Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

Revolutionizing Sentiment Prediction: A Cutting-Edge Deep Learning Approach

B. Dilip Chakravarthy

Assistant Professor, Computer Science Department, Annamacharya Institute of Technology and Science, Rajampet, Andhra Pradesh

Article History:

Received: 15-10-2022

Revised: 27-10-2022

Accepted: 22-11-2022

Abstract

In light of its interactive and real-time nature, the exploration of public sentiment through the analysis of extensive social data has gained significant attention. Recent studies leverage sentiment analysis and social media to track major events by monitoring user behavior. This article presents an adaptable approach to sentiment analysis that promptly extracts user opinions from social media posts and assesses them. Over time, an escalating number of individuals have shared their opinions on social media, fostering increased communication among people. Despite these advantages, certain drawbacks have led to resentment in some individuals. One such drawback is the potential for hate speech, which can have a negative impact on the community when it involves insulting or threatening language. Recognizing and removing such speech from social media platforms before it spreads is crucial. The process of determining whether a text conveys feelings of hatred involves sentiment analysis. The Twitter dataset, comprising 5000 tweets, underwent analysis using the Python language. Deep learning techniques were employed to enhance the accuracy of the machine learning model. In both cases of the Twitter dataset, the Random Forest approach yielded an impressive 99 percent accuracy rate, showcasing promising experimental outcomes.

Keywords: interactive and real-time nature, threatening language.

1. Introduction

For instance, by using SA and taking into account factors like favourable or negative thoughts about the goods, suggestions of things provided by a recommendation system may be anticipated. Emotional analysis is the technique that uses Natural Language Processing to automatically mine attitudes, opinions, perspectives, and emotions from text, audio, tweets, and database sources (NLP). In a sentiment analysis, views in a text are categorised into "positive," "negative," and "neutral" categories. Subjectivity analysis, opinion mining, and assessment extraction are other names for it. Kajal et al [4] illustrated a cross cyber detection mechanism using monitored machine learning techniques with Project heuristic architecture of swarm intellect artificial bee hive with separate spectral transition and neural network with svm classifier as dual classification to create evaluation metrics to recognise internet backbone invaders. The accuracy rate of the DDoS attack-designed system is 98 percent, with an accuracy of 0.9[3]. Kushanket.al [5]. two gathering-based methodologies are used in this work. The first is reliant on logistic regression, k-nearest neighbour (KNN), and random forest, while the second is dependent on LSVM, and random forest. We have divided the material into three categories, such as Misogynist, Racist, and None, before using the procedures [6]. We discover that the second model had a better degree of precision than the other one we were using. The degree of accuracy that we obtained is 85%. Khalid et.al [7], presented a variety of algorithms, including gradient boosting, decision trees, RF, and logistic regression. The technique used to show if the provided material is negative, positive, or neutral is narrative text. Support Vector Machine (SVM), Naive Bayes (NB), and K-nearest neighbour are a few examples of classifier procedures that MeylanWongkar et al illustrated for separating hate speech. Nave-Bayes had the highest level of precision (80%) compared to K nearest neighbours (75%), while K nearest neighbours had a greater level of accuracy than Support Vector Machine (63 percent). In the current article, we look at popular perceptions of the Republic of Indonesia's official candidate for the years 2019–24. For next work, we must analyse the current President's survey using a variety of internet media[8]. How to cope with identifying keywords on a Twitter account is suggested by Hajime Watanabe et al[9] . We must make use of the instances and unigram recognition that were afterwards learned from training sets. Violent, aggressive, or animosity-inducing language is a component of hate speech. We must verify the data from 2010 tweets, which reveals the accuracy of 78% of insulting tweets and the exactness of 87% of analyst tweets that are hostile or neutral. To identify the presence of hatred speech across diverse age, religion, gender, and other categories, we must do a quantitative and subjective analysis [10], In their work, Dagar et al [11]. reported their attempt to assess several types of health-related tweets for depression and anxiety. We must identify the many types of emotions, such as negative. Kshirsagar et al [12] use the LSTM, GBDT and Word Embedding approach to identify hate speech on social media. The hate speech includes spreading false information, inciting violence, and other negative behaviours. The dataset was tested, and the findings showed an accuracy of 84% using a mixture of 90% receiving prepared information and 10% test data. Additional datasets must be tested exposed to the delayed effects of blended data in order to get amazing findings. Numerous deep knowledge replicas, including support course machines, XGBoost, and additional ML methods, were detailed by Zamani et al[13] in their study. They dealt with datasets that were mixed-code Hindi-English and English-Hindi. The classifier's results favour Hindi and English in large part, with F scores of 55, 68, and 54 [14]. Kokatnoor et al. [15]. A model called Stacked Weighted Ensemble is suggested for the identification of hate speech. Along with certain separate classifiers. Various tagged grouping approaches are usedfor the order of emotions, such as joy, disgust, anger, and fear. After gathering information, classifiers provide results in the range of 80 to 92 percent Liu et al[16]. Hate keywords Fear, Disgust, and Anger were reported by [17]. as having a greater rate. 9984 positive, 34177 negative, and 4658 neutral terms were obtained from Twitter [18]. The larger percentage of unfavourable tweets demonstrates the type of content that is prevalent in online media. Future efforts should not only take into account a single attempt[19]. 2. Method The Data set used for the analysis is shown in table below which have 200000 rows and 3 columns as shown in Tabel 1 with the target, text and tweet as the header. The Process Flow

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022**

chart is given in Fig 1 with the stages of import data, text cleaning, TF IDF vectorization, train test split, train and evaluate, adjust class imbalance, model evaluation, model building, regularization, hyperparameter tuning, get best parameter & evaluate and find recall and f1 score and top 10 terms used in the tweets is shown in Fig 2 Along with count of hate vs not hate speech used for sentiment analysis is shown in Fig 3

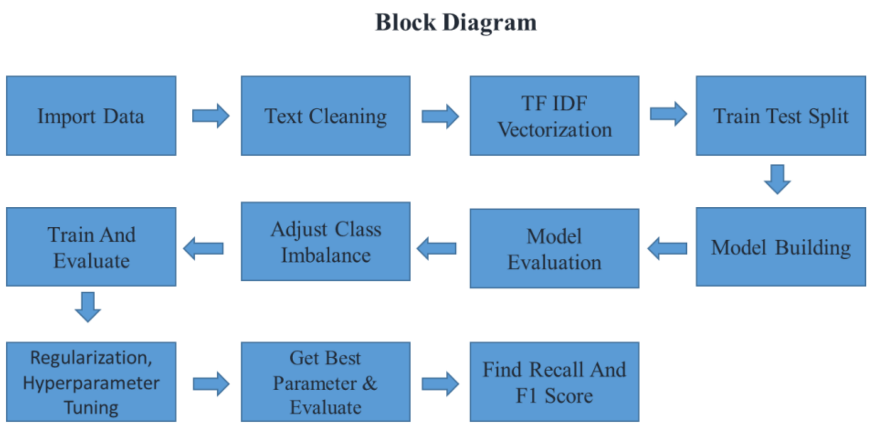

Fig 1. Flowchart of the process used

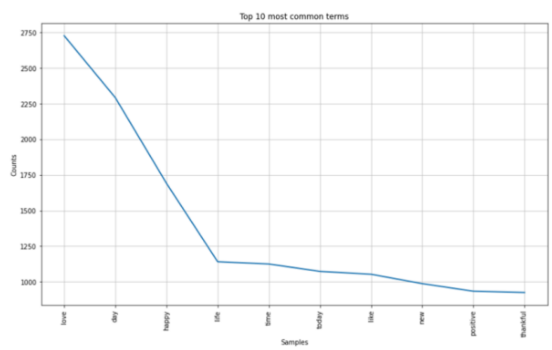

Fig 2. Top 10 Common terms used in the tweets

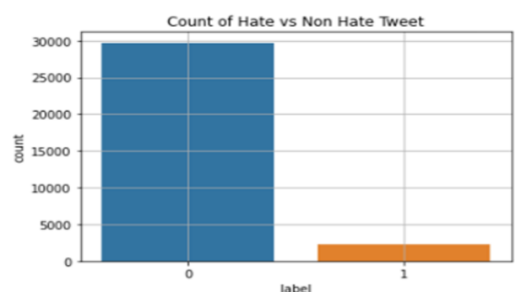

Fig 3. Count of Hate Vs Non Hate Tweets

3. Results and Discussion

3.1. AdaBoost Classifier

AdaBoost is a powerful technique for enhancing the performance of machine learning algorithms, particularly effective when training weak learners. Weak learners, in the context of classification problems, typically yield accuracy only slightly better than random chance. AdaBoost excels when paired with decision trees of minimal complexity, often with just one level.

As an ensemble approach in machine learning, AdaBoost, short for adaptive boosting, earns its name from the process of adaptively reassigning weights to each instance, assigning heavier weights to those instances that were misclassified. The algorithm is employed in supervised learning to mitigate bias and variance, operating on the principle that learners improve incrementally. In boosting, each subsequent learner, excluding the first, is built upon the knowledge of learners created before it. This iterative process transforms weak learners into strong ones, steadily improving the model's overall performance.

While AdaBoost shares similarities with boosting, there are nuanced differences in how it operates [20]. Figure 4 illustrates the stacking strategy used in machine learning, depicting the process of combining multiple learners to create a robust and accurate model. This stacking strategy is a visual representation of how AdaBoost leverages an ensemble of weak learners to collectively achieve superior predictive performance

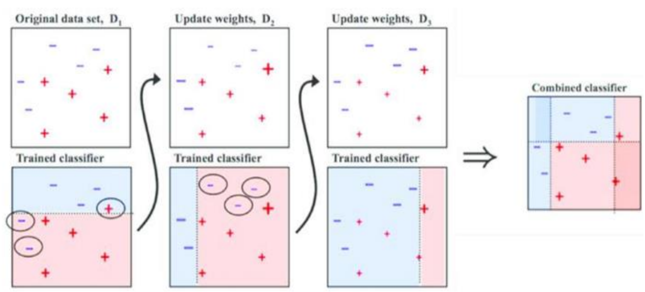

Fig 4. Adaboost approach used in Machine learning

3.2. Neural Net

A neural network operates by transforming an input vector through layers of units, commonly known as neurons, constituting the network. Each neuron processes input through a function and forwards the result to the subsequent layer. Neural networks are often characterized as feed-forward, indicating a unidirectional flow without feedback to the preceding layer, where each component transmits its output to all components on the layer above it. The connections between units involve weighted signals, and the adjustment of these weights during the training phase enables the neural network to adapt to specific problem domains [21].

Inspired by the learning process of the human brain, neural networks consist of a predefined set of parameters enabling the system to learn and adapt by analyzing new data. Each parameter, resembling neurons or inputs, receives one or more inputs, processes them through a function, and produces an output. These outputs then become inputs for the subsequent layer of neurons, forming a cascading process. After each neural layer is evaluated, and the inputs are received by the neurons, the results are propagated to the next layer of neurons [22].

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

Figure 5 illustrates a basic neural network, showcasing the interconnected layers of neurons and the flow of information. Additionally, Figure 6 presents different levels of sentiment analysis, demonstrating the application of neural networks in understanding and categorizing sentiments at various levels of complexity.

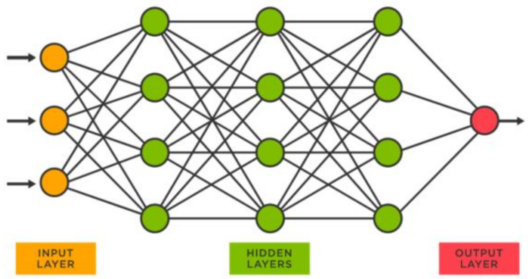

Fig 5. Simple Neural Net

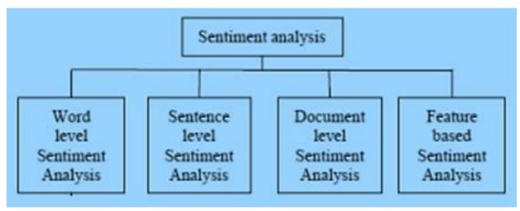

Fig 6. Different Levels of Sentiment analysis

3.3. The Decision Tree Classifier

Stands out for its versatility and effectiveness across various applications. Its inherent strength lies in the ability to extract insightful decision-making information from provided data [23]. The construction of a decision tree involves utilizing training sets, and this approach falls under the guided study algorithm family.

As illustrated in Figure 7, the decision tree method distinguishes itself by its applicability to a range of problems, including both classification and regression tasks. The model's objective is to generate an educational model based on training data, providing predictions about the class or value of the destination variable. This model serves as a valuable resource for individuals learning fundamental decision-making principles [24].

In the decision tree structure, the anticipation of a class label for a record begins at the hierarchy's root. By comparing the values of the current node with those of the root attribute, the algorithm follows a series of decisions and branches, ultimately reaching a prediction for the class label.

The flow chart of sentiment analysis tasks, depicted in Figure 8, showcases the decision tree's role in guiding the analytical process. This visual representation underscores the sequential nature of decision-making within the decision tree framework for sentiment analysis.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022**

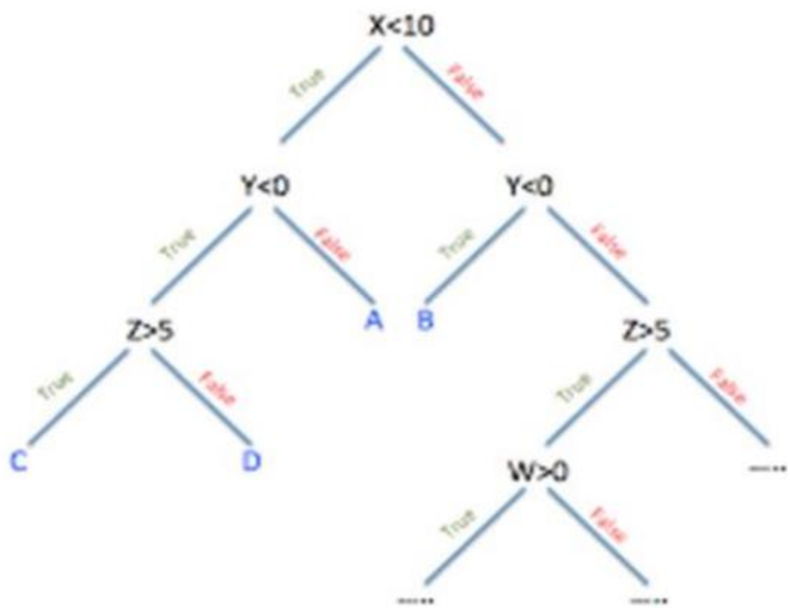

Fig 7. Working of DT

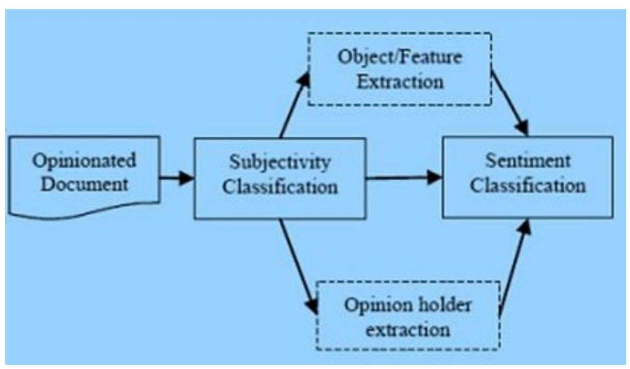

Fig 8. Flow chart of Sentiment analysis tasks

3.4. RBF SVM

In the realm of Support Vector Machines (SVM), the Radial Basis Function (RBF) kernel introduces a nonlinear element to the decision region. Even within the RBF kernel SVM, there coexists a linear decision area, offering versatility in handling datasets with varying characteristics.The RBF Kernel SVM is adept at creating nonlinear combinations of samples, projecting them into a higher-dimensional space where a linear boundary can be applied for class-splitting decisions. This proves especially valuable for datasets that exhibit nonlinear patterns or are inherently inseparable by a linear boundary.

When dealing with datasets that can be linearly separated, the option to employ the linear kernel function (kernel="linear") is available, tailored for scenarios with linearly structured data.

Understanding the nature of the dataset is key in determining the appropriate kernel function to utilize. For datasets with nonlinear characteristics, such as those that are linearly inseparable, kernel functions like RBF are recommended. Conversely, for datasets amenable to linear separation, the linear kernel function is a suitable choice.

The process of training an optimal model using the SVM approach is elucidated in Figure 4, providing a visual guide for practitioners who possess a clear understanding of when and how kernel functions should be applied. The existence of a linear decision area in the RBF kernel SVM decision region, as reiterated in Figure 9, reinforces the adaptability of the RBF kernel in handling diverse dataset structures.

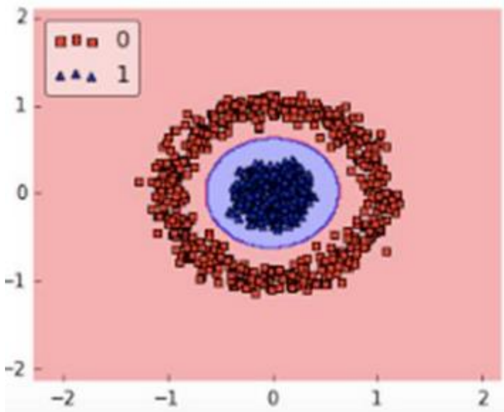

Fig 9. classification by RBF SVM

3.5. Nearest Neighbors

Classifier A machine learning technique called nearest neighbour classification seeks to categorise previously unobserved query items while differentiating between two or more destination classes. It is an example of supervised learning since, like any classifier, it needs some training data with predetermined labels [26]. K-Nearest Neighbors is a straightforward yet crucial machine learning classification method. It is a well-known supervised learning technique used in intrusion detection, data mining, and pattern identification. Since it is non-parametric, or assumes no underlying assumptions about data distribution, it is broadly applicable in real-world situations. Consider that there are two categories, A and B, and that we have a new data point, x1, and we are unsure which of the two categories this data point belongs to. To handle this kind of issue, a K-NN approach is necessary[27]. As seen in Fig 10, K-NN may be used to quickly identify the category or class of a certain dataset.

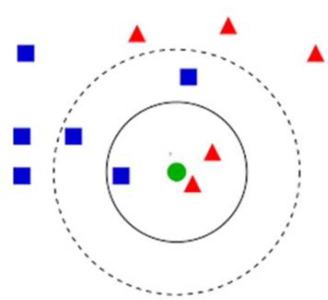

Fig 10. Nearest Neighbors Classifier

3.6. Random Forest Classifier

A well-known machine learning algorithm Random Forest is a technique used in the supervised learning process. It may be used to solve ML problems requiring both classification and regression. It is based on the concept of ensemble learning, which is a way of combining several classifiers to handle complex problems and improve model performance[28]. Rf is a classifier that, as the name indicates, utilises a number of decision trees on different sets of the input dataset and combines them to improve the dataset's prediction accuracy. Rather of relying on a single decision tree, the random forest uses predictions from each tree to estimate the plurality of predictions' amount of votes [29]. Even without adjusting the hyperparameters, random forest is a flexible and simple strategy that produces excellent results. Due to its ease and variety, it is also one of the greatest often rummagesale procedures. To provide a additional accurate and reliable forecast, random forests create and combine many decision trees [30]. Given that organization and reversion glitches make up the mainstream of modern machine learning schemes, the chance forests offer a significant advantage in many applications. Because classification is frequently seen as the machine's foundation, let's examine how random forests classify data. Fig 11 shows how two trees appear to be a random forest from below.

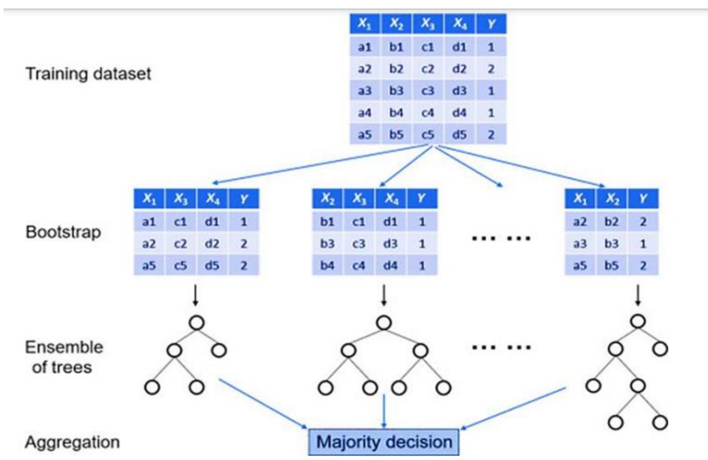

Fig 11. Random Forest Process

As trees grow, the random forest model becomes more unpredictable. Instead than focusing on the primary feature when dividing a node, it considers a random assortment of features. This results in a wide range, which frequently results in a better model [31]. 3.8. Ludwig Classifier Without creating a single line of code, models may be predicted and utilised using Ludwig, a powerful learning toolset. It is developed on top of TensorFlow and makes use of an abstraction based on data types to build a large number of applications. Ludwig's declarative file format enables extremely speedy prototype and model iteration. It allows experienced users to be much more productive by speeding up tasks that would otherwise take months or minutes. It is suitable for novices to train profound learning models without knowing all TensorFlow intricacies or profound learning in general [32]

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

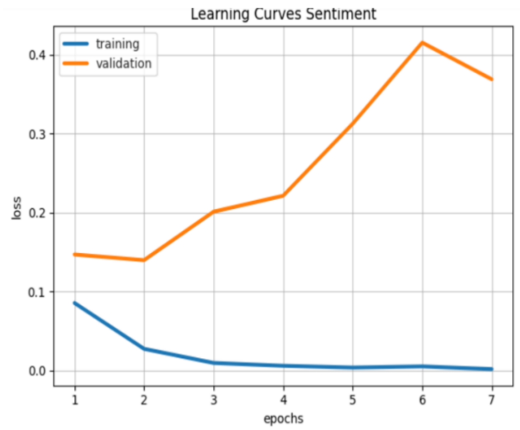

Fig 12. Learning Curve Sentiment

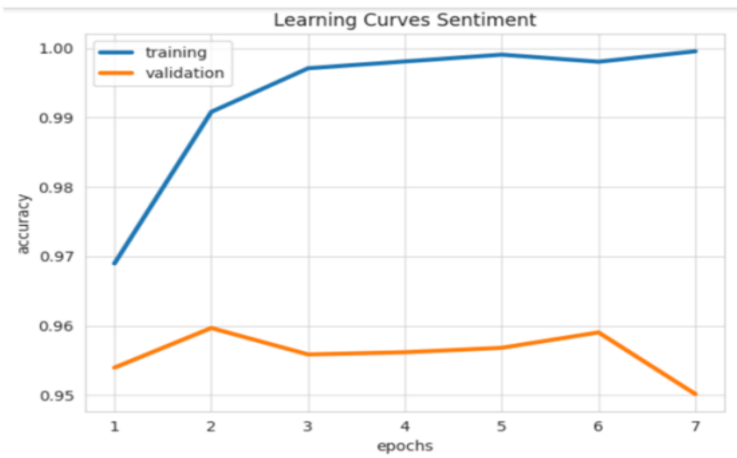

Fig 13. Learning Curve using Ludwig Classifier

A very excellent indication at 7 epochs, the learning curve obtained using deep learning guarantees about 99.9% accuracy. Fig 12 loss vs. epoch curve illustrates how a loss in the training curve decreases to virtually zero with a rise in epoch, which in and of itself demonstrates the system's accuracy. Fig 13 makes it quite evident that using the Ludwig classifier with many epochs boosts accuracy. Based on the pre-processed data sets, a number of Classification techniques were developed in the Python programming language on Google Colab. These methods included Boosting, Logistic Regression, SVM, and Naive Bayes. The top 8 algorithms' conclusions show positive outcomes for the situation. On the clean data set, our classifier was trained using the default functionality. Precision, F1-score, recall, accuracy, and ROC area were the four metrics used for assessment. A crucial confusion matrix is a 2-dimensional matrix that offers details about the actual classes and anticipated classes of a classifier. Model summary is Shown in Table 2.

Table 2. Model Summary

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 30)]              0
_________________________________________________________________
embedding (Embedding)        (None, 30, 300)           20047200
_________________________________________________________________
dropout (Dropout)            (None, 30, 300)           0
_________________________________________________________________
conv1d (Conv1D)              (None, 24, 128)           268928
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 18, 128)           114816
_________________________________________________________________
global_max_pooling1d (Global (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 512)               66048
_________________________________________________________________
dropout_1 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 513
=================================================================
```

The various results are combined and shown in Table 3

Table 3. Comparison among different classifiers

Algorithm used	Precision	Recall	f1-score	Accuracy
Linear SVM	1	0.93	0.96	0.93
Nearest Neighbors	0.99	0.94	0.96	0.93
Random Forest	0.97	0.96	0.97	0.96
Linear SVC	0.97	0.96	0.97	0.96
AdaBoost	0.97	0.94	0.95	0.94
Neural Net	1	0.93	0.96	0.92
Decision Tree	0.95	0.94	0.94	0.94
RBF SVM	0.95	0.96	0.96	0.95

Inference from table 3 1. Neural Net and Linear SVM are discovered to have the maximum value for accuracy, which is unity. Except for the Decision Tree, all other classifiers have a score over 0.95. 2. Recall indicates how accurately a classifier can predict future events. It reaches its maximum in RBF SVM, Random Forest, and Linear SVC, and it is more than 0.93 in all other cases, indicating how accurate the prediction is. 3. The F1 Score, which is the harmonic average of accuracy and recall, has a maximum in both Random Forest and Linear SVC and virtually similar value. 4. One of the key metrics for any machine learning application is accuracy, which is the percentage of correctly categorised predictions. In this study, the maximum accuracy for Random Forest and Linear SVC was determined to be around 97 percent. Additionally, it is discovered that employing a deep learning model with Ludwig Classifier, as illustrated in Fig 14, an accuracy of roughly around 99.9 percent is attained.

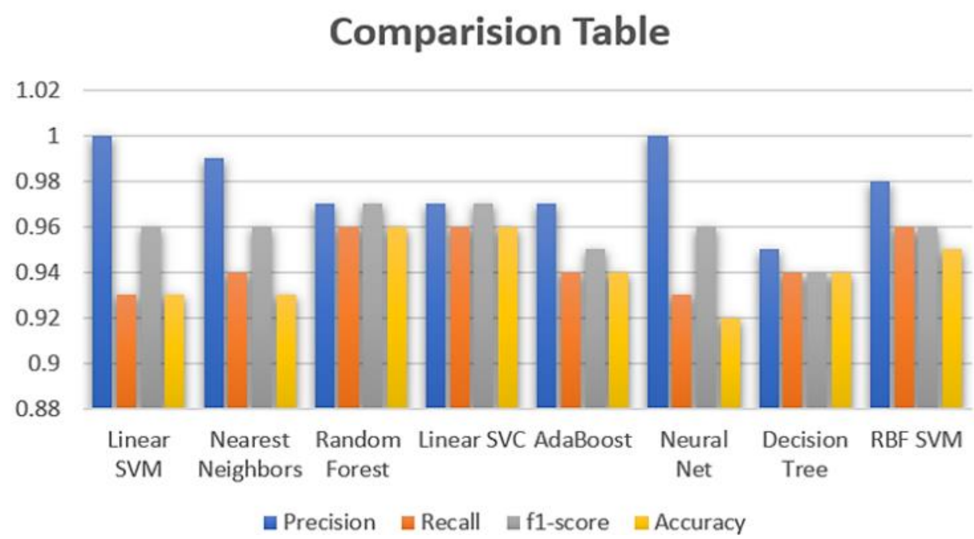

Fig 14. Summary of different classifiers

4. Conclusion

In this study, we have effectively used deep learning and a variety of categorization techniques to predict human behavior. Performance metrics including accuracy, recall, precision, and f1-score were obtained during the study, which was done in Python. The results show that Random Forest and Linear SVC have maximum accuracy of about 97 percent. A deep learning network using Ludwig Classifier also achieves an accuracy of nearly 99.9%, as shown in Fig 9. Researchers in the same field will find this study article useful.

5. References

1. V. Sundararaj and M. Rejeesh, "A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites", Journal of Retailing and Consumer Services, vol. 58, p. 102190, 2021. https://doi.org/10.1016/j.jretconser.2020.102190
2. Sultan, M.T.; Sharmin, F.; Badulescu, A.; Gavrilut, D.; Xue, K. Social Media-Based Content towards Image Formation: A New Approach to the Selection of Sustainable Destinations. Sustainability 2021, 13, 4241. https://doi.org/10.3390/su13084241
3. Salminen J, Mustak M, Corporan J, Jung S, Jansen BJ. Detecting Pain Points from User-Generated Social Media Posts Using Machine Learning. Journal of Interactive Marketing. 2022;57(3):517-539. https://doi:10.1177/10949968221095556
4. Zhang, T., Schoene, A.M., Ji, S. et al. Natural language processing applied to mental illness detection: a narrative review. npj Digit. Med. 5, 46 (2022). https://doi.org/10.1038/s41746-022-00589-7
5. Q. Pham et al., "Swarm intelligence for next-generation networks: Recent advances and applications", Journal of Network and Computer Applications, vol. 191, p. 103141, 2021. https://doi.org/10.1016/j.jnca.2021.103141

6. Hashemi, Mohammad J. Detecting Anomalies in Network Systems by Leveraging Neural Networks. University of Colorado at Boulder ProQuest Dissertations Publishing, 2021. 28652078.
7. Gaur, V., Kumar, R. Analysis of Machine Learning Classifiers for Early Detection of DDoS Attacks on IoT Devices. Arab J SciEng 47, 1353–1374 (2022). https://doi.org/10.1007/s13369-021-05947-3
8. M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), 2021, pp. 41-44, https://doi.org/10.1109/EIConCIT50028.2021.9431845
9. J. Sachdeva, K. K. Chaudhary, H. Madaan and P. Meel, "Text Based Hate-Speech Analysis," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 661-668, https://doi.org/10.1109/ICAIS50930.2021.9396013
10. Q. Pham et al., "Swarm intelligence for next-generation networks: Recent advances and applications", Journal of Network and Computer Applications, vol. 191, p. 103141, 2021. https://doi.org/10.1016/j.jnca.2021.103141
11. M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter", 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019.
12. Rajput, G., Punn, N.S., Sonbhadra, S.K., Agarwal, S. Hate Speech Detection Using Static BERT Embeddings. In: Srirama, S.N., Lin, J.CW.,Bhatnagar, R., Agarwal, S., Reddy, P.K. (eds) Big Data Analytics. BDA 2021. Lecture Notes in Computer Science(), vol 13147. 2021. Springer, Cham. https://doi.org/10.1007/978-3-030-93620-4_6
13. G. Liu, S. Wang, B. Wan, Z. Wang and C. Wang, "ML-Stealer: Stealing Prediction Functionality of Machine Learning Models with Mere Black-Box Access," 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 2021, pp. 532-539, https://doi.org/10.1109/TrustCom53373.2021.00083
14. M. Dagar, A. Kajal and P. Bhatia, "Twitter Sentiment Analysis using Supervised Machine Learning Techniques," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1-7, https://doi.org/10.1109/ISCON52037.2021.9702333
15. Aljero MKA, Dimililer N. A Novel Stacked Ensemble for Hate Speech Recognition. Applied Sciences. 2021; 11(24):11684. https://doi.org/10.3390/app112411684
16. Ankita, S. Rani, A. Bashir, A. Alhudhaif, D. Koundal and E. Gunduz, "An efficient CNN-LSTM model for sentiment detection in #BlackLivesMatter", Expert Systems with Applications, vol. 193, p. 116256, 2022. https://doi.org/10.1016/j.eswa.2021.116256
17. Madukwe, K.J., Gao, X., Xue, B. What Emotion Is Hate? Incorporating Emotion Information into the Hate Speech Detection Task. In: Pham, D.N., Theeramunkong, T., Governatori, G., Liu, F. (eds) PRICAI 2021: Trends in Artificial Intelligence. PRICAI 2021. Lecture Notes in

Computer Science(), vol 13032. 2021. Springer, Cham. https://doi.org/10.1007/978-3-030-89363-7_21

18. Ankita, S. Rani, A. Bashir, A. Alhudhaif, D. Koundal and E. Gunduz, "An efficient CNN-LSTM model for sentiment detection in #BlackLivesMatter", Expert Systems with Applications, vol. 193, p. 116256, 2022. https://doi.org/10.1016/j.eswa.2021.116256
19. I. Anderson and W. Wood, "Habits and the electronic herd: The psychology behind social media’s successes and failures", Consumer Psychology Review, vol. 4, no. 1, pp. 83-99, 2020. https://doi.org/10.1002/arcp.1063
20. H. Sihotang, M. Zarlis, S. Efendi, D. Jollyta and Husain, "Evaluation of Maturity Level of Information and Communication Technology (ICT) Governance with CobIT 5.0 Case Study: STMIK Pelita Nusantara Medan", Journal of Physics: Conference Series, vol. 1255, no. 1, p. 012046, 2019. https://doi.org/10.1088/1742-6596/1255/1/012046
21. V. Rybalkin, N. Wehn, M. R. Yousefi and D. Stricker, "Hardware architecture of Bidirectional Long Short-Term Memory Neural Network for Optical Character Recognition," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017, pp. 1390-1395, https://doi.org/10.23919/DATE.2017.7927210
22. C. Aggarwal, "Neural Networks and Deep Learning", 2018. https://doi.org/10.1007/978-3-319-94463-0
23. Polat, Kemal, and SalihGüneş. "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform." Applied Mathematics and Computation 187, no. 2 (2007): 1017-1026. https://doi.org/10.1016/j.amc.2006.09.022
24. Mueen A, Zafar B, Manzoor U. Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. International journal of modern education & computer science. 2016 Nov 1;8(11). https://doi.org/10.5815/ijmecs.2016.11.0
25. Fouad IA. A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms. Ain Shams Engineering Journal. 2022 Aug 1:101895. https://doi.org/10.1016/j.asej.2022.101895
26. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," in IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153- 1176, Secondquarter 2016, https://doi.org/10.1109/COMST.2015.2494502
27. Schulte, Matthias, and Katharina Morik. "Real-time feature extraction from video stream data for stream segmentation and tagging." an der TechnischenUniversit¨at Dortmund Fakult¨at f¨urInformatik, Dortmund, January 22, 2013.
28. Osisanwo FY, Akinsola JE, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J. Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT). 2017 Jun;48(3):128-38.
29. Dev VA, Eden MR. Formation lithology classification using scalable gradient boosted decision trees. Computers & Chemical Engineering. 2019 Sep 2;128:392-404.
30. Dev VA, Eden MR. Formation lithology classification using scalable gradient boosted decision trees. Computers & Chemical Engineering. 2019 Sep 2;128:392-404.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 11, 2022**

31. Abidi SM, Xu Y, Ni J, Wang X, Zhang W. Popularity prediction of movies: from statistical modeling to machine learning techniques. Multimedia Tools and Applications. 2020 Dec;79(47):35583-617. https://doi.org/10.1007/s11042-019-08546-5
32. Jupalle H, Kouser S, Bhatia AB, Alam N, Nadikattu RR, Whig P. Automation of human behaviors and its prediction using machine learning. Microsystem Technologies. 2022 Aug;28(8):1879-87. https://doi.org/10.1007/s00542-022-05326-4