Improving Drug Review Categorization using Sentiment Analysis and Machine Learning

1st Rezaul Haque
Department of CSE
East West University
Dhaka 1212, Bangladesh
rezaulh603@gmail.com

2nd Piyush Kumar Pareek
Artificial Intelligence and Machine Learning
Nitte Meenakshi Institute of Technology
Bengaluru, India
piyush.kumar@nmit.ac.in

3rd Md Babul Islam
School of ICT
University of Calabria
Calabria, Italy
mdbabul.islam@dimes.unical.it

4th Ferdaus Ibne Aziz
Department of Management
University of Ulsan
Ulsan, Republic of Korea - 44610
shifatshaheen92@gmail.com

5th Swakshar Das Amarth
Department of CSE
BRAC University
Dhaka 1212, Bangladesh
swakshar.das.amarth@g.bracu.ac.bd

6th Katura Gania Khushbu
Department of CSE
East West University
Dhaka 1212, Bangladesh
katurakhushbu@gmail.com

***Abstract*—Content created by users from social media platforms has become increasingly important as the internet has grown, and it now contains a plethora of knowledge regarding medications, diagnostics, treatments, and illnesses. Customers' opinions regarding previously used medicines are contained in the data in the form of comments, which can be used to detect crucial adverse drug reactions. Using SA techniques like SA, this information can be used to get important insights. It is usually impossible for a prospective buyer to read through all of the comments prior to making a purchasing decision. Drug evaluations benefit both healthcare providers and the general population since they provide vital healthcare data. Assessing opinions regarding many aspects of drug evaluations can provide substantial insights, facilitate choice-making, and improve public monitoring systems by revealing collective perspective. Another key issue is the unstructured and linguistic nature of the evaluations, which finds it challenging for users to categorize comments into useful insights. Previous research has used ML algorithms to perform categorization on drug reviews. So, using appropriate NLP and ML algorithms, our key objective in this study is to acquire a higher categorization score than earlier research studies. We achieved our aim by applying SA on medicine reviews to detect positive, negative, and neutral user comments training five ML algorithms on TF-IDF and CV feature extraction technique. Our results show that out of all ML algorithms, RF trained on CV surpass previous study results, with accuracy and F1 score of 96.65% and 96.42% respectively.**

***Index Terms*—drug sentiment analysis, machine learning, text classification**

I. Introduction

With the emergence of Web 2.0, the internet has become a more dynamic and user-driven platform, resulting in an explosion of content created by users on social networking and online retail websites [1]. The tremendous proliferation of online data has spurred numerous industries, including healthcare and pharmaceuticals, to utilize Sentiment Analysis (SA) to harvest insights from user comments. Manufacturers may improve their pharmacovigilance systems, identify problems more quickly, compare brand reputation, carry out aftermarket drug surveillance, and guarantee the secure administration of medications with no side effects by automatically analyzing massive amounts of user-generated content. The early detection of adverse drug reactions is vital for risk management and to avoid financial and reputational losses. While SA alone cannot directly identify adverse drug reactions, it provides valuable insights into user sentiment towards drug experiences. The categorization of user comments into positive, negative, or neutral categories helps identify potential instances of negative sentiment, which may indicate adverse drug reactions or side effects. By focusing on negative sentiment comments, we can identify instances where users express concerns, complaints, or experiences of undesirable effects related to the drugs.

SA is a form of analysis that examines emotions in written content to ascertain the degree of agreement or disagreement among viewpoints on various aspects of a subject [2]. It's not widely used in the medical domain due to privacy and ethical concerns [3]. In healthcare, SA needs to be precise to understand patient satisfaction on factors like staff cordiality and treatment effectiveness. In addition, accurately extracting sentiments from user reviews is a significant challenge due to the immense growth of reviews. SA on drug reviews can potentially save lives by identifying unknown adverse drug effects and alerting healthcare professionals to potential risks associated with specific medications. Research on SA of drug reviews [4], [5] focuses on improving patient satisfaction and feedback. The performance of Machine Learning (ML) classifiers in SA is greatly influenced by text preprocessing and feature extraction techniques. Therefore, it is crucial to carefully choose appropriate methods for text cleansing and feature extraction specifically for drug SA. However, traditional ML algorithms and feature extractors like TF-IDF and BoW have encountered low accuracy scores due to class imbalance, high dimensionality, and lack of preprocessing.

The major objective of this study is to use ML algorithms

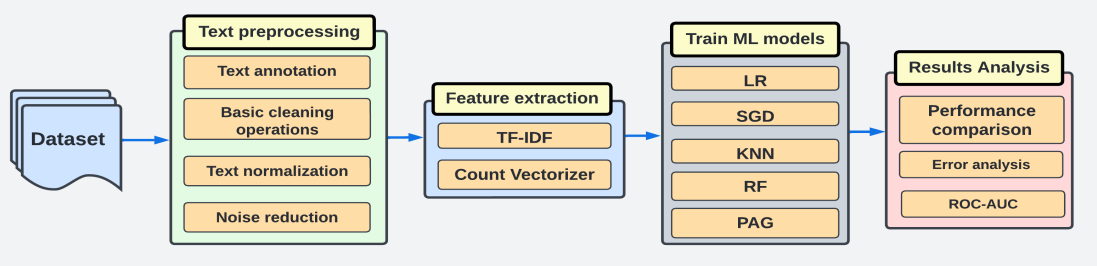

Fig. 1. Proposed methodology.

that have been trained on the TF-IDF and CV to increase the accuracy of earlier studies. The suggested methodology for the current investigation is shown in Figure 1. The 215,063 medicine reviews we gathered for our study were divided into three categories based on user ratings: positive, negative, and neutral. To improve the quality of the text, text pre-processing techniques using Python's NLTK package were applied. Five different ML algorithms, including Logistic Regression (LR) [12], Random Forest (RF) [13], K-Nearest Neighbors (KNN) [11], Passive Aggressive (PAG), and Stochastic Gradient Descent (SGD) [16], were trained on both TF-IDF and CV. The accuracy, error analysis, and model performance of these classifiers and feature extractors were then compared, offering important insights into how effective they were for drug SA. Our approach complements existing methods of drug surveillance and can significantly contribute to patient safety by enabling proactive measures to mitigate risks associated with medications. Our contributions toward the accomplishment of the objective are as follows:

- Offered a thorough comparative study of ML algorithms, enhancing the knowledge and understanding of drug SA, empowering researchers and practitioners in their future endeavors.
- Improved the accuracy of earlier studies by choosing suitable feature extraction methods and applying ML algorithms to use.

II. Related Works

ML algorithms are being used to analyze user evaluations, particularly in the restaurant and E-commerce industries. Despite its widespread use in various applications, SA has received relatively little attention in the pharmaceutical industry. However, recent developments have shown the relevance of mining medication reviews for various healthcare stakeholders.

Numerous research investigations have utilized various methods to analyze drug sentiment analysis (SA) through diverse preprocessing and feature extraction techniques. However, the lack of comprehensive research with high accuracy in drug SA can be attributed to inadequate preprocessing, feature extraction techniques, and unsuitable training parameters. Due to limited access to authentic datasets, researchers often rely on the drugs.com dataset for training machine learning/deep learning algorithms in binary or multiclass classification tasks. In a study by [6], an ML-based drug recommendation system was introduced, employing several ML algorithms for binary sentiment classification. The LR model achieved the highest accuracy score of 91% among the models trained using four different feature extraction methods. Similarly, in [5], a fuzzy-rough feature-selection-based ML model was proposed, using BOW and TF-IDF techniques to classify sentiments into three categories. The RF model with TF-IDF achieved the highest accuracy score of 67%. Another approach discussed in [7] focused on a linguistic method for drug sentiment analysis, utilizing a multiclass dataset collected from WebMD. The proposed approach outperformed two SVM models, achieving an accuracy score of 69%, surpassing the baseline by 7%. Lastly, [8] investigated the impact of sentiment analysis features in detecting adversarial drug reactions from online posts, achieving an 80% accuracy score through binary classification using a dataset compiled from Twitter and DailyStrength.

It is evident from existing literature that earlier ML models do not exhibit a satisfactory level of accuracy on ML algorithms. Therefore, it is of utmost importance to devise appropriate techniques to enhance the performance of models in order to meet the demands of the field. We are confident that by meticulously choosing preprocessing methods, feature extraction techniques, and ML models, we can enhance the accuracy of previous findings. To the best of our knowledge, there is currently no research that attains high accuracy scores in multi-class SA of drug reviews obtained from drugs.com using ML algorithms.

III. Methodology

A. Data Collection

"The Drug Review Dataset", acquired from the UCI repository, is described in this section. It comprises user reviews for different drugs, information on diseases that specific treatments are thought to treat, and a scale of ratings from zero to ten stars that represent the degree of satisfaction experienced by users. The dataset comprises a collection of 215,063 assessments, each categorized as negative (ratings between 0 and 4), neutral (falling within the range of 4 and 7), or positive (exceeding 7). The dataset also contains the name of the medicine, the user's condition, the usable count, and the count of votes. In Figure 2, the distribution of reviews per class is illustrated, with 105,433 positive, 100,071 negative, and 9,559 neutral reviews. To train ML models, the class

labels were converted to numerical format using Sklearn's Label Encoder package.

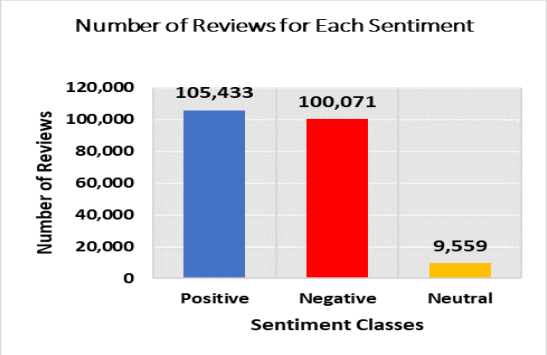

Fig. 2. Number of reviews for each class.

B. Text Pre-processing

Since data gathered from the internet frequently contains many types of noise, such as adverts, HTML components, punctuation marks, and empty spaces, text pre-processing plays an important role in SA. Pre-processing removes erratic, raucous, and incorrect information, which reduces text noise and improves the efficacy and precision of classification models [9]. We used specific data cleansing methods during this research, such deleting stop words and numerals. We also conducted text normalization, which involved expanding English contractions, lemmatization, and spell correction to reduce the dimensionality of the data. These techniques were implemented using Python's NLTK and RegEx libraries. Noise reduction and text normalizing should be approached cautiously since they have a risk of accidentally eliminate number of rows, which can lead to reduced accuracy.

C. Feature Extraction

To use ML algorithms for classification, textual data needs to be transformed into numerical input [10]. This is done through feature extraction techniques such as TF-IDF and CV.

1) TF-IDF Vectorizer: To determine the significance of a phrase inside a text, the TF-IDF approach is utilized. The inverse frequency is calculated as the logarithm of the entire count of documents divided by the volume of documents including the phrase. It also calculates the term frequency, which is the frequency of a term in a document split by the entire number of words. TF-IDF makes it easier to find important words inside a document by transforming textual data into a vector space. The formulae for TF-IDF are shown in formulas 1-3, where the word is represented by a and the document by b.

$$TF(a,b) = \frac{\text{Term } a \text{ frequency in document } b}{\text{Total words in document}} \quad (1)$$

$$IDF(a) = \log_2\left(\frac{\text{Total documents}}{\text{documents with term } a}\right) \quad (2)$$

$$TFIDF(a) = TF(a,b) \times IDF(a) \quad (3)$$

Supervised ML algorithms will utilize N-Grams to extract text features, where N represents the number of tokens taken sequentially from the text. In this particular research, Sklearn feature extraction library was employed, utilizing TF-IDF on sanitized reviews. The maximum feature count was set to 12,000, with an N-gram range encompassing unigrams and bigrams.

2) CV: Utilizing the CV approach, the majority of common terms are identified, and a limited representation of the words across this set of words is created. With the help of the CV, the documents are converted into an array that includes the counts of each token. The documents are first divided into segments, and a dense grid is then built using the token counts found in the texts. Stop phrases found in the document collection are eliminated before removing terms which appear in less than four different texts to create the matrix. By using this strategy, seldom used phrases are removed. Utilizing a variety of N-grams and the default maximum feature value, the CV was used in this particular study to process sanitized reviews.

D. Training ML Models

The dataset is first subjected to feature extraction and then divided into 80% for training and 20% for testing. The model is trained on the labeled 172,044 drug reviews in the training set and evaluated on the unlabeled 43,011 drug reviews in the testing set. The selected ML models are described below.

1) RF: The RF Classifier is a classification technique that uses many decision trees. Randomness is used to construct each tree to create statistically independent forests. To apply the RF, two parameters need to be set: the number of trees as estimators (n estimators) and the criteria to be used. Default settings can yield good outcomes, but an excessive number of trees may result in inconsistent outcomes. Setting the n estimator factor to 100 and choosing the entropy criterion parameter in this specific study inquiry allowed for the identification of the best RF model for categorization over a series of tests.

2) PAG: PAG strategies are frequently used in software programs that manage large data collections. Such algorithms are widely used in a variety of settings involving substantial learning. In contrast to batch learning, which uses the full training data at once, PAG receives data input consecutively as needed and updates the ML model automatically. The algorithm was developed in this study using default settings, with a cap of 200 cycles allowed.

3) LR: In the basic form of LR, a binary dependent variable is modeled using a logistic function. However, there are many more complex variations of this model. This method, which is widely used for data classification, falls under the Generalized Linear Models category. LR is a statistical tool that can be employed to forecast the probabilities of experimental outcomes. It is also referred to as the Ensemble Learning algorithm. For training the LR model, we used a cost parameter of 1, a maximum iteration of 200, a tolerance of 0.001, and the liblinear solver settings.

4) SGD: When dealing with convex loss functions, SGD, a well-known ML approach, is frequently used to improve linear

models. It is an improved kind of gradient descent which employs interpolated gradients created by picking random subsets of the training data. In datasets with duplicate observations, this strategy is effective and beneficial. The maximum number of iterations in this study's model, which has an LR loss function, is set at 200.

5) KNN: KNN algorithm is a type of supervised learning method used for classification and regression.In this study, value of K was set to 3 to train the model. When K=3, the algorithm classifies data by selecting the 3 nearest data points and assigning the most common class among them as the predicted class for the input data point. This approach is based on the idea that similar data points tend to be in the same class.

E. Evaluation

The performance of the suggested models was assessed using metrics such as accuracy, precision, recall, and F1 score. Equations 4-7 were employed to compute these measurements. TP denotes the count of accurately identified positive data, while TN represents the count of accurately identified negative data. FN signifies the count of erroneously identified negative data, and FP indicates the count of incorrectly identified positive data.

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

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{f1-score} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \tag{7}$$

IV. Results Analysis

A. Performance Based on Evaluation Metrics

Table I presents the performance of experimental models trained on two different feature extractors, TF-IDF and CV. Among the models, the RF model consistently outperforms the others, achieving the highest accuracy scores for both feature extractors. For the TF-IDF feature extractor, the RF classifier achieves the highest accuracy of 93.32%. It also demonstrates high precision (93.39%), recall (94.85%), and F1 score (94.16%). The LR classifier follows with an accuracy of 89.62% and competitive classification scores. The KNN classifier achieves an accuracy of 89.16% with similar classification scores. The SGD and PAG classifiers show slightly lower accuracy but still perform reasonably well. In contrast, when using the CV feature extractor, the classifiers consistently achieve higher accuracy scores. The RF classifier stands out as the best performer with an accuracy of 96.65% and excellent precision, recall, and F1 score values. The LR, KNN, PAG, and SGD classifiers also achieve high accuracy scores above 95%. This indicates that the RF model effectively identifies positive instances while minimizing false positives and false negatives.

In terms of feature extractor comparison, Figure 3 reveals that the CV consistently outperforms TF-IDF. When TF-IDF is employed as the feature extractor, the models achieve accuracy scores ranging from 86.01% to 93.32%. On the other hand, when using CV, the models achieve higher accuracy scores ranging from 95.31% to 96.65%. Notably, the RF model obtains the highest accuracy of 96.65% when coupled with CV.

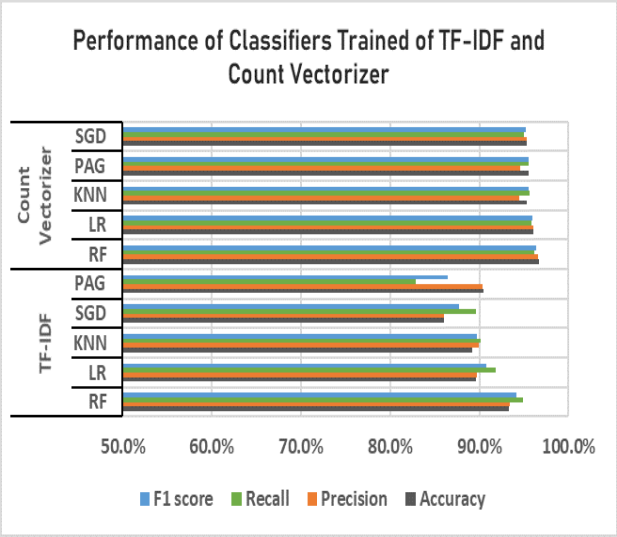

Fig. 3. Results of ML algorithms trained on TF-IDF vectorizer.

B. Error Analysis

The confusion matrix was utilized to assess how well the top-performing model (RF trained on CV) performed in our classification test. Figure 4 shows the results, and upon analyzing the matrix, we can observe that the RF model accurately predicts positive and negative classes of drug reviews more effectively than the neutral class. Specifically, out of the 20,972 positive reviews, the model correctly predicts 20,425 as positive, and out of the 20,117 negative reviews, it correctly predicts 19,457 as negative. Furthermore, the neutral class achieves a respectable true positive rate, with 1,428 classes correctly classified out of 1,922 reviews. However, the model encounters challenges in accurately predicting the neutral class. Despite this, the RF model trained with CV outperforms RF model trained on TF-IDF in terms of accuracy and fewer incorrect predictions for all three classes. When tested on 20% of the data, the model makes incorrect predictions of 3.2% for the negative class, 25.7% for the neutral class, and 2.6% for the positive class.

C. ROC-AUC Curve

The ROC Curve, used for classification, balances sensitivity and specificity. TPR and FPR for different thresholds are needed to generate the curve, with FPR and TPR values shown on the x and y axes. The diagonal of the ROC space is the starting point for analysis, with increased precision as the curve moves towards it. The range of values for AUC-ROC

TABLE I
RESULTS OF EACH CLASSIFIER'S TF-IDF AND CV TRAINING

Feature Extractor	Model	Accuracy	Precision	Recall	F1 score
TF-IDF	RF	93.32%	93.39%	94.85%	94.16%
	LR	89.62%	89.68%	91.89%	90.76%
	KNN	89.16%	89.99%	90.11%	89.77%
	SGD	86.01%	86.02%	89.58%	87.76%
	PAG	90.43%	90.40%	82.90%	86.50%
CV	RF	96.65%	96.60%	96.13%	96.42%
	LR	96.06%	96.04%	95.84%	95.97%
	KNN	95.33%	94.46%	95.61%	95.52%
	PAG	95.57%	94.56%	95.51%	95.48%
	SGD	95.31%	95.36%	94.95%	95.17%

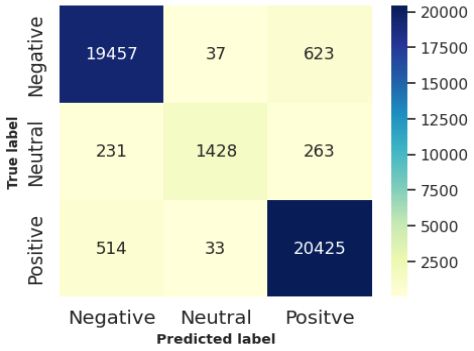

Fig. 4. Confusion matrix of RF trained on CV.

curve area is from 0 to 1, where 0 signifies complete misclassification and 1 represents perfect classification accuracy. Figure 5 shows the mean AUC score for RF models trained with CV, with label 0, 1, and 2 obtaining AUC scores of 99% 92% and 99% respectively. The higher AUC indicates better discrimination between different classes, suggesting that our approach is very effective in classifying drug reviews.

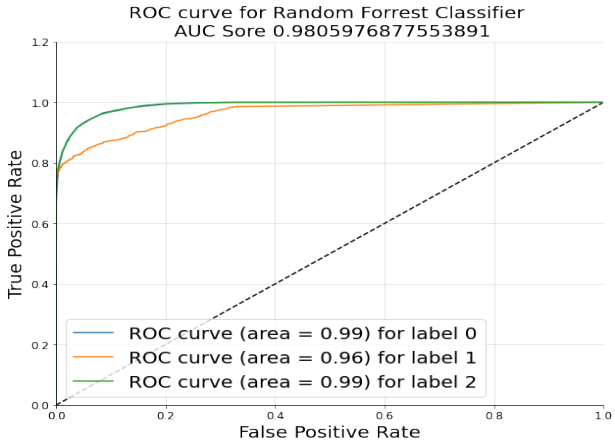

Fig. 5. ROC-AUC curve of highest performing classifier (RF).

D. State-of-the-Art Comparison

In the comparative analysis presented in Table II, we can see that our research demonstrates superior performance compared to the cited previous works, achieving higher accuracy on a larger dataset and providing improved classification capabilities as evidenced by the ROC-AUC analysis. In contrast, the cited research works achieve lower accuracy scores ranging from 66.41% to 82.86% on smaller datasets and a different number of classes. The higher accuracy and the larger dataset used in our research indicate that our approach performs better in SA of drug reviews, enabling more accurate identification of sentiment and potentially uncovering unknown adverse drug effects.

TABLE II
COMPARATIVE ANALYSIS WITH PRIOR RESEARCH WORKS

Paper	Number of class	Total reviews	Accuracy
[5]	3	4142	66.41%
[6]	2	212,141	93%
[7]	3	4200	82%
[8]	4	6279	82.86%
Our approach	**3**	**215,063**	**96.65%**

V. DISCUSSION

In this research, the dataset employed exhibited a smaller quantity of neutral reviews, posing challenges for ML models to accurately predict this class in comparison to the other two classes. While the obtained results surpassed previous research findings, there is still room for improvement by augmenting the number of neutral reviews. In this study, the RF classifier emerged as the best performer among the different classifiers. RF's success can be attributed to several key factors. Firstly, RF utilizes ensemble learning, combining multiple decision trees to make predictions. This ensemble approach allows RF to capture complex relationships and interactions between features in drug reviews, enabling it to accurately identify sentiment patterns. Additionally, RF provides a measure of feature importance, allowing it to identify the most influential features in determining sentiment. This capability helps RF leverage informative keywords or phrases, enhancing its predictive accuracy. Furthermore, RF is robust to noise and overfitting, making it well-suited to handle subjective and diverse drug reviews. Its ability to capture nonlinear relationships between features and sentiments is particularly valuable in SA tasks.

Additionally, the better performance of the CV feature extractor compared to TF-IDF in the task of drug SA can be

attributed to several factors. In drug SA, the exact frequency of words plays a crucial role in determining the sentiment of a review. CV captures the exact occurrence of words in each document, allowing it to effectively represent the frequency of positive, negative, and neutral words. This direct word counting approach is particularly relevant in SA, where the frequency of sentiment-bearing words is crucial for accurate classification. On the other hand, TF-IDF considers the relative importance of words across the entire corpus. While TF-IDF is effective in capturing the importance of words in a general context, it may not be as effective in capturing the specific sentiment-bearing words that are essential for drug SA. TF-IDF scores can become sparse when there is limited word overlap between documents, potentially leading to reduced accuracy in sentiment classification.

VI. Conclusion

In today's fast-growing technological world, examining reviews using AI technology has become crucial. Whether we are going shopping, making online purchases, or choosing a restaurant, reviews are often the first thing we look at to ensure that we are making the right decision. Consequently, the significance of reviews is growing in our everyday routines. Among these, drug reviews hold particular importance as they provide insights into user preferences, thereby supporting healthcare professionals in their decision-making process and facilitating improved well-being. Therefore, this study aims to explore and compare different classifiers and feature extraction techniques to develop a reliable and efficient model for drug SA. By addressing this problem, healthcare professionals, pharmaceutical companies, and policymakers can gain valuable insights into public sentiment towards drugs, enabling them to make informed decisions and enhance patient care. The text pre-processing was performed using Python's NLTK package to enhance the text quality, and five different ML algorithms were trained separately on TF-IDF and CV and compared based on opted evaluation metrics. The experimental results demonstrated the effectiveness of the RF classifier when trained on the CV feature extractor. RF exhibited superior performance in terms of classification scores compared to other classifiers, regardless of the feature extractor used. The CV feature extractor proved to be more effective than TF-IDF in capturing sentiment information from drug reviews. The RF classifier's ability to handle nonlinear relationships, robustness to noise and overfitting, and resilience to imbalanced data contributed to its exceptional performance. These discoveries offer significant knowledge to professionals and researchers involved in the realm of SA, particularly in the context of drug-related applications. The study's results suggest that selecting appropriate text preprocessing techniques and employing RF with CV can lead to accurate and reliable sentiment classification of drug reviews, thus facilitating decision-making processes in the healthcare domain.

In future work, creating a hybrid model that combines deep learning (DL) and machine learning (ML) techniques can improve drug sentiment analysis. This hybrid approach leverages DL's feature extraction abilities and ML's interpretability. Developing a user-friendly web application based on this model can assist doctors in making informed decisions by providing sentiment analysis results for specific drugs. This integration enables convenient access to valuable insights, enhancing patient care and optimizing treatment plans.

References

[1] R. K. Atkinson, K. Sabo, and Q. Conley, "The participatory web," Handbook of Technology in Psychology, Psychiatry and Neurology: Theory, Research, and Practice, pp. 91–120, Aug. 2012, doi: 10.1080/1369118X.2012.665935.

[2] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," Decis Support Syst, vol. 53, no. 4, pp. 675–679, Nov. 2012, doi: 10.1016/J.DSS.2012.05.022.

[3] K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," Artif Intell Med, vol. 64, no. 1, pp. 17–27, May 2015, doi: 10.1016/J.ARTMED.2015.03.006.

[4] M. N. Uddin, M. F. bin Hafi, S. Hossain, and S. M. M. Islam, "Drug Sentiment Analysis using Machine Learning Classifiers," International Journal of Advanced Computer Science and Applications, vol. 13, no. 1, pp. 92–100, 2022, doi: 10.14569/IJACSA.2022.0130112.

[5] T. Chen, P. Su, C. Shang, R. Hill, H. Zhang, and Q. Shen, "Sentiment Classification of Drug Reviews Using Fuzzy-rough Feature Selection," IEEE International Conference on Fuzzy Systems, vol. 2019-June, Jun. 2019, doi: 10.1109/FUZZ-IEEE.2019.8858916.

[6] S. Garg, "Drug recommendation system based on sentiment analysis of drug reviews using machine learning," Proceedings of the Confluence 2021: 11th International Conference on Cloud Computing, Data Science and Engineering, pp. 175–181, Jan. 2021, doi: 10.1109/CONFLUENCE51648.2021.9377188.

[7] J.-C. Na and W. Y. M. Kyaing, "Sentiment Analysis of User-Generated Content on Drug Review Websites," Journal of Information Science Theory and Practice, vol. 3, no. 1, pp. 6–23, Mar. 2015, doi: 10.1633/JISTAP.2015.3.1.1.

[8] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, and G. H. Gonzalez, "Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts," J Biomed Inform, vol. 62, pp. 148–158, Aug. 2016, doi: 10.1016/J.JBI.2016.06.007.

[9] P. Chandrasekar and K. Qian, "The Impact of Data Preprocessing on the Performance of a Naïve Bayes Classifier," Proceedings - International Computer Software and Applications Conference, vol. 2, pp. 618–619, Aug. 2016, doi: 10.1109/COMPSAC.2016.205.

[10] S. M. and A. R. Pais, "Classification of Phishing Email Using Word Embedding and Machine Learning Techniques," Journal of Cyber Security and Mobility, vol. 11, no. 3, pp. 279–320–279–320, May 2022, doi: 10.13052/JCSM2245-1439.1131.

[11] Islam, Md Babul, et al. "Detect deception on banking credit card payment system by machine learning classifiers." Second International Conference on Cloud Computing and Mechatronic Engineering (I3CME 2022). Vol. 12339. SPIE, 2022.

[12] Islam, Md Babul, et al. "Cost Reduce: Credit Card Fraud Identification Using Machine Learning." 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2022.

[13] Islam, Md Babul, et al. "A fiber wireless improved 5G network-based virtual networking system focused on equal bandwidth." 2021 2nd international symposium on computer engineering and intelligent communications (ISCEIC). IEEE, 2021.

[14] Islam, Md Babul, et al. "Twitter Opinion Mining on COVID-19 Vaccinations by Machine Learning Presence." Proceedings of Third Doctoral Symposium on Computational Intelligence: DoSCI 2022. Singapore: Springer Nature Singapore, 2022.

[15] Islam, Md Babul, et al. "Pandemic Outbreak Time: Evaluation of Public Tweet Opinion by Machine Learning." 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET). IEEE, 2022.

[16] Islam, Khandaker Sajidul, et al. "Blockchain Based New E-voting Protocol System without Trusted Tallying Authorities." 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT). IEEE, 2022.

[17] Sayeed, Md Abu, et al. "Bangladeshi Traffic Sign Recognition and Classification using CNN with Different Kinds of Transfer Learning through a new (BTSRB) Dataset." 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE). IEEE, 2023.