Drug Quality Classification Using Sentiment Analysis of Drug Reviews

Devesh Parmar[a] Harsh Katariya[b] Arpit Dobariya[c] Santosh Bharti[d] Rajeev Gupta[e]

[abcde] Computer Engineering, SOT, Pandit Deendayal Energy University, Gandhinagar, 382007 (E-mail: deveshparmar2003@gmail.com, santosh.bharti@sot.pdpu.ac.in)

ABSTRACT

Since the pandemic has emerged, there is shortage in remote access of medical resources, a lack of medical experts and healthcare professionals, absence of adequate instruments and medicines etc. As a result, large number of individuals have died. Due to unavailability of services, people started taking medical prescriptions on their own without appropriate consultation, which worsens the state of health than normal. Machine Learning in recent times has invaluable numerous applications. This research is intendent to present a drug review prediction having capability to reduce the pile of specialists. In this exploration we are building a drug review system which analyse patient's ratings to forecast sentiment based on different vectorization process like Count Vectorizer, TF-IDF, Bag of Words and manual analysis of characteristics that can help recommending the best medication for a disease given by various classification procedures. The predicted results were appraised by precision, accuracy score, AUC score and F1 score. The outcome show's that Random Forest Classifier outperforms all the other model with 94% accuracy.

Keywords *– Drugs, Sentiment Analysis, Natural Language Processing, Count Vectorizer, Machine Learning, Random Forest Classifier*

I. INTRODUCTION

The primary objective of sentiment analysis in NLP is to ascertain individuals' feelings, opinions, and attitudes regarding products, services, individuals, organizations, and other entities. An admirable task in analysing polarity known as aspect-level sentiment analysis levels to examine the sentiment polarity of a particular aspect in its verdict.

The world is experiencing a physician shortage as a result of the exponential increase in Pandemic, particularly in village areas where there are a smaller number of specialists than that of urban localities. It takes between six and twelve years to earn a doctor's license. As an outcome, the quantity of doctors can't be increased briskly in a little amount of time. In this challenging time, an e-medicine framework ought to be activated as much as possible [3].

Nowadays, clinical errors are very common. Every year, prescription errors affect over many peoples around the world. Because specialists compose the solution based on their very limited knowledge, specialists make mistakes when prescribing over 40% of medicines. Patients who need medical experts with substantial understanding of microscopic organisms, antibacterial assistance, and patients must select the top-level medication. Day to day, advanced research comes out with increasing number of drugs and trials that are available to clinical staff. As a result, it becomes increasingly difficult for doctors to select the appropriate medication or treatment for a patient firmed on incisions and previous medical records. Reviews of products had become an essential and intrinsic part of global product acquisition considering the rapid growth of the online products and the web-based profession industry. Individuals all over the world had utilized to look at reviews and websites first before making a purchase decision. Most of the previous research focused on E-Commerce rating expectations and proposals, but medical care or clinical therapies have received little attention. There is increment in the number of people worried about their health and diagnostics. According to a 2013 survey conducted by the Pew American Research Center, approximately 60% of adults searched online for health-related topics and 35% searched online for health condition diagnosis. The proposed framework is crucial because it can assist doctors and patients learn more about drugs for discrete health circumstances. A recommender structure is a standard framework that initiates a thing to the client, reliant upon their benefit and need. The following frameworks appoints the customers' surveys to interrupt down their sentiment and encourage a recommendation for their demanding need. With the help of feature engineering and sentiment analysis, the drug recommender system recommends medications for a particular condition based on feedback from patients. The process of determining and pulling out emotional data from language, including opinions and viewpoints, is known as sentiment analysis. Featuring engineering, on the alternative hand, is the method of creating additional features from existing ones. It makes models perform better. This investigation was broken up into five parts: The Introduction section explains the requirement for this research, the Related Work segment explains previous research in this area, the Methodology segment explains the methods utilized in this study, the Result segment uses a variety of metrics to evaluate the results of the applied model, the Discussion section explains the framework's limitations, and the Conclusion section.

II. RELATED WORK

An effort has been made to apply machine learning algorithms and deep learning algorithms considering the rapid advancement of AI recommendations for frameworks. Recommender frameworks are now widely used in restaurants, e-commerce, the travel industry, and other sectors. Unfortunately, there are only a few studies in the sector of drug proposition structure that use thought analysis. This is because medication reviews include medical terminology like infection label, reactions, and mock names that were utilized in manufacturing the drug. As a result, they are much more difficult to analyse.

An Online semantic based framework GalenOWL[5] was framed to assist physicians in locating medication-specific information, is presented in the study. A framework that uses a patient's infection, sensitivities, and drug interactions to recommend medications is shown in the paper. Utilizing international standards like ICD-10 and UNII, clinical data and terminology were first correctly combined with the clinical information before being converted to ontological terms for the purpose of empowering GalenOWL. Leilei Sun [2] looked at a lot of treatment records to figure out which prescription for treatment was best for patients. The motivation was to estima te the similarities between medical prescription records using a constructive opinionated clustering algorithm. In a similar methodology, the inventor developed a foundation for evaluating the suitability of the encouraged medication. New patients will get the leading treatment intention from this structure hinge on their demographics, location, and medical concerns. A patient's electronic medical record (EMR) will be utilized for testing from different clinics. The outcome demonstrates that this structure raises the rate of cure. Naïve Bayes (NB), Logistic Regression (LR), Random Forest Classifier (RF) and a Decision Tree (DT) were used to conduct multilingual sentiment analysis in this study.

The idea that the drug that should be prescribed should be based on the patient's capacity is the foundation of the study. For instance, if the patient's immunity is low, trustworthy medications should be suggested. Proposed a method for classifying risk levels to determine the patient's immunity. Over sixty threat constituents, including alcoholism, diabetes and hypertension, have been selected to determine the patient's extent to protect themselves from different infections. Additionally, a freely available paradigm system that makes use of a decision support system to assist physicians in determining first-line medical assistance was developed. Jiang Xiaohong et al [9]. focused at three distinct algorithms on treatment data: the DT algorithm, the RF, and the finely trained neural networks was selected for the medication proposal module due to its outstanding performance in the three distinct boundaries—model expectance, model competence, and model adaptability. In addition, the flaw inspecting system was proposed to guarantee administration quality, precision, and analysis. Hassan, Mohammad Mehedi et al. [10], and others came up with a proposal for a cloud-managed drug system (CADRE). CADRE can suggest drugs with top N associated prescriptions set up on the side effects of the patient. The medications were initially grouped into clusters using collaborative filtering techniques, as shown by the functional description data, for the foundation of this proposed framework. However, the model is switched to a cloud supported outlook utilizing tensor decomposition in order to improve the grade of the experience of medication implication after taking into consideration its shortcomings, such as being computationally expensive, having a cold start, and having insufficient information.[1]

III. METHODOLOGIES

A. Data Collection and Information

The dataset utilized for this research is Drug Review Dataset (Drugs.com) grabbed from the Kaggle Datasets. This dataset holds total eight features they are, patient id (numerical), name of drug used (text), date (text), review (text) of a patient, condition (text) of a patient, useful count (numerical) which suggest the number of individuals who found the review helpful, date (date) of review entry, and a 10-star patient rating (numerical) determining overall patient satisfaction. It has a total of 156919 instances.

B. Data Pre-processing and Visualisation

Enforced standard data pre-processing methods which includes inspecting null values, eliminating correlations, checking of outliers, discarding unnecessary features and values. Checking of mathematical relation between the features etc.

The following figure Fig. 1 shows top 10 drugs having rating of 10.

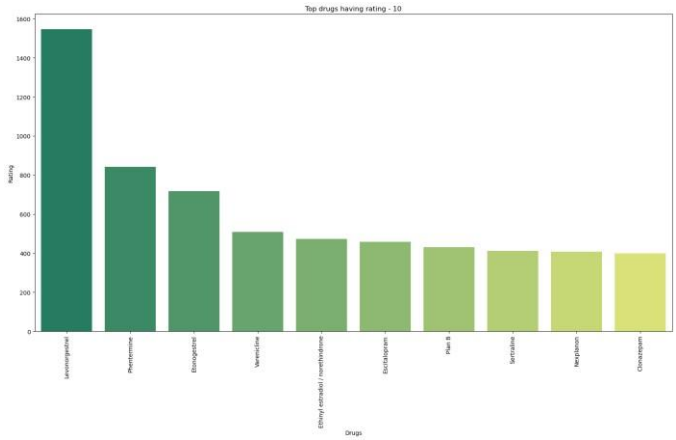

Fig-1 Drugs with highest Rating

From the above observation we are able to know that Levonorgestrel is most favoured drug having maximum positive reviews among all the other drugs respectively.

Fig 2. Shows top 10 conditions people are facing throughout the dataset.

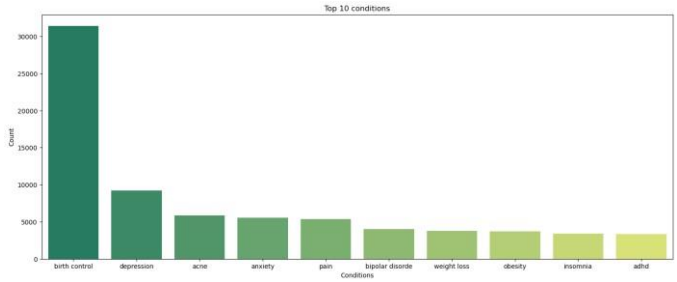

Fig-2 Top 10 Conditions

From the above figure we are able to get that birth control is the topmost condition having greater drug reviews among all the other conditions in the dataset.

Fig. 3 shows amount of positive and negative sentiments present in the review.

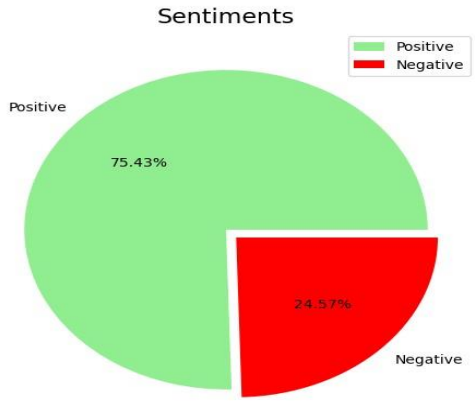

Fig-3 Sentiments Distribution

The review and drug name column were joined because they have conditions and review which helps is prediction. Also, it is necessary to clean up the review part to get maximum desired output from vectorization this procedure is defined as text processing. The cleaned reviews were in lower case and are free from special symbols to bypass the replication, and tokenization was executed for turning them into small parts known as tokens. Apart from these stopwords like “a, an, the, we, with” etc. were discarded from the corpus. We have taken word ‘not’ into consideration for better accuracy for negative sentiments of the data. The tokens were obtained by performing stemming and lemmatization on all the tokens. Our aim is sentiment analysis so the rating feature is labelled with 1 (5-10 rating) and 0 (0-4 rating) for positive and negative reviews.

C. Feature Extraction

A proper data pre-processing, setup is required to construct sentiment analysis classifiers following text pre-processing. Algorithms for machine learning cannot simply work with text; It ought to be transformed into numerical form. Specifically, vectors of numbers. The Bow, TF-IDF, and count vectorizer are three straightforward methods used in this study for feature extraction from text data. In addition to Bow, TF-IDF, and Count Vectorizer, a manual feature model was created by employing feature engineering methods to manually extract useful data from the review feature in the dataset.

1) Bow: It is an algorithm known as "Bag of words" which is responsible of counting the number of times each token appears in a review or document. Unigram or any arbitrary number of words can be used to describe a term or token. The fact that the Bow model takes into description all of the terms by not considering into report the fact that some of the terms in the corpus are extremely successive creates a massive matrix that is expensive to compute.

2) TF-IDF: A well-liked weighting method is TF-IDF, in which words are given weight other than. The aim was to indicate the terms that come up frequently in the data little weight, which means that T F-IDF evaluates significance rather than recurrence. The possibility of detecting a word in a document is referred to as term frequency.

3) Count Vectorizer: Based on the frequency of each word that appearing in the text, it is utilized to mutate a given text into a vector. When we want to convert each word in multiple such texts into vectors (for use in subsequent text analysis), this is helpful. Let’s look at a few examples of text from a document, each representing a list item: A matrix is created by Count Vectorizer, with each unique word described by a column and each text sample from the document represented by a row. The count of each word in that particular text sample is all that matters in determining the value of each cell.

4) Manual Features: A well-known idea that contributes to improving the model's accuracy is feature engineering. We utilized eight features, counting the condition column, useful count which was label encoded using the Scikit library's label encoder function. To determine the polarity of cleaned and uncleaned reviews, the Text blob toolkit was utilized.

D. K fold cross split

In this, we merged two datasets using various vectorizers and manual features. The input dataset than divided into K groups of samples of equivalent sizes. These samples are called folds. The prediction function employs K-1 folds for each learning set, while the remaining folds are utilized for the test set. Because it is simple to comprehend and produces results that are less biased than those produced by other methods, this approach is widely used in CVs. Here we have taken K = 30 for this research. By this its splits our dataset into K successive folds. After that, each fold is approached and used once as an approval while other K-1 persisting folds are used to structure the training set.

E. Smote

For avoiding the issue of class imbalance, narrowly the training data were subjected to a Synthetic Minority Over-Sampling Technique (Smote) following the K fold cross split. It is a procedure for synthesizing new data from existing data using oversampling. Linear interpolation is used for the selecting random minority instance 'x' and its k closest neighbour 'y' in the feature space to create the new class data for minor instances. Distribution of data after smote analysis is given below in Table-I.

Smote	Class	Train	Test
No	Negative	37254	1294
	Positive	114435	3936
Yes	Negative	114435	3936
	Positive	114435	3936

Table-I Smote Analysis

F. Classifier Model

To construct a classifier that could predict the sentiment, various machine-learning classification algorithms were utilized. The Bow, TF-IDF and Count Vectorizer model was used for experiments in NB, LR, perceptron and ridge classifier because the matrices are very scattered. On Word tokenizer and the manual features model, the Decision Tree, Random Forest classifiers were utilized. We only chose machine learning classification algorithms with faster predictions and shorter training times. For our research Random Forest classifier performed best by enabling parameter tunning of number of estimators and criterion. For this we have taken estimators as and criterion. One of the major challenges is computational power of the different classifiers so we have gone forward with those machine learning algorithms which has workable training time and can able to predict the outputs rapidly.

G. Parameter Tunning

Tunning for Hyperparameters is essential because they control overall etiquette of the machine learning model. Each models have different parameters depending upon the requirements. It is kind of values to be set for the machine learning model before the process for increasing the accuracy of model. For our research we have applied parameter tunning to random forest classifier with GridSearchCV which came out with following best parameters – (n_estimators =40, criterion= log_loss). The following parameters helped the model to outperform among all the other models.

H. Metrics

Five metrics were used to evaluate the predicted sentiment: precision, recall, f1score, accuracy and the AUC score.TP = True Positive or instances in which the positive class was predicted correctly; TN = True Negative or instances in which the negative class was predicted correctly; FP = False Positive or instances in which the positively class was predicted falsely; FN = False Negative or instances in which the model predicted the negative class falsely. The AUC score is used as a check on the ROC curve and helps to differentiate the algorithm's ability to compare classes. The ROC curve depicts the association between the True Positive Rate and the False Positive Rate at various thresholds.

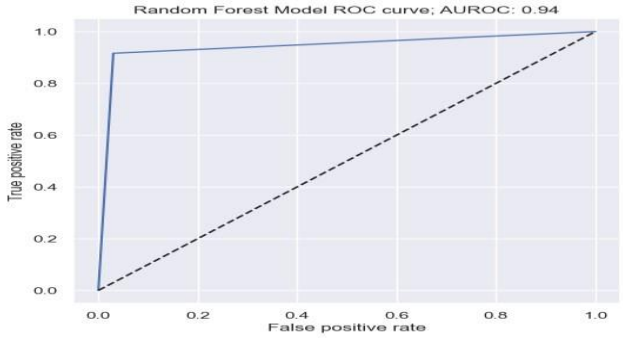

Fig-4 ROC curve of Random Forest Classifier

I. Drug Recommending System

Looking on to the metrics, the foremost predicted results were taken and concatenated together to obtained definite prediction. The normalized useful count was then multiplied by the merged results to build a comprehensive score for a given drug. In this system user have to enter drug name and review in order to predict the sentiment of the drug. As a result, output is obtained whether drug is preferable is or not. Overall system (webapp) is made with help of streamlit library.

IV. RESULTS

In this analysis, each review is either labelled as positive sentiment or negative sentiment relying upon the people's review. People rated 0-5 are rated as negative while rating greater than 5 are rated as positive. The reviews are passed through stemmer to remove last few characters from words. After that it is passed though lemmatizer to convert word to meaningful form. The Methods like TFIDF and Count vectorizer is used to represent the text and to enhance the model. Below shown Table I is the result of metrics from 4 different classifiers with the help of TFIDF Vectorization. From this Random Forest outperforms all the models with 93% accuracy while Multinomial Naïve Bayes underperformed with accuracy of 78%.

Model	Precision	Recall	F1	Accuracy	AUC
Multinomial Naïve Bayes	0.75	0.80	0.78	0.78	0.78
Logistic Regression	0.87	0.83	0.85	0.84	0.84
Decision Tree Classifier	0.89	0.85	0.87	0.87	0.83
Random Forest Classifier	0.97	0.91	0.94	0.93	0.94

Table II – TFIDF Vectorizer metrics

Table II shows the results of metrics of all four classifiers with the help of Count vectorization. From this similar to above Multinomial Naïve Bayes underperformed with accuracy of 79% which is greater than the TFIDF vectorizer results while Random Forest classifier outperforms from all the above classifiers with 94% accuracy.

Model	Precision	Recall	F1	Accuracy	AUC
Multinomial Naïve Bayes	0.78	0.80	0.79	0.79	0.80
Logistic Regression	0.85	0.90	0.87	0.88	0.87
Decision Tree Classifier	0.90	0.89	0.89	0.89	0.89
Random Forest Classifier	0.94	0.94	0.94	0.94	0.94

Table III – Count Vectorizer metrics

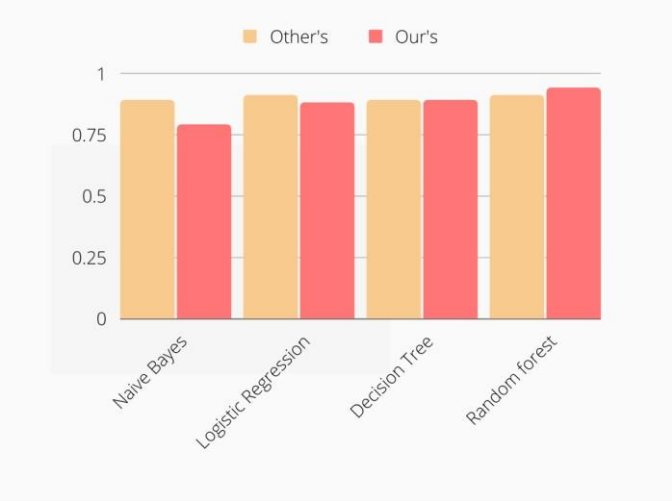

Fig -5 Comparison with others existing model

V. DISCUSSION

Although each all the five methods produced good and satisfactory results, this didn't indicate that the recommender system is prepared for use in real life problems. It still needs to be enhanced and improved. As predicted, the positive and negative sentiments standards differ, indicating that pre-processing of data should be enhanced by using definite methods. Classification algorithms must also use proper hyperparameter tunning to optimize and improve model's accuracy. The leading predicted outcome of each method was simply added to the recommendation framework. For superior results and comprehension, an actual representation of various predicted sentiment is required. The methodology that can be utilized to build a recommender system by extracting results from the features and performing classification is the sole focus of this paper.

VI. CONCLUSION

Reviews are increasingly becoming necessary for making decisions. Before we go outside, we read reviews to make sure we are making the right choices. Influenced by this research problem sentiment analysis of the reviews of drug was examined for building a recommendation system applying various classification algorithms, such as Multinomial Naïve Bayes, Logistic Regression applied to Count Vectorizer, TF-IDF, and algorithms such as Decision Tree, Random Forest applied to different method, sentiment analysis of drug analysis was examined in this research to assemble a review system. We tried to evaluate models based on five parameters: Precision, Accuracy, Recall, F1score, and AUC score. The Random Forest classifier with Count Vectorizer outperforms all the models with 94.5% Accuracy, having precision score of 0.9499, recall score of 0.9413 and roc-auc score of 0.9454 and F1 score of 0.9456. Whereas Naïve Byes, Logistic Regression and Decision Tree Classifiers only gave 79%, 88% and 89% respectively.

VII. REFERENCES

[1] Satvik Garg, 2021 Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning, 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 175-181.

[2] Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong, and Yanming Xie. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 1865–1874. DOI: https://doi.org/10.1145/2939672.2939866

[3] Jin-Cheon Na, 2015 Journal of Information Science Theory and Practice.

[4] Telemedicine, https://www.mohfw.gov.in/pdf/Telemedicine.pdf

[5] Shuvendu Das, 2022 Predicting Effectiveness of Drug from Patient's Review Proceedings of the Advancement in Electronics & Communication Engineering, available at SSRN.

[6] T. Venkat Narayana Rao, 2022 Medicine Recommendation System Based on Patient Reviews, international journal of scientific & technology research volume 9, issue 02.

[7] Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Semant 3, 14 (2012). https://doi.org/10.1186/2041-1480-3-14

[8] Sumit Mishra, 2021 Drug Review Sentiment Analysis using Boosting Algorithms. ISSN: 2456-6470, Volume-5 | Issue-4, pp.937-941.

[9] Mohammed Nazim Uddin, Md. Ferdous Bin Hafiz Sohrab Hossain, 2022 Drug Sentiment Analysis using Machine Learning Classifiers International Journal of Advanced Computer Science and Applications 13(1).

[10] Y. Bao and X. Jiang," An intelligent medicine recommender sys- tem framework," 2016 IEEE 11th Conference on Industrial Elec- tronics and Applications (ICIEA), Hefei, 2016, pp. 1383-1388, doi: 10.1109/ICIEA.2016.7603801.

[11] Zhang, Yin & Zhang, Dafang & Hassan, Mohammad & Alamri, Atif & Peng, Limei. (2014). CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies. Mobile Networks and Applications. 20. 348-355. 10.1007/s11036-014-0537-4.

[12] Felix Gräßer, 2018 Surya Kallumad Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and CrossData Learning DH '18: Proceedings of the International Conference on Digital Health Pages 121–125.

[13] Rajput A., 2019 Natural Language Processing, Sentiment Analysis and Clinical Analytics. arXiv.