DRUG RECOMMENDATION SYSTEM BASED ON SENTIMENTAL ANALYSIS OF DRUG REVIEWS

Tejaswini.P
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai

Vidhyasree.C
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai

Vilasini.G
Department of Computer Science and Engineering Panimalar Engineering College, Chennai

Dr.K. Valarmathi, ME., Ph.D., Professor
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai

Abstract - Machine learning-based drug recommendation system is based on emotive analysis of drug reviews. During the coronavirus outbreak, legitimate clinical resources were in low supply, such as specialists and healthcare workers, as well as suitable equipment and medicines. Individuals are dying because the entire medical society is under trouble. The number of doctors cannot be increased quickly. As a result, people began taking medications without contacting a doctor. This aggravated the situation. In this tough period, a telemedicine system should be activated as quickly as feasible. Machine Learning has more useful applications and creative work in the automation field. The goal of this project is to create a medicine recommendation system that will lessen the workload of specialists. Medicine recommendation system leverages patient evaluations to predict sentiment using various vectorization methods such as BOW, TF-IDF, and Word2Vec, and then employs multiple algorithms to assist patients in selecting the best drug for a certain ailment. The drug recommender system uses sentiment analysis and feature engineering to prescribe medication based on patient reviews. Sentiment analysis is a set of tactics, methods, and tools for detecting and extracting emotional information from text [1]. Feature engineering is the process of creating extra features from existing ones, with the goal of improving model performance. Precision, recall, f1score, accuracy, and AUC score were used to evaluate the anticipated sentiments. This result is 93% accurate.

Keywords:
Smote, Bow, TF-IDF, Word2Vec, Sentiment analysis, Drug, Recommender system, Machine Learning, NLP.

I. INTRODUCTION

Since the number of coronavirus infections is rapidly increasing, the countries are experiencing a doctor shortage, particularly in rural areas when compared to urban ones. Clinical errors are common these days. Every year, medication errors harm around 200 thousand people in China and 100 thousand in the United States. Over 40% of doctors make mistakes while prescribing because they are guided by their knowledge [2][3]. Patients who require doctors with broad knowledge of microscopic organisms, antibacterial drugs, and patients require top-level medication [4]. Every day, a new study uncovers new medications and testing. This makes it more difficult for doctors to determine treatment or drugs for a patient based on their symptoms. Item reviews have grown increasingly significant as the internet and web-

based technology have evolved. Individuals have become accustomed to reading reviews and visiting websites before making a purchasing decision. The number of people concerned about their health and looking for answers online. According to a Pew American Research Center survey conducted in 2013[5], 60 percent of adults searched online for health-related topics, and 35 percent searched for diagnosing health disorders. A medication recommender framework is required in order for specialists or doctors to aid patients in expanding their understanding of medications for certain health concerns. A recommender framework is a system that gives a user an item based on their benefit, need, and requirement. These frameworks use customer surveys to break down their feelings and make recommendations tailored to their specific requirements. The drug recommender system uses sentiment analysis and feature engineering to offer drugs based on a specific condition. Sentiment analysis is a set of tactics, methods, and instruments for recognising and extracting emotional information. Five sections of examination work: The introduction section gives a brief overview of why this research is needed. The related works section provides a succinct overview of past studies in this field, while the methodology section details the methods used in this study. The Result section examines model results using multiple metrics, while the Discussion part discusses the framework's constraints.

II. LITERATURE SURVEY

Intends to provide a medicine recommendation system that will significantly minimise the number of specialists needed. In this study, medicine recommendation system that uses patient reviews to predict sentiment using various vectorization processes such as Bow, TF-IDF, Word2Vec, and Manual Feature Analysis, which can help different classification algorithms recommend the best drug for a given disease. Precision, recall, f1score, accuracy, and AUC score were used to analyse the results, which revealed that the Linear SVC on TF-IDF outperforms all other models by 93 percent[6].

A health technology assessment is a multidisciplinary strategy to systematically and thoroughly evaluating medical and social issues associated to the usage of a health technology. Surveillance of social media platforms can provide vital information to the clinical community and decision makers on the effectiveness and safety of using health technologies on a patient, which can aid HTA recommendations[7].

The majority of related feature selection strategies for sentiment classification are unable to overcome issues with evaluating significant features, which lower classification performance. This research provides an improved hybrid feature selection technique based on machine learning approaches to improve sentiment categorization. Finally, the suggested technique's performance is evaluated using the Support Vector Machine (SVM) classifier. The accuracy, precision, recall, and F-measure are used to evaluate the performance[8].

Sentiment analysis is a research topic that involves categorising these views, opinions, and remarks. Researchers are currently investigating sentiment quantification, which deals with calculating relative frequency of a class of interest under the umbrella of sentiment analysis[9].

Sentiment Analysis (SA) is concerned with extracting sentiment (identification and classification) from unstructured text data such as product evaluations and microblogs. The use of supervised machine learning (SML), a method that employs datasets with predetermined class labels based on mathematical learning from a training dataset, is one of the most popular approaches for SA. To detect similarities and differences between the train and actual datasets, the Tree Similarity Index (TSI) and Tree Differences Index (TDI), a formula generated from tree structure, have been proposed[10].

We use supervised machine learning algorithms to develop and compare the performance of five classifiers for categorization issues. The top performing classifier was then used to predict the sentiment polarity of reviews, with an F1-score of 89.42 percent. Then, using a thematic analysis of positive and negative evaluations, we find themes that represent numerous aspects that influence the effectiveness of mental health apps in both positive and negative ways[11].

The usefulness of various sentiment categorization strategies, ranging from simple rule-based and lexicon-based approaches to more advanced machine learning algorithms, has been detailed in recent studies. Machine learning approaches have

fallen short in terms of accuracy, while lexicon based systems have suffered from a shortage of dictionaries and labelled data. This research presents an integrated framework for improving accuracy and scalability by bridging the gap between lexicon-based and machine learning approaches[12].

The current research looks at how sentiment analysis can be used in language acquisition. We built RESOLVE, a context-aware emotion synonym suggesting system for educational purposes, to achieve this goal. Importantly, the usage information for each emotion term is provided, including situation descriptions, definitions, and example sentences, in order to aid vocabulary development and word use[13].

The pharmaceuticals given in appropriate medical items suitable for the patient's current diagnostic are the foundation of effective pharmacotherapy. The fuzzy method to healthcare database analysis is a new tool for generating information that may be used in the ultimate decision-making process of drug selection in medical practise for a defined polymorbid group of patients[14].

Naive Bayes and Recurrent Neural Networks were used to accomplish multilingual sentiment analysis in this study (RNN). The data show that RNN outperformed Naive Bayes 95.34 percent of the time [15].

III. EXISTING SYSTEM

Many recommender systems use collaborative filtering (CF) as a successful strategy. The ratings provided to objects by users are the sole source of information for learning to produce recommendations in traditional CF-based techniques. However, in many applications, the ratings are typically relatively sparse, causing CF-based algorithms to severely decrease their recommendation performance. Deep neural networks have seen tremendous success in speech recognition, computer vision, and natural language processing in recent years. The use of deep neural networks in recommender systems, on the other hand, has gotten less attention.

IV. PROPOSED SYSTEM

We want to create neural network-based solutions to address the key challenge of recommendation via collaborative filtering with implicit feedback in this project. The algorithm will first obtain Drug reviews from the supplied URL, then parse and sanitise the reviews. For each review of the Drug, determine the positive and negative polarity. The Drug is then graded on the various aspects again, and the overall sentiment distribution of the Drug is provided. We create aesthetically beautiful and easy-to-understand graphs that provide summarised feedback using a combination of data aggregation techniques, NLP, linguistic analysis, and popular visualisation approaches. This is accomplished through a thorough sentiment analysis of the data.

V. METHODOLOGIES

1. LinearSVC

A Linear SVC's goal is to fit data you provide and provide a "best fit" hyperplane that divides or categorises your data. The "predicted" class can then be determined by feeding some features to your classifier.

2. Logistic Regression

This is a supervised machine learning technique for classification issues; it is a predictive analytic algorithm based on the probability notion.

3. Random Forest Classifier

This algorithm is made up of several decision trees. When creating each individual tree, it employs bagging and feature randomization in order to generate an uncorrelated forest of trees whose committee prediction is more accurate than that of any one tree.

4. Multinomial NB Classifier

The multinomial Naive Bayes classifier is good for discrete feature classification. Normally, integer feature counts are required for the multinomial distribution. Fractional counts, such as tf-idf, may also function in practise.

VI. RESULTS

Depending on the user's star rating, each review was categorised as positive or negative. Positive ratings range from one to five stars, while negative ratings range from one to five stars. All of the algorithms had similar findings, ranging from 89 to 91 percent accuracy. The LGBM model has the

best accuracy of 91 percent. After analysing all of the models, the combined model predictions of Perceptron (Bow), LinearSVC (TF-IDF), LGBM (Word2Vec), and Random Forest (Manual Features) were incorporated. The major goal is to ensure that each of the four models accurately classifies the recommended top medications.

VII. CONCLUSION

A novel deep learning-based framework for drug recommender systems has been suggested. The following are the paper's major contributions:
Deep Drug brings together modules for candidate generation, ranking, community detection, matrix factorization, and review mining.The presented framework is generic, and with a few tweaks, it may be applied to a variety of circumstances besides drug suggestion. It's simple to add a new data source to the framework. For instance, one may sample a set of frames from each Drug, input them to a convolutional neural network, and then add the final feature map to the Drug's representation vector.

VIII. FUTURE ENHANCEMENT

In this study, sentiment analysis of drug reviews was used to develop a recommender system employing a variety of machine learning classifiers, including Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, Linear SVC, and Bow, TF-IDF. Precision, recall, f1score, accuracy, and AUC score were used to evaluate them, and the Linear SVC on TF-IDF outperforms all other models by 93 percent. Future study will include a comparison of other oversampling strategies, as well as algorithm tuning to improve the recommender system's performance.

IX. REFERENCES

[1] Duggan, Fox, Maeve, Susannah, et al." Health online (2013) Pew Research Internet Project Report.

[2] Burkle CM , , Lanier WL ,Wittich CM, et al.(2014)"Medication errors: an overview for clinicians", Vol. 89, pp.1116-25.

[3] CHEN, M. R., WANG, H. F, et al. (2013), "The reason and prevention of hospital medication errors. Practical Journal of Clinical Medicine", Vol. 4.

[4] Bartlett JG, Dowell SF, Fine MJ , Mandell LA, File TM Jr, Musher DM, et al.(2000), "Practice guidelines for the management of community-acquired pneumonia in adults. Infectious Diseases Society of America". Vol.31, pp.347-82.

[5] Duggan , Fox, Maeve,Susannah, et al." Health online (2013) Pew Research Internet Project Report.

[6] Satvik Garg, (2021), "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning".

[7] Julie Polisena , Martina Andellini, Piergiorgio Salerno , Simone Borsci, Leandro Pecchia , Ernesto Iadanza, (2021), "Case Studies on the Use of Sentiment Analysis to Assess the Effectiveness Safety of Health Technologies: A Scoping Review".

[8] Nur Sysfiqah Mohd Nafis, Suryanti Awang, (2021), "An Enhanced Hybrid Feature Selection Technique Using Term
Frequency-Inverse Document Frequency and Support Vector Machine- Recursive Feature Elimination for Sentiment Classification".

[9] Kashif Ayyub, Saqib, Ehsan Ullah Munir, Muhammad Wasif Nisar, Momna Abbasi ,(2020)," Exploring Diverse Features for Sentiment Quantification Using Machine Learning Algorithms".

[10] Azwa Abdul Aziz, Andrew Starkey, (2020) ," Predicting Supervise Machine Learning Performances for Sentiment analysis Using Contextual-Based Approaches".

[11] Oladapo Oyebode, Felwah Alqahtani, Rita Orji, (2020), "Using Machine Learning and Thematic Analysis Methods to Evaluat Mental Health Apps Based on User Reviews".

[12] Farkhund Iqbal, Jahanzeb Maqbool Hashmi, Benjamin C.M.Fung,
Rabia Batool, Asad Masood Khattak, Saiqa, (2019), "A Hybrid Framework for Sentiment Analysis Using Genetic
Algorithm Based Feature Reduction".

[13] Mei-Hua Chen, Wei-Fan Chen, Lun-Wei Ku, (2019), "Application of Sentiment Analysis to Language Learning".

[14] Monika Fedorova , Daniela Perdukova , Zdenko Pirnik, Viliam Fedak Ondrej Sukel , Padmanaban Sanjeevikumar, (2019), "Fuzzy system as a promising tool for drugs Selection in medical practise".

[15] V. Goel, A. K. Gupta ,N. Kumar, (2019), "Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing.