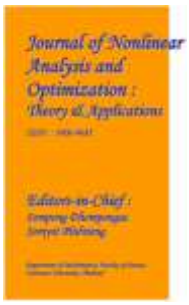
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2 : 2024
ISSN : **1906-9685**



DRUG RECOMMENDATION SYSTEM BASED ON SENTIMENT ANALYSIS OF DRUG REVIEWS USINGMACHINE LEARNING

[1]SK. NAYAB RASOOL, [2]MR.B. SURESH REDDY

[1]PG Scholar in the department of MCA at QIS College of Engineering & Technology, Vengamukkapalem, Ongole, AP, India.

[2] Assistant Professor in the department of MCA at QIS College of Engineering & Vengamukkapalem, Ongole, AP, India

ABSTRACT:

As per the report of the World Health Organization (WHO), diabetes has become one of the rapidly expanding chronic diseases that has affected the life of 422 million people all over the world. The number of deaths in Bangladesh due to diabetes has reached 28,065, which is 3.61% of the total deaths of Bangladesh, according to the latest data published by the WHO in 2018. So, we need to be concerned about the risks of diabetes disease. If we cannot take proper steps to diagnose diabetes at an early stage, eventually we have to face serious health issues. In this paper, we have shown the relation of different symptoms and diseases that cause diabetes so that we can help a person to diagnose diabetes at an early stage. Nowadays, machine learning classification approaches are well accepted by researchers for developing disease risk prediction models. Therefore, eleven machine learning classification algorithms such as Logistic Regression (LR), Gaussian Process (GP), Adaptive Boosting (AdaBoost), Decision Tree (DT), K-Nearest Neighbors (KNN),

Multilayer Perceptron (MLP), Support Vector Machine (SVM), Bernoulli Naive Bayes (BNB), Bagging Classifier (BC), Random Forest (RF), and Quadratic Discriminant Analysis (QDA) have been used in this study. Among all these machine learning classifiers, Random Forest (RF) classifier has showed the best accuracy of 98%. And its Area Under Curve (AUC) is also, the highest

Index: Logistic Regression (LR), Gaussian Process (GP), Adaptive Boosting (AdaBoost), Decision Tree (DT), K-Nearest Neighbors (KNN), Multilayer Perceptron

(MLP), Support Vector Machine (SVM), Bernoulli Naive Bayes (BNB), Bagging Classifier (BC), Random Forest (RF), and Quadratic Discriminant Analysis (QDA)

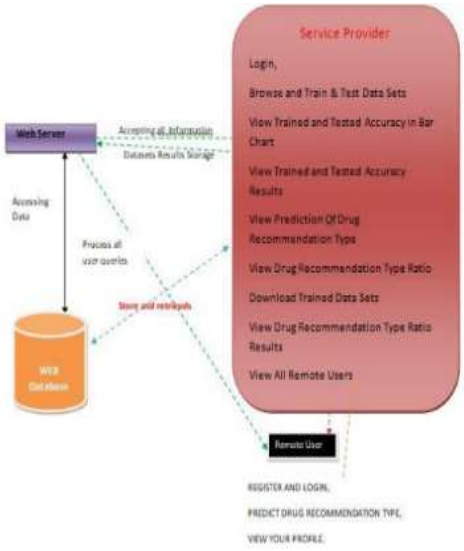
INTRODUCTION

With the number of corona virus cases growing exponentially, the nations are facing a shortage of doctors, particularly in rural areas where the quantity of specialists is less compared to urban areas. A doctor takes roughly 6 to 12 years to procure the necessary qualifications. Thus, the number of doctors can't be expanded quickly in a short time frame. A Telemedicine framework ought to be energized as far as possible in this difficult time [1].

Clinical blunders are very regular nowadays. Over 200 thousand individuals in China and 100 thousand in the USA are affected every year because of prescription mistakes. Over 40% medicine, specialists make mistakes while prescribing since specialists compose the solution as referenced by their knowledge, which is very restricted [2][3]. Choosing the top-level medication is significant for patients who need specialists that know wide based information about microscopic organisms, antibacterial medications, and patients [6]. Every day a new study comes up with accompanying more drugs, tests, accessible for clinical staff every day. Accordingly, it turns out to be progressively challenging for doctors to choose which treatment or medications to give to a patient based on indications, past clinical history. With the exponential development of the web and the web-based business industry, item reviews have become an imperative and integral factor for acquiring items worldwide. Individuals worldwide become adjusted to analyze reviews and websites first before settling on a choice to buy a thing. While most of past exploration zeroed in on rating expectation and proposals on the E-Commerce field, the territory of medical care or clinical therapies has been infrequently taken care of. There has been an expansion in the number of individuals worried about their well-being and finding a diagnosis online. As demonstrated in a Pew American Research center survey directed in 2013 [5], roughly 60% of grown-ups searched online for health related subjects, and around 35% of users looked for diagnosing health conditions on the web. A medication recommender framework is truly vital with the goal that it can assist specialists and help patients to build their knowledge of drugs on specific health conditions. A recommender framework is a customary system that proposes an item to the user, dependent on their advantage and necessity. These frameworks employ the customers' surveys to break down their sentiment and suggest a recommendation for their exact need. In the drug recommender system, medicine is offered on a specific condition dependent on patient reviews using sentiment analysis and feature engineering. Sentiment analysis is a progression of strategies, methods, and tools for distinguishing and extracting emotional data, such as opinion and attitudes, from language [7]. On the other hand, featuring engineering is the process of making more features from the existing ones; it improves the performance of models.

This examination work separated into five segments: Introduction area which provides a short insight concerning the need of this research, Related works segment gives a concise insight regarding the previous examinations on this area of study, Methodology part includes the methods adopted in this research, The Result segment evaluates applied model results using various metrics, the Discussion section contains limitations of the framework, and lastly, the conclusion section.

SYSTEM ARCHITECTURE



METHODLOGY

Proposed System Limitations:

Here are some proposed system limitations in a drug recommendation system based on sentiment analysis of drug reviews using machine learning:

Data Quality and Bias:

Sentiment Analysis Accuracy: Machine learning models for sentiment analysis may not perfectly capture the nuances of user reviews, especially those with sarcasm or mixed emotions.

Review Bias: Reviews may be skewed towards positive or negative experiences, leading the system to favor certain drugs.

Missing Information: Reviews may not mention all side effects or important details about a drug's effectiveness for specific users.

Limited Scope:

Individual Needs: The system cannot account for individual patient factors like medical history, allergies, and drug interactions.

Focus on Reviews: Reviews are subjective and may not reflect a drug's overall effectiveness or safety profile established through clinical trials.

Technical Limitations:

Limited Training Data: The system's performance depends on the quality and quantity of available drug review data.

Evolving Language: Language use and slang can change over time, requiring the model to be constantly

updated.

Overall System Limitations:

Not a Substitute for Medical Advice: This system should not replace consultations with qualified healthcare professionals.

Potential Misuse: Misinterpreting recommendations or using the system for self-diagnosis could be dangerous.

Ethical Considerations:

Privacy Concerns: The system should ensure user privacy and anonymize review data appropriately.

Transparency and Explainability: The system should be transparent about its limitations and how it arrives at recommendations. These limitations highlight the importance of using such a system as a potential starting point for discussion, not a definitive source for medical advice

EXPECTED RESULTS

Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning Now- a-days new diseases are attacking human world and corona virus is such disease and this diseases require lots of medical systems and medical human experts and due to growing disease medical experts and systems are not sufficient and patients will take medicines on their risk which can cause serious death or serious damage to patient body.

1) Upload Drug Review Dataset: using this module we will upload dataset to application

2) Read & Preprocess Dataset: using this module we will read all reviews, drug name and ratings from dataset and form a features array.

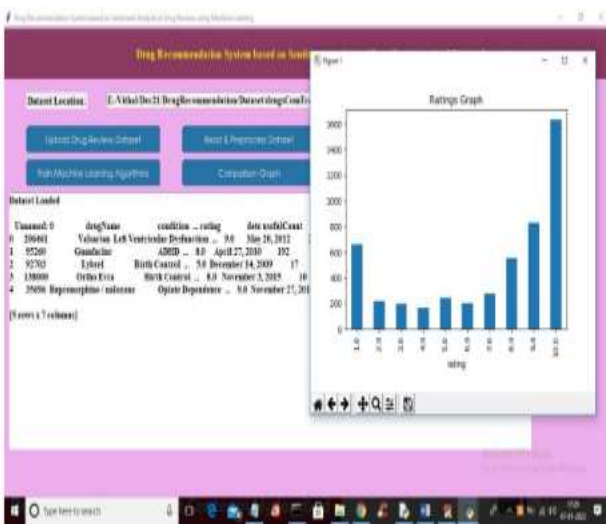
3) TF-IDF Features Extraction: features array will be input to TF-IDF algorithm which will find average frequency of each word and then replace that word with frequency value and form a vector. If word not appear in sentence, then 0 will be put. All reviews will be consider as input features to machine learning algorithm and RATINGS and Drug Name will be consider as class label.

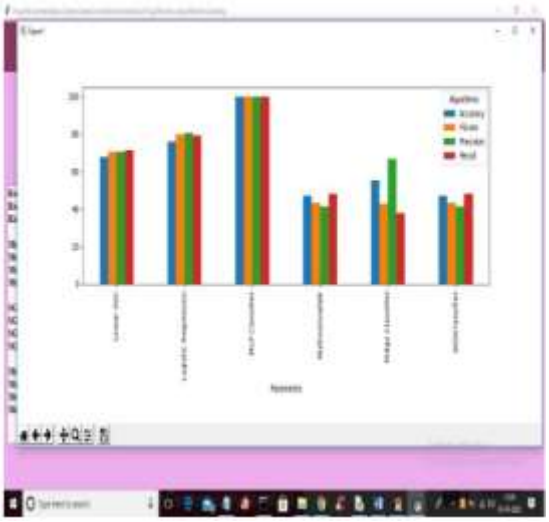
4) Train Machine Learning Algorithms: using this module we will input TF-IDF features to all machine learning algorithms and then trained a model and this model will be applied on test data to calculate prediction accuracy of the algorithm.

5) Comparison Graph: using this module we will plot accuracy graph of each algorithm

6) Recommend Drug from Test Data: using this module we will upload disease name test data and then ML will predict drug name and ratings.

To run project double click on ‘run.bat’ file to get b





VIII. CONCLUSION

Reviews are becoming an integral part of our daily lives; whether go for shopping, purchase something online or go to some restaurant, we first check the reviews to make the right decisions. Motivated by this, in this research sentiment analysis of drug reviews was studied to build a recommender system using different types of machine learning classifiers, such as Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, Linear SVC, applied on Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Cat boost were applied on Word2Vec and Manual features method. We evaluated them using five different metrics, precision, recall, f1score, accuracy, and AUC score, which reveal that the Linear SVC on TF-IDF outperforms all other models with 93% accuracy. On the other hand, the Decision tree classifier on Word2Vec showed the worst performance by achieving only 78% accuracy. We added best-predicted emotion values from each method, Perceptron on Bow (91%), Linear SVC on TF-IDF (93%), LGBM on

Word2Vec (91%), Random Forest on manual features (88%), and multiply them by the normalized useful Count to get the overall score of the drug by condition to build a recommender system.

Future work involves comparison of different oversampling techniques, using different values of n-grams, and optimization of algorithms to improve the performance of the recommender system.

IX. FUTURE ENHANCEMENT

Here's a concise take on future enhancements for the drug recommendation system:

Data Powerhouse: Integrate social media reviews, anonymized EHR data (regulations permitting), and realtime drug interaction data for richer recommendations.

Explainable and Interactive: Provide reasons behind recommendations, allow user feedback to improve the model, and consider a chatbot interface for guidance.

Advanced AI Techniques: Explore XAI for transparency, deep learning for complex sentiment, and potentially multimodal analysis for a more holistic understanding.

Global Reach and Trust: Expand language support, fight fake reviews, and prioritize user privacy throughout the system.

These enhancements can strengthen the system's capabilities while keeping user privacy and responsible data practices at the forefront.

REFERENCE

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

[2] Wittich CM, Burkle CM, Lanier WL. Medication errors: an overview for clinicians. Mayo Clin Proc. 2014 Aug;89(8):1116-25.

[3] CHEN, M. R., & WANG, H. F. (2013). The reason and prevention of hospital medication errors. Practical Journal of Clinical Medicine, .

[4] Drug Review Dataset, https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529# [5] Fox, Susannah, and Maeve Duggan. ”Health online 2013. 2013.” URL: http://pewinternet.org/Reports/2013/Health-online.aspx

[6] Bartlett JG, Dowell SF, Mandell LA, File TM Jr, Musher DM, Fine MJ. Practice guidelines for the management of community-acquired pneumonia in adults. Infectious Diseases Society of America. Clin Infect Dis. 2000Aug;31(2):347-82. doi: 10.1086/313954. Epub 2000 Sep 7. PMID: 10987697; PMCID: PMC7109923.

[7] Fox, Susannah & Duggan, Maeve. (2012). Health Online 2013. Pew Research Internet Project Report.

[8] T. N. Tekade and M. Emmanuel, ”Probabilistic aspect mining approach for interpretation and evaluation of drug reviews,” 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, 2016, pp. 1471-1476, doi: 10.1109/SCOPES.2016.7955684.

[9] 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

[10] 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

AUTHOR PROFILE

Mr. B. SURESH REDDY, done his M. Tech (Masters of Technology) in Arjun College of Technology & Sciences. At JNTU Hyderabad. Assistant Professor in the department of CSE at QIS College Of Engineering and Technology(Autonomous),Vengamukkapalem(V), Ongole, Prakasam. His areas of interest are Data Structures, Machine learning, and Web technologies.

Mr.Sk.Nayab Rasool currently pursuing Master of Computer Applications at QIS College of engineering andTechnology (Autonomous), Ongole,Andhra Pradesh. He Completed Bsc (Computer Science) from Acharya Nagarjuna University,Guntur, Andhra Pradesh. His areas of interests are Block chain, Artificial Intelligence& Data Science.