DRUG RECOMMENDATION SYSTEM BASED ON SENTIMENT ANALYSIS OF DRUG REVIEWS USING MACHINE LEARNING

MR.MAHEEDHAR VARA PRASAD[1], SAMANVITHA PERALA[2], SANJANA[3], ALEKYA VADIYALA[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT— Since coronavirus has shown up, inaccessibility of legitimate clinical resources is at its peak, like the shortage of specialists and healthcare workers, lack of proper equipment and medicines etc. The entire medical fraternity is in distress, which results in numerous individual's demise. Due to unavailability, individuals started taking medication independently without appropriate consultation, making the health condition worse than usual. As of late, machine learning has been valuable in numerous applications, and there is an increase in innovative work for automation. This paper intends to present a drug recommender system that can drastically reduce specialists heap. In this research, we build a medicine recommendation system that uses patient reviews to predict the sentiment using various vectorization processes like Bow, TF-IDF, Word2Vec, and Manual Feature Analysis, which can help recommend the top drug for a given disease by different classification algorithms. The predicted sentiments were evaluated by precision, recall, f1score, accuracy, and AUC score. The results show that classifier LinearSVC using TF-IDF vectorization outperforms all other models with 93% accuracy.

Index Terms— Drug, Recommender System, Machine Learn- ing, NLP, Smote, Bow, TF-IDF, Word2Vec, Sentiment analysis.

I. INTRODUCTION

With the number of coronavirus cases growing exponentially, nations are facing a shortage of doctors, particularly in rural areas where the number of specialists is lower compared to urban areas. A doctor typically requires 6 to 12 years to procure the necessary qualifications, meaning that the number of doctors cannot be expanded quickly in a short time frame. Therefore, a Telemedicine framework should be maximized during this challenging time.Clinical errors are common today. Over 200,000 people in China and 100,000 in the USA are affected each year due to prescription mistakes. More than 40% of doctors make mistakes when prescribing medication because they rely on their limited knowledge. Choosing the right medication is critical for patients, especially those who need specialists with broad knowledge about microorganisms, antibacterial medications, and patients. Every day, new studies and medications are introduced, making it increasingly difficult for doctors to choose the appropriate treatment or medications for a patient based on symptoms and past medical history.With the exponential growth of the internet and the e-commerce industry, product reviews have become an important factor in purchasing decisions worldwide. People are now accustomed to reading reviews and

checking websites before making a decision. While much of the previous research has focused on rating prediction and recommendations in the e-commerce field, healthcare and clinical therapies have not been as extensively covered. The number of people concerned about their health and searching for diagnoses online has increased. A Pew Research Center survey conducted in 2013 revealed that approximately 60% of adults searched online for health-related topics, and around 35% of users searched for health condition diagnoses online. A medication recommender system is crucial as it can assist specialists and help patients increase their knowledge about drugs for specific health conditions.

A recommender system is a conventional system that suggests items to the user based on their interests and needs. These systems use customer reviews to analyze sentiment and make recommendations that meet the user's exact needs. In the context of drug recommendation systems, medications are suggested based on patient reviews, using sentiment analysis and feature engineering. Sentiment analysis is a series of techniques, methods, and tools for identifying and extracting emotional data, such as opinions and attitudes, from language. On the other hand, feature engineering is the process of creating new features from existing ones to improve the performance of models.This research is divided into five sections: the Introduction, which provides an overview of the need for this study; the Related Works section, which summarizes previous research in this area; the Methodology section, which outlines the methods used in this study; the Results section, which evaluates the applied model using various metrics; the Discussion section, which highlights the limitations of the framework; and the Conclusion section.

II. LITERATURE SURVEY

A) TELEMEDICINE, HTTPS://WWW.MOHFW.GOV.IN/PDF/TELEMEDICINE.PDF

With a sharp increment in AI advancement, there has been an exertion in applying machine learning and deep learning strategies to recommender frameworks. These days, recom- mender frameworks are very regular in the travel industry, e-commerce, restaurant, and so forth. Unfortunately, there are a limited number of studies available in the field of drug proposal framework utilizing sentiment analysis on the grounds that the medication reviews are substantially more intricate to analyze as it incorporates clinical wordings like infection names, reactions, a synthetic names that used in the production of the drug

B) WITTICH CM, BURKLE CM, LANIER WL. MEDICATION ERRORS: AN OVERVIEW FOR CLINICIANS. MAYO CLIN PROC. 2014 AUG;89(8):1116-25.

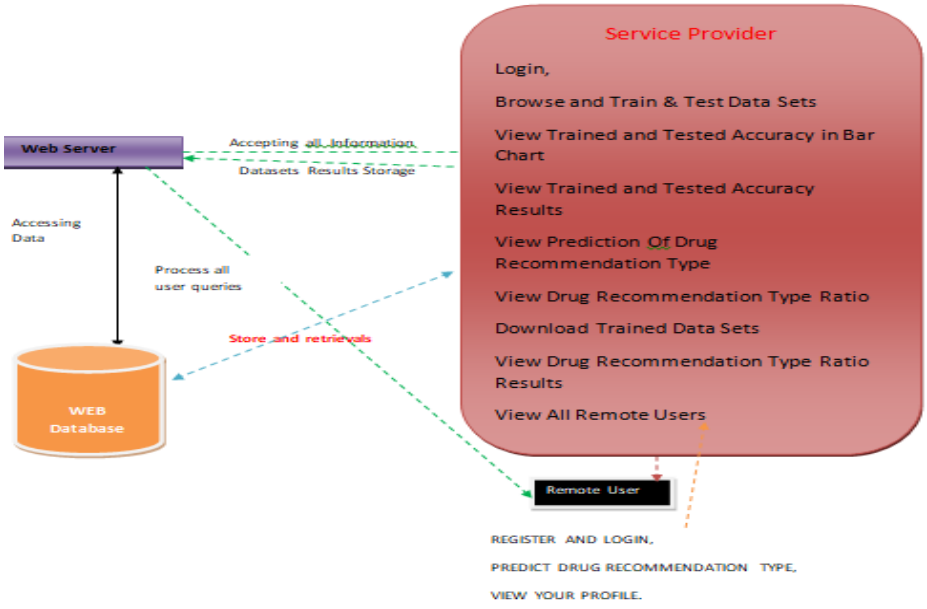
This study The study presents GalenOWL, a semantic-empowered online framework designed to assist specialists in discovering details about medications. The paper describes a system that suggests drugs for a patient based on the patient’s infection, sensitivities, and drug interactions. To enable GalenOWL, clinical data and terminology are first converted into ontological terms using global standards, such as ICD-10 and UNII, and then accurately integrated with clinical information.Leilei Sun examined large-scale treatment records to identify the best treatment prescriptions for patients. The approach involved

using an efficient semantic clustering algorithm to estimate similarities between treatment records. The author also developed a framework to assess the effectiveness of the recommended treatments. This system can recommend the best treatment regimens to new patients based on their demographic data and medical complications. An Electronic Medical Record (EMR) system was used, collecting data from multiple clinics for testing. The results demonstrated that this framework improved the cure rate.In another research, multilingual sentiment analysis was conducted using Naive Bayes and Recurrent Neural Network (RNN) models. Google Translator API was employed to convert multilingual tweets into English. The results showed that the RNN model, with an accuracy of 95.34%, outperformed Naive Bayes, which had an accuracy of 77.21%.

C) CHEN, M. R., & WANG, H. F. (2013). The reason and prevention of hospital medication errors. Practical Journal of Clinical Medicine, 4.

This paper study is based on the fact that the recommended drug should depend upon the patient's capacity. For example, if the patient's immunity is low, at that point, reliable medicines ought to be recommended. Proposed a risk level classification method to identify the patient's immunity. For example, in excess of 60 risk factors, hypertension, liquor addiction, and so forth have been adopted, which decide the patient's capacity to shield himself from infection. A web-based prototype system was also created, which uses a decision support system that helps doctors select first-line drugs.Xiaohong Jiang et al examined three distinct al- gorithms, decision tree algorithm, support vector machine (SVM), and backpropagation neural network on treatment data. SVM was picked for the medication proposal module as it performed truly well in each of the three unique bound- aries - model exactness, model proficiency, model versatility. Additionally, proposed the mistake check system to ensure analysis, precision and administration quality.Mohammad Mehedi Hassan et al. developed a cloud-assisted drug proposal (CADRE). As per patients' side effects, CADRE can suggest drugs with top-N related prescriptions. This proposed framework was initially founded on collabora- tive filtering techniques in which the medications are initially bunched into clusters as indicated by the functional description data. However, after considering its weaknesses like com- putationally costly, cold start, and information sparsity, the model is shifted to a cloud-helped approach using tensor decomposition for advancing the quality of experience of medication suggestion.Considering the significance of hashtags in sentiment anal- ysis, Jiugang Li et al.constructed a hashtag recommender framework that utilizes the skip-gram model and applied con- volutional neural networks (CNN) to learn semantic sentence vectors. These vectors use the features to classify hashtags using LSTM RNN. Results depict that this model beats the conventional models like SVM, Standard RNN. This explo- ration depends on the fact that it was undergoing regular AI methods like SVM and collaborative filtering techniques; the semantic features get lost, which has a vital influence in getting a decent expectation.

III. PROPOSED SYSTEM



Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

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 analy- sis 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, LinearSVC, ap- plied on Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Catboost 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%), LinearSVC on TF-IDF (93%), LGBM on Word2Vec (91%), Random Forest on manual features (88%), and multiply them by the normalized usefulCount to get the overall score of the drug by condition to build a recommender system. Future work involves comparison of different over- sampling techniques, using different values of n-grams, and optimization of algorithms to improve the performance of the recommender system.

REFERENCES

[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.

[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. 2000 Aug;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 Embed- ded 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

[11] V. Goel, A. K. Gupta and N. Kumar, "Sentiment Analysis of Mul- tilingual Twitter Data using Natural Language Processing," 2018 8th International Conference on Communication Systems and Net- work Technologies (CSNT), Bhopal, India, 2018, pp. 208-212, doi: 10.1109/CSNT.2018.8820254.

[12] Shimada K, Takada H, Mitsuyama S, et al. Drug-recommendation system for patients with infectious diseases. AMIA Annu Symp Proc. 2005;2005:1112.