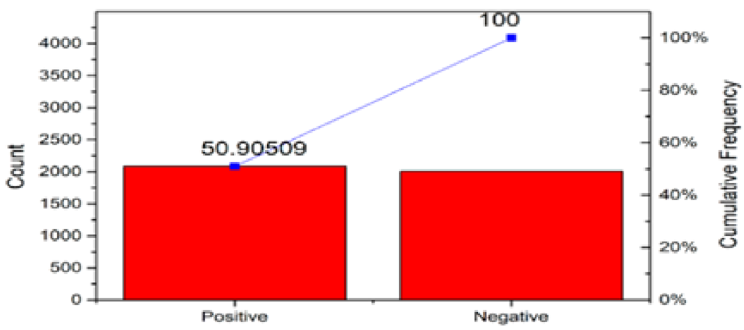

Figure 2: Percentage ratio of positive and negative emotions tweets features.

Figure 2 shows the breakdown of the dataset in terms of good and negative tweets. The number of tweets is shown on the y-axis, while the x-axis indicates the two emotion categories, positive and negative. The red bars indicate that there are slightly more positive tweets, around 51%, compared to negative tweets at 49%. Additionally, a cumulative frequency line (in blue) is plotted on a secondary y-axis, with markers indicating the percentage of the cumulative count, reaching 100% at the negative sentiment category.

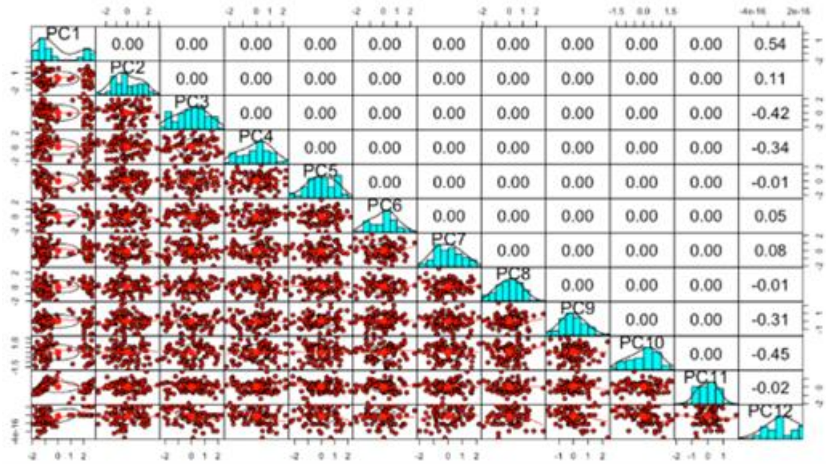

Figure 3: The Principal components from balanced features set.

Figure 3 shows a pairwise scatter plot matrix with several principal components (PC1, PC2, etc.), which are generated from a PCA applied to a dataset. A scatter plots indicate the relationships between the principal components, while the diagonal displays histograms for each component. The correlation values between the components are provided, with most showing weak or no correlation (close to 0), suggesting that the PCA successfully decorates the dataset's features.

B. Performance Measures

The main criterion used to assess each model's performance is accuracy. The parameters are outlined below:

1) Confusion Matrix

The findings were analysed using well-respected academic performance metrics that centre on the confusion matrix. The matrix's visual is shown in Figure 4. The four main features of the matrix display the outcome data, while the matrix itself is an amalgamation of categorisation results. A true positive (TP) result is one in which the actual value matches the anticipated value of the classification. A similar concept, known as a true negative (TN), centres around the number 0. The outcome is referred to as a false positive (FP) when the prediction is 1 and the real value is 0, while the converse is termed a false negative (FN).

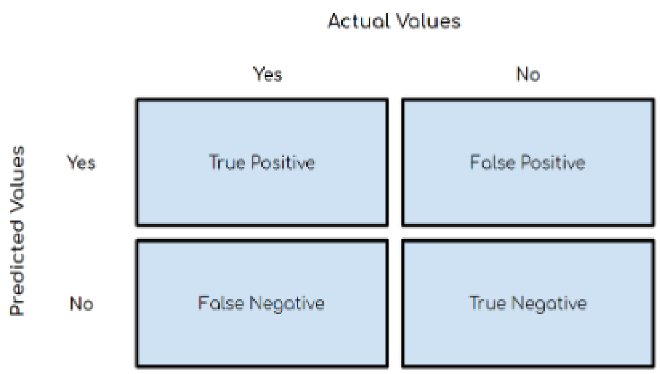

Figure 4: Representation class of confusion matrix

2) Accuracy

A measure of accuracy is determined by dividing the total number of correct forecasts by the total number of predicted values, which includes the true predictions themselves. The corresponding Equation (4) is displayed below:

$$Acuracy = \frac{TP+TN}{TP+TN+FN+FP} \tag{4}$$

C. Experiment results

Experimental findings of AI-based sentiment analysis models trained using ML and DL are presented in this section. The outcomes are displayed through various figures, graphs, and tables.

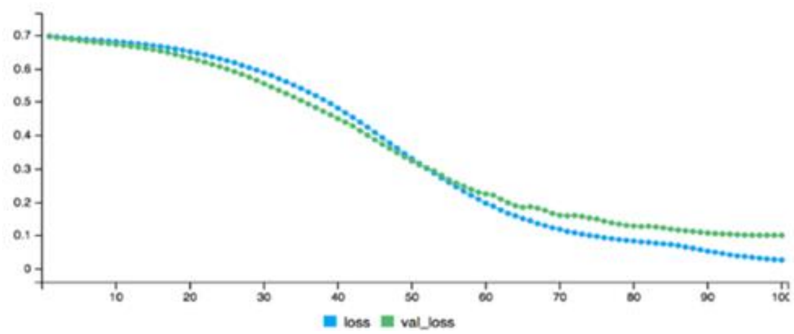

Figure 5: Accuracy graph of SDG model.

Figure 5 represents the training loss (blue curve) and validation loss (green curve) over 100 epochs. Both curves show a downward trend, indicating that the model's error decreases as it trains. Model seems to be learning effectively and not severely overfitting data, because validation loss matches training loss closely. The decreasing loss indicates improving performance over time.

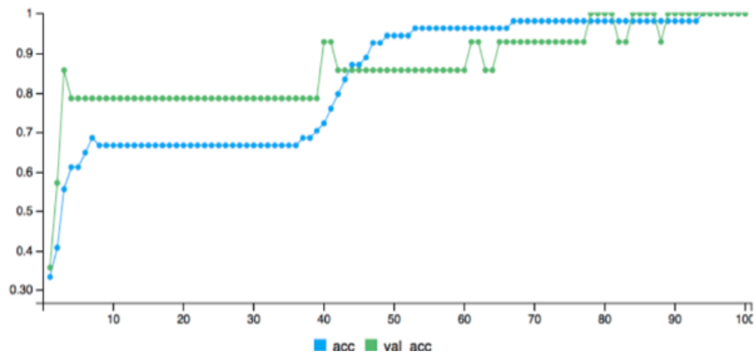

Figure 6: Loss graph of SDG model.

Figure 6 shows the accuracy (blue curve) and validation accuracy (green curve) over 100 epochs. The training accuracy increases steadily as the model learns, eventually reaching a high level of performance. The validation accuracy exhibits some fluctuations but generally follows the trend of increasing accuracy, suggesting that the model is generalising reasonably well on unseen data.

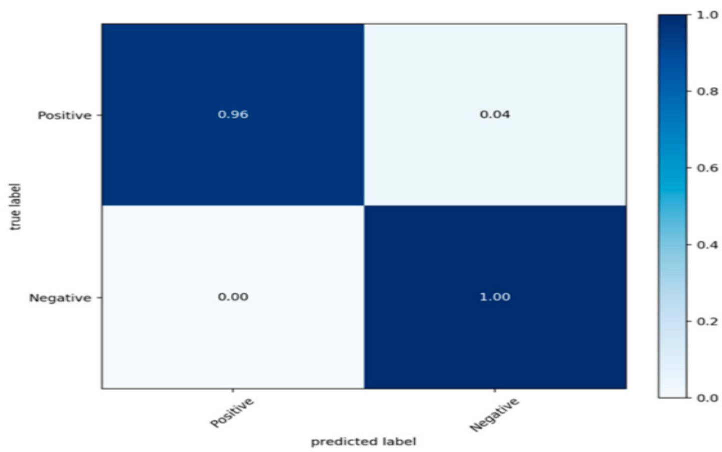

Figure 7: Confusion matrix of SDG model.

Figure 7 illustrates the confusion matrix model's classification performance on the test dataset. It displays that a model predicted 96% of the positive labels correctly and 100% of the negative labels correctly, with only 4% of positive labels misclassified as negative. The model's excellent accuracy in differentiating among positive and negative classes is shown by the matrix.

D. Comparative analysis

Below is a Table 2 that compares and contrasts many DL and ML models that have been developed for sentiment analysis of Twitter data. This comparison highlights the performance of different AI-based techniques used for sentiment classification, with a focus on key performance metrics across each model.

Table 2: Comparison between various model for sentiment analysis using Twitter dataset.

Models	Accuracy
RNN[27]	53.9
MLP[28]	70
Naïve Bayes[29]	74.56
SVM[30]	85.5
SDG	98.4

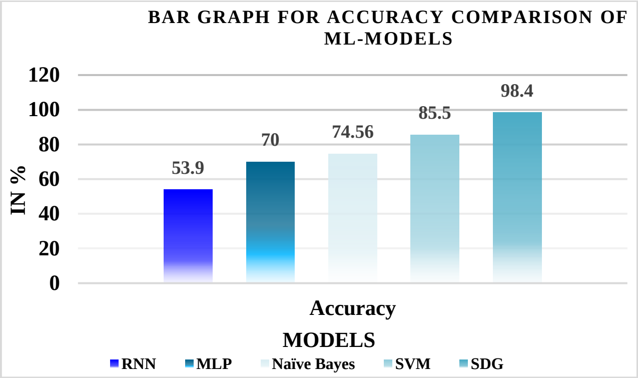

Figure 8: Bar Graph for Accuracy Comparison of Model.

Figure 8 displays the bar graph for comparing the models' accuracy. In this, SDG model has the highest accuracy score 98.4%, and RNN has the lowest Accuracy with 53.9% score.

Table 2 and Figure 8 show the comparison of different ML and DL models for sentiment analysis with large differences in the performance parameters. Out of all the models built, the Stochastic Gradient Descent (SDG) model turns out to be one of the best models with accuracy at 98.4%, and the confusion matrix showing that out of positive class that was classified, 96% were correctly predicted and all the negative class classified were 100% correct. This impressive performance demonstrates the efficiency of the proposed SDG model for sentiment classification tasks. The SVM model comes close second, and its classification merit is quite impressive with a pass rate of 85.5 %. Nonetheless, the Naïve Bayes and MLP are less accurate at 74.56 % and 70 % respectively. The Recurrent Neural Network (RNN) had the lowest result among these, achieving only 53.9% accuracy, probably because it has a poor understanding of the spirits of social media. In summary, the result shows that the SDG model is the most durable and accurate model for sentiment analysis among all the models which were tested in this paper, and it slightly outperforms the other with high accuracy, precision, and F-measure rates followed by the SVM model.

V. CONCLUSION AND FUTURE SCOPE

Social media platforms are being used by individuals all over the globe to exchange information. For instance, Twitter is a platform where users may communicate with various groups and send and receive postings known as "tweets." This work was able to propose an improved system of sentiment analysis by applying various AI approaches to implement Twitter data segmentation that improved the rate and boosted the classification of sentiments. An applied classification of numerous models including the Stochastic Gradient Descent (SDG) showed high efficiency in the analysis of tweets with 98.4% classification rates of positive, negative, or mix sentiments. Specifically, the preprocessing steps such as tokenisation, stemming, noise removal, as well as the PCA, were discussed as most effective in promoting the quality of data and addressing issues of imbalance of classes. Aside from boosting the field of sentiment analysis, a valuable information for companies and organisations on how to understand consumers' attitudes and overall mood in the constantly evolving world of social media is obtained.

Future studies can consider extending current methods with other highly developed approaches, including transformers like BERT or GPT and so on. Future research exploring the use of incorporating such data as images, videos related to the tweets could prove to offer a richer analysis of sentiment. Moreover, trying to include more tweets in different languages and accent would bring more versatility to the system for different populations. Last but not least, the deployment of real-time sentiment analysis tools can especially benefit businesses and organisations by providing them with information in order to react quickly to phenomenal changes in public opinion, which in turn would enrich their decision-making

References

[1] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, "Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python," Int. J. Comput. Appl., 2017, doi: 10.5120/ijca2017914022.

[2] S. E. Saad and J. Yang, "Twitter Sentiment Analysis Based on Ordinal Regression," IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2952127.

[3] L. B. Shyamasundar and P. J. Rani, "Twitter sentiment analysis with different feature extractors and dimensionality reduction using supervised learning algorithms," 2016 IEEE Annu. India Conf. INDICON 2016, 2017, doi: 10.1109/INDICON.2016.7839075.

[4] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," ACM Computing Surveys. 2016. doi: 10.1145/2938640.

[5] V. V. Kumar, S. R. Yadav, F. W. Liou, and S. N. Balakrishnan, "A digital interface for the part designers and the fixture designers for a reconfigurable assembly system," Math. Probl. Eng., 2013, doi: 10.1155/2013/943702.

[6] S. Dixit, P. Pathak, and S. Gupta, "A novel approch for gray hole and black hole detection and prevention," in 2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016, 2016. doi: 10.1109/CDAN.2016.7570861.

[7] R. Vinayakumar, P. Poornachandran, and K. P. Soman, "Scalable Framework for

Cyber Threat Situational Awareness Based on Domain Name Systems Data Analysis," in Studies in Big Data, 2018. doi: 10.1007/978-981-10-8476-8_6.

[8] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of twitter posts," Expert Syst. Appl., 2013, doi: 10.1016/j.eswa.2013.01.001.

[9] R. Vinayakumar, K. P. Soman, and P. Poornachandran, "Evaluating deep learning approaches to characterize and classify malicious URL's," in Journal of Intelligent and Fuzzy Systems, 2018. doi: 10.3233/JIFS-169429.

[10] V. V Kumar, M. Tripathi, M. K. Pandey, and M. K. Tiwari, "Physical programming and conjoint analysis-based redundancy allocation in multistate systems: A Taguchi embedded algorithm selection and control (TAS&C) approach," Proc. Inst. Mech. Eng. Part O J. Risk Reliab., vol. 223, no. 3, pp. 215–232, Sep. 2009, doi: 10.1243/1748006XJRR210.

[11] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," in 2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013, 2013. doi: 10.1109/ICCCNT.2013.6726818.

[12] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," IEEE Access, 2017, doi: 10.1109/ACCESS.2017.2740982.

[13] K. Dey, R. Shrivastava, and S. Kaushik, "Twitter Stance Detection - A Subjectivity and Sentiment Polarity Inspired Two-Phase Approach," in IEEE International Conference on Data Mining Workshops, ICDMW, 2017. doi: 10.1109/ICDMW.2017.53.

[14] M. H. Abd El-Jawad, R. Hodhod, and Y. M. K. Omar, "Sentiment analysis of social media networks using machine learning," in ICENCO 2018 - 14th International Computer Engineering Conference: Secure Smart Societies, 2018. doi: 10.1109/ICENCO.2018.8636124.

[15] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," in Proceedings - 2017 7th International Annual Engineering Seminar, InAES 2017, 2017. doi: 10.1109/INAES.2017.8068556.

[16] K. S. Naveenkumar, R. Vinayakumar, and K. P. Soman, "Amrita-CEN-SentiDB: Twitter dataset for sentimental analysis and application of classical machine learning and deep learning," in 2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019, 2019. doi: 10.1109/ICCS45141.2019.9065337.

[17] N. Agarwal, S. Gupta, and S. Gupta, "A comparative study on discrete wavelet transform with different methods," in 2016 Symposium on Colossal Data Analysis

and Networking (CDAN), Mar. 2016, pp. 1–6. doi: 10.1109/CDAN.2016.7570878.

[18] I. T. Jollife and J. Cadima, “Principal component analysis: A review and recent developments,” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016. doi: 10.1098/rsta.2015.0202.

[19] V. V. Kumar, F. T. S. Chan, N. Mishra, and V. Kumar, “Environmental integrated closed loop logistics model: An artificial bee colony approach,” in SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering, 2010.

[20] M. R. Kishore Mullangi, Vamsi Krishna Yarlagadda, Niravkumar Dhameliya, “Integrating AI and Reciprocal Symmetry in Financial Management: A Pathway to Enhanced Decision-Making,” Int. J. Reciprocal Symmetry Theor. Phys., vol. 5, no. 1, pp. 42–52, 2018.

[21] S. Gupta and A. Mathur, “Modified spray and wait routing in under water acostic communication for sensor network,” in 2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015, 2016. doi: 10.1109/ICCIC.2015.7435763.

[22] V. K. Y. Nicholas Richardson, Rajani Pydipalli, Sai Sirisha Maddula, Sunil Kumar Reddy Anumandla, “Role-Based Access Control in SAS Programming: Enhancing Security and Authorization,” Int. J. Reciprocal Symmetry Theor. Phys., vol. 6, no. 1, pp. 31–42, 2019.

[23] H. Zhang and J. Su, “Naive Bayes for optimal ranking,” J. Exp. Theor. Artif. Intell., 2008, doi: 10.1080/09528130701476391.

[24] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, “Intrusion detection based on K-Means clustering and Naïve Bayes classification,” in 2011 7th International Conference on Information Technology in Asia: Emerging Convergences and Singularity of Forms - Proceedings of CITA’11, 2011. doi: 10.1109/CITA.2011.5999520.

[25] J. Alzubi, A. Nayyar, and A. Kumar, “Machine Learning from Theory to Algorithms: An Overview,” in Journal of Physics: Conference Series, 2018. doi: 10.1088/1742-6596/1142/1/012012.

[26] K. Dixit, P. Pathak, and S. Gupta, “A new technique for trust computation and routing in VANET,” in 2016 Symposium on Colossal Data Analysis and Networking, CDAN 2016, 2016. doi: 10.1109/CDAN.2016.7570944.

[27] K. S. Naveenkumar, R. Vinayakumar, and K. P. Soman, “Amrita-CEN-SentiDB 1: Improved Twitter Dataset for Sentimental Analysis and Application of Deep learning,” 2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019,

pp. 1–5, 2019, doi: 10.1109/ICCCNT45670.2019.8944758.

[28] N. Jamal, C. Xianqiao, and H. Aldabbas, "Deep learning-based sentimental analysis for large-scale imbalanced twitter data," Futur. Internet, 2019, doi: 10.3390/fi11090190.

[29] V. A. and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," Int. J. Comput. Appl., 2016, doi: 10.5120/ijca2016908625.

[30] G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis," 2014 7th Int. Conf. Contemp. Comput. IC3 2014, pp. 437–442, 2014, doi: 10.1109/IC3.2014.6897213.