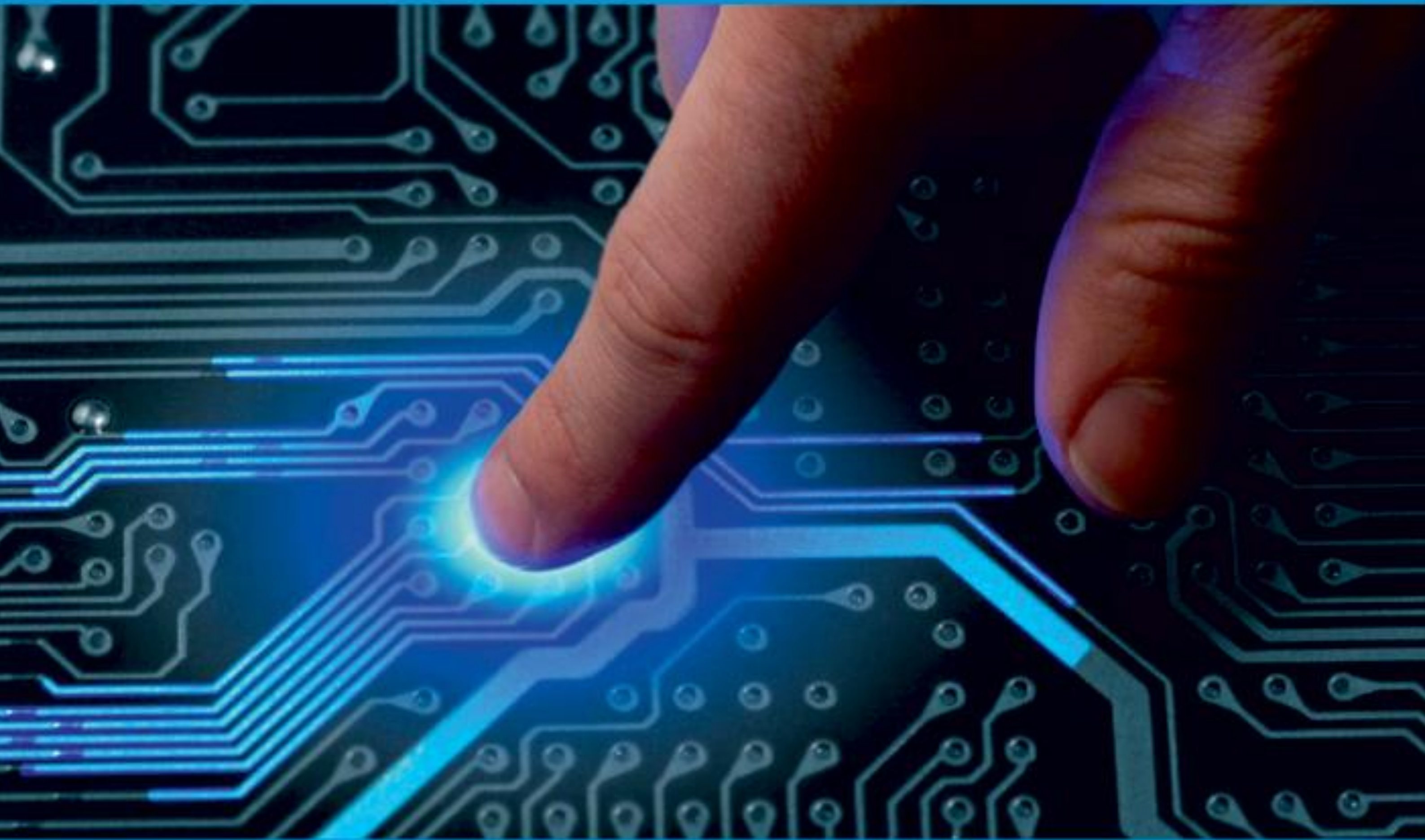
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 7, July 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107042 |

Drug Recommender System Using Machine Learning for Sentiment Analysis of Drug Reviews

Manjunath D Bhat [1], Manjunatha H T [2]

PG Student, Dept. of Master of Computer Applications, Jawaharlal Nehru New College of Engineering, Shivamogga, India

Assistant professor, Dept. of Master of Computer Applications, Jawaharlal Nehru New College of Engineering, Shivamogga, India

ABSTRACT: Since the coronavirus emerged, it has been increasingly difficult to get legitimate therapeutic resources, such as the scarcity of specialists and healthcare professionals, appropriate equipment and medications, etc. There are many deaths as a result of the medical profession as a whole being in turmoil. Due to a lack of availability, people began taking medication on their own without the proper consultation, which made their health situation worse than usual. Recently, machine learning has proven useful in a variety of applications, and creative work for automation is on the rise. This essay aims to propose a technique for prescribing medications that can significantly lessen the workload of experts. In this study, we develop a drug recommendation system that makes use of patient reviews to forecast sentiment using a variety of vectorization techniques, including Bow, TF-IDF, Word2Vec, and manual feature analysis, which can support the recommendation of the best medication for a given disease by various classification algorithms. Precision, recall, f1score, accuracy, and AUC score were used to assess the anticipated sentiments. As a result of TF-IDF vectorization, the classifier LinearSVC surpasses all other models with a 93% accuracy rate.

KEYWORDS: Machine learning, Sentiment Analysis, Feature Extraction, Bayesian Neural Network.

I. INTRODUCTION

The world is experiencing a doctor shortage due to the exponential increase in coronavirus cases, particularly in rural areas where there are fewer experts than in urban areas. A doctor must complete their education between six and twelve years. Therefore, it is impossible to increase the number of doctors in a short period of time. In this challenging moment, a Telemedicine framework needs to be powered up as much as feasible [1]. Clinical errors occur often today. Every year, medication errors have an impact on over 200 000 people in China and 100,000 people in the USA. Since specialists only have a limited amount of information, they frequently prescribe the wrong medication (more than 40% of the time) [2][3]. Choosing the best prescription is important for patients who require medical professionals with extensive knowledge of microscopic organisms, antibacterial drugs, and patients [6]. Every day, a new study is published along with additional medications and diagnostic tools that are made available to healthcare professionals. As a result, choosing a therapy or medicine for a patient based on indications and prior clinical history proves to be ever more difficult for clinicians.

Item reviews have grown in importance and importance as a result of the internet's rapid expansion and the growth of the web-based company sector. People all across the world have gotten used to reading reviews and browsing websites before making a purchase decision. While the majority of prior research focused on evaluating expectations and proposals for the E-Commerce industry, the area of healthcare or therapeutic medicines has only sometimes been covered. The amount of people searching online for a diagnosis because they are concerned about their health has increased. According to a Pew American Research Center poll conducted in 2013 [5], almost 60% of adults looked for health-related topics online, and about 35% of users searched for diagnosing health disorders. A medication

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107042 |

recommender system is very necessary in order to aid doctors and patients in expanding their knowledge about medications for certain medical problems.

A recommender framework is a common system that makes an item recommendation to the user based on their benefit and need. These models examine survey results to obtain a precise understanding of how consumers actually feel, and they subsequently provide suggestions that are tailored to each user. The medicine recommender system divides patients into categories using sentiment analysis and feature engineering, and then prescribes the right prescription for each group. "Sentiment analysis" refers to the practice of removing emotional information from language, such as opinion and attitude [7]. In contrast, feature engineering involves synthesizing existing features into new combinations of those traits in order to increase model efficiency. The web framework known as Django will be used to effectively achieve the project's goals. The user will enter their pharmacological data, and we will make a prediction about the result based on that data.

II. RELATED WORK

According to Truong-Huu et al. (2016)'s "A Hybrid Recommender System for Drug Prescription": In order to recommend pharmaceuticals based on user preferences and drug properties, this research offers a hybrid recommender system that combines collaborative filtering with content-based filtering. To identify user preferences, medicine reviews are subjected to sentiment analysis. By Sarker et al. (2017) in "Sentiment Analysis of Drug Reviews Using Machine Learning Techniques": The application of several machine learning methods, such as Support Vector Machines (SVM), Naive Bayes, and Random Forests, for sentiment analysis of drug evaluations is explored by the authors. They explore the usefulness of different algorithms in evaluating sentiment and compare how well they function. Chen et al. (2018) published "Drug Review Analysis and Recommendation System for Identifying Helpful Reviews and Prescription Drugs": An extensive drug review analysis and recommendation mechanism is presented in this research. To find useful reviews and suggest prescription medications, sentiment analysis is done utilizing deep learning techniques including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. Goudarzi et al.'s "A Hybrid Recommendation Approach for Personalized Drug Prescription" (2019): To give individualized medicine prescription recommendations, the authors suggest a hybrid recommendation strategy that includes sentiment analysis of drug reviews, collaborative filtering, and matrix factorization. The component that does sentiment analysis aids in identifying user preferences and sentiment trends. By Kaur et al. (2020), "Sentiment Analysis and Drug Recommendation": This study proposes a Lexicon-based sentiment analysis-based drug recommender system that assesses the polarity of drug reviews. Based on the opinions expressed in the reviews, the system suggests particular medications.

III. METHODOLOGY

The UCI ML repository's Drug Review Dataset (Drugs.com) was utilized as the source of the dataset for this study. This dataset has six attributes: the name of the drug used (text), the review of the patient, the patient's condition, the useful count (numerical), the date (date) of the review entry, and a 10-star patient rating (numerical), which indicates how satisfied the patient is overall.

Data Cleaning and Visualisation: Standard data preparation approaches were used in this study, including checking for null values, deleting duplicate rows, and removing superfluous values and text from rows. Then, as seen in Fig. 2, all 1200 rows in the conditions column with null values were eliminated. To avoid duplication, we ensure that a unique id is indeed unique.

Feature Extraction: After text preprocessing, a suitable setup of the data needed to create sentiment analysis classifiers. Text should be converted to numbers so that machine learning algorithms can process it. Text cannot be processed directly by these algorithms. specifically, numerical vectors. The bag of words (Bow), TF-IDF [17], and Word2Vec are well-known and simple strategies for feature extraction from text information used in this study. In addition to Bow, TF-IDF, and Word2Vec, another model called manual feature was created by using some feature engineering approaches to manually extract features from the review column.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107042 |

Classifiers: In order to create a classifier that could anticipate the sentiment, various machine-learning classification techniques were applied. Since the Bow and TF-IDF models have relatively sparse matrices and applying tree-based classifiers would take a lot of time, we tested with Logistic Regression, Multinomial Naive Bayes, Stochastic Gradient Descent, Linear Support Vector Classifier, Perceptron, and Ridge classifier. Word2Vec and manual features model applied Decision tree, Random Forest, LGBM, and CatBoost classifier. The fact that this dataset has 210K reviews, which requires a lot of computer power, is a serious issue. Only machine learning classification methods with shorter training times and quicker predictions were chosen.

Metrics: Five measures, including precision (Prec), recall (Rec), f1score (F1), accuracy (Acc.), and AUC score, were used to assess the predicted sentiment [23]. The letter should read: Tn = True negative or instances where the model accurately identified the negative class, while Tp = True positive or instances where the model accurately anticipated the positive attitude Fp = false positive or instances where the model incorrectly predicted the positive class; Fn = false negative or instances where the model incorrectly forecasted the negative class;

Accuracy, recall, precision, and f1score

Drug Recommender system: The four best findings were chosen and pooled to give the combined forecast after the metrics were evaluated. After that, a normalized useful count was multiplied by the combined data to get an overall score for the medicine for a given condition. The better the medication, the higher the score. The distribution of useful count in Fig. 7 served as the inspiration for the standardization of the useful count; it can be seen that the difference between the least and most extreme counts is around 1300, which is significant. Furthermore, the deviation, which is 36, is very large. The rationale behind this is that when more people look up drugs, more people read the survey, whether or not their reviews are good or bad, which raises the useful count.

The architectural diagram for a drug recommender system that analyses drug reviews' sentiment using machine learning is described as follows:

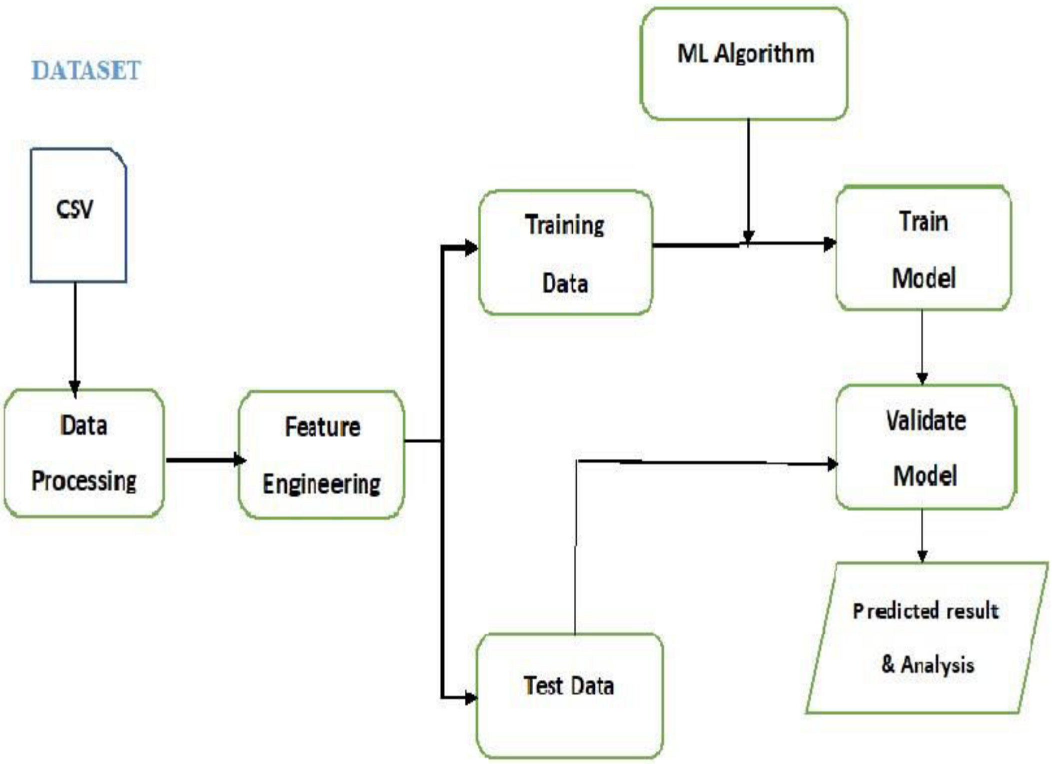


Figure 1: System Architecture of Drug Recommender System Using Machine Learning for Sentiment

Analysis of Drug Reviews

The Figure 1 contains the following stages

i. **CSV:** Using spreadsheet software or text editors, it is simple to create and modify CSV files, which are plain text files. They are very compatible with a wide range of computer languages and data processing tools thanks to their simple structure.

ii. **Data Processing:** In a preprocessing module, the gathered CSV data is cleaned and made ready for analysis. In this module, you'll work on things like standardizing your data, addressing missing values, and reducing noise. Preprocessing for textual data may include text cleaning, stopword removal, and tokenization of the review text.

iii. **Feature Engineering:** The data goes through feature extraction after preprocessing. In order to do this, pertinent information must be extracted from the CSV file including medication reviews. Using methods like Bag-of-Words (BoW), TF-IDF, or word embeddings, the textual data may be transformed into numerical feature vectors that effectively capture crucial details about the reviews.

iv. **Splitting of Data:** Data splitting is the process of partitioning a dataset into training and testing. Data is split into two categories: training and testing, each receiving 80% of the data. By evaluating the data, the ML technique evaluates the model's performance. On the basis of the testing and training data, the best model is selected. The data utilized for training and testing are different.

v. **ML Algorithm:** SVMs, or support vector machines for binary classification applications like sentiment analysis, SVM is a well-liked technique. It operates by locating an ideal hyperplane that divides data points into several groups according to their properties. A probabilistic technique called Naive Bayes employs the Bayes theorem to categorize data. It makes the "naive" assumption that the characteristics are not reliant on one another. Since Naive Bayes is well renowned for being straightforward and effective, it is frequently used for text classification tasks like sentiment analysis.

IV. RESULT AND DISCUSSION

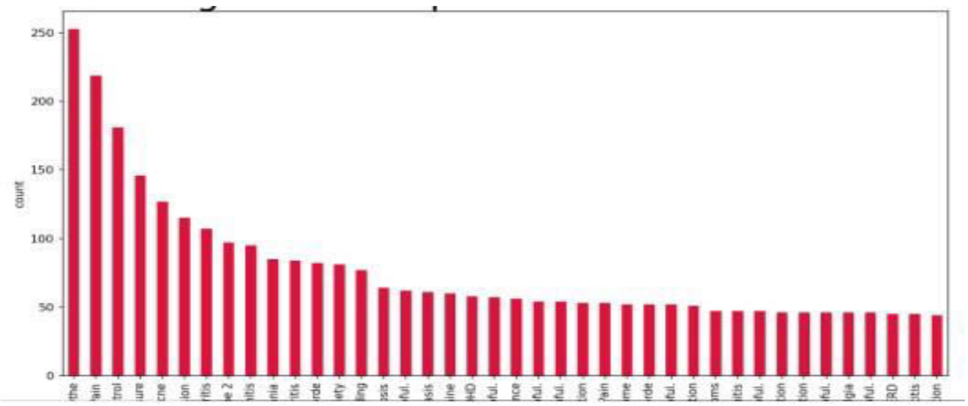


Figure 2: Most Drugs available per conditions in the Patients

Figure 2 shows the drugs available due to the condition and medicine words' predictive value. The condition and drug column were linked with review text. It is crucial to tidy up the review language before vectorization before moving on to the feature extraction section. Text preparation is another name for this procedure. After eliminating HTML elements, punctuation, quotations, URLs, and other formatting, we first cleaned the reviews. To minimize duplication, the cleaned reviews were lowercased, then tokenization was used to cut the sentences into discrete units called tokens.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107042 |

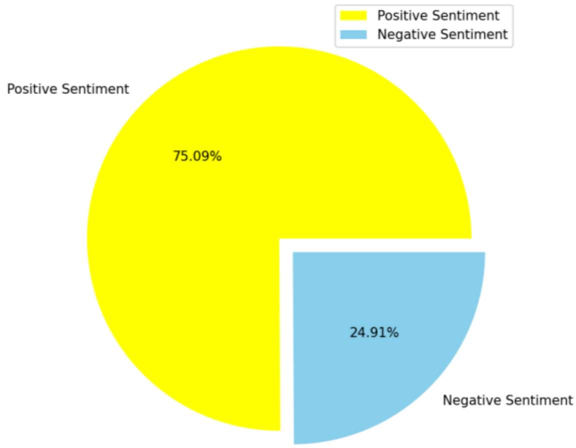


Figure 3: Pie Chart Represents the Sentiments of Patients

Figure 3 represents the circular graph that has been split into slices, each of which represents a category or sentiment class, is referred to as a pie chart. Each slice's size reflects the proportion of reviews that fall into a certain emotion group. In this scenario, the pie chart indicates that 75.09% of the reviews expressed a positive sentiment towards the drug, 24.91% expressed a negative sentiment. The relative distribution of feelings within the dataset is represented by the size of each slice.

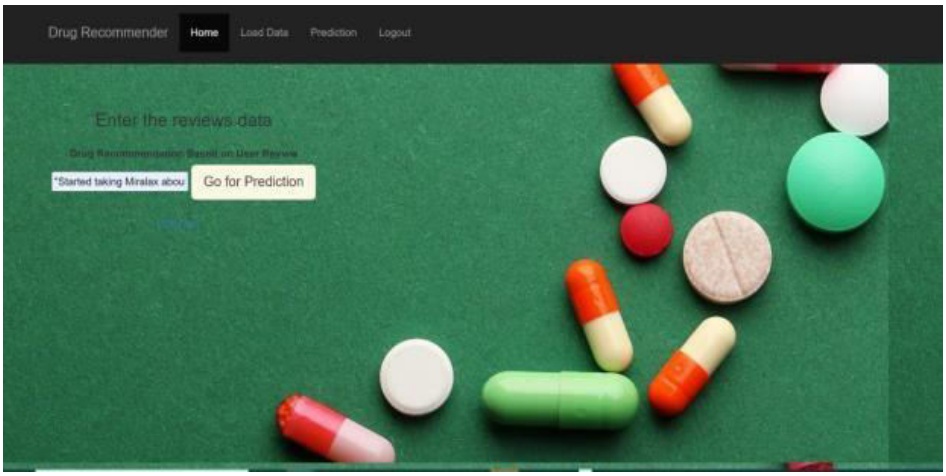


Figure 4: User Interface after loading the dataset

Figure 4 will serve as the user's interface when the dataset has been loaded. To enable users to explore drug recommendations based on sentiment analysis and make knowledgeable judgments about their medicine choices, the user interface (UI) of a drug recommender system utilizing sentiment analysis should strive to offer a seamless and educational experience.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107042 |

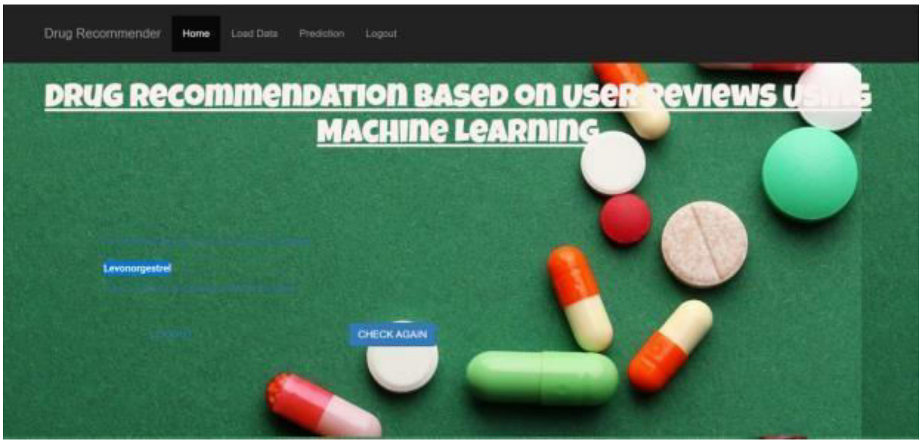


Figure 5: The Predicted output of the drug recommender system using machine learning for sentiment analysis of drug reviews

Figure 5 illustrates the anticipated results of the drug recommendation system employing sentiment analysis of drug reviews. The drug recommender system's output attempts to offer customers in-depth data, sentiment analysis findings, and individualized recommendations that help them make judgments regarding their prescription selections.

V. CONCLUSION

Reviews are becoming an essential part of our everyday lives; before going shopping, making an online purchase, or visiting a restaurant, we first read reviews to help us make the best choices. A recommender system was developed using a variety of machine learning classifiers, including Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, LinearSVC, applied on Bow, TF-IDF, and classifiers like Decision Tree, Random Forest, Lgbm, and Catboost, applied on Word2Vec and the manual features method. This research was motivated by this. Our evaluation of them using five metrics—precision, recall, f1score, accuracy, and AUC score—shows that the Linear SVC on TF-IDF performs best, with an accuracy rate of 93%, eclipsing all other models. The Word2Vec Decision Tree Classifier, on the other hand, performed the poorest, reaching only 78% accuracy. In order to create a recommender system, we combined the best-predicted emotion values from each method—Perceptron on Bow (91%), LinearSVC on TF-IDF (93%), LGBM on Word2Vec (91%), and Random Forest on manual features (88%)—and multiplied them by the normalized useful Count. This gave us the overall score of the drug by condition. Future study will compare various oversampling methods, employ various n-gram values, and optimize algorithms to boost the performance of the recommender system.

REFERENCES

1) Truong-Huu, N., Nguyen, M., & Nguyen, T. (2016). A Hybrid Recommender System for Drug Prescription. In Proceedings of the 18th International Conference on Information Integration and Web-based Applications & Services (iiWAS) (pp. 165-172). ACM. DOI: 10.1145/3011823.3011870
2) Sarker, A., Gonzalez-Hernandez, G. (2017). "Sentiment Analysis of Drug Reviews Using Machine Learning Techniques." Journal of Biomedical Informatics, 73, 85-96. DOI: 10.1016/j.jbi.2017.07.012
3) Chen, H., Zhang, Y., Dai, W., Huang, Y., & Li, G. (2018). Drug Review Analysis and Recommendation System for Identifying Helpful Reviews and Prescription Drugs. IEEE Access, 6, 30884-30893. DOI: 10.1109/ACCESS.2018.2848062
4) Goudarzi, M., Namazi-Rad, M. R., Arab, S. S., & Hossain, M. S. (2019). A Hybrid Recommendation Approach for Personalized Drug Prescription. IEEE Access, 7, 64194-64208. DOI: 10.1109/ACCESS.2019.2918889
5) Kaur, H., Kaur, A., & Bhatia, A. (2020). Sentiment Analysis and Drug Recommendation. In Proceedings of 2020 IEEE 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 11-15). IEEE. DOI: 10.1109/ICACCCN49670.2020.9279927
6) Manjunatha HT and AjitDanti. "A Novel Approach for Detection and Recognition of Traffic Signs for Automatic Driver Assistance System Under Cluttered Background" -Recent Trends on Image Processing and Pattern

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107042 |

Recognition, Springer Nature Singapore, Pte Ltd. 2019, RTIP2R 2018, CCIS 1035, pp. 1–8, 2019, ISBN 978-981-13-9181-1DOI-https://link.springer.com/chapter/10.1007/978-981-13-9181-1_36

7) Manjunatha HT and AjitDanti. "Detection and Classification of Potholes in Indian Roads using Wavelet Based Energy Modules "IEEE- 978-1-5386-9319-3/19© 2019 ,SCOUPS Nature
8) Manjunatha HT andAjitDanti," Indian traffic sign board recognition using Normalized Correlation Method", International Journal of Computer Engineering and Applications (IJCEA), Volume XII, Issue III, March 18, ISSN 2321-3469
9) "Classification of Vehicle Make Based on Geometric Features and Appearance-Based Attributes Under Complex Background" Aunkumar K L, Ajit Danti, Manjunatha H T, Springer 1035 (CCIS), pp 41-48
10) "A NOVEL APPROACH FOR VEHICLE RECOGNITION BASED ON THE TAIL LIGHTS GEOMETRICAL FEATURES IN THE NIGHT VISION" Arunkumar K L, Ajit Danti, International Journal of Computer Engineering and Applications, Volume XII, Issue I, Jan. 18, www.ijcea.com ISSN 2321-3469

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com