International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

System for Recommending Drugs Based on Machine Learning Sentiment Analysis of Drug Reviews

Ankitha S[1], Dr. H N Prakash[2]

[1]Student, Dept. of CSE, Rajeev Institute of Technology, Hassan, Karnataka, India
[2]Professor and Head, Dept. of CSE, Rajeev Institute of Technology, Hassan, Karnataka, 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. The purpose of this paper is to present a system for prescribing drugs that can considerably minimize the workload of experts. Thus, a proposed system provides a drug recommendation platform that requires patient feedback to predict sentiment using a number of vectorization techniques.*

Key Words: Drug Recommendation, Machine Learning, Sentiment Analysis, Reviews, Ratings.

1. INTRODUCTION

As a result of the massive rise in coronavirus cases, there is a worldwide doctor shortage 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 amount of time. An infrastructure for Telemedicine needs to be pushed up as soon as possible in this difficult moment. Medical errors occur often today. Over 200 thousand people per year in China and 100 thousand people per year in the USA are associated with drug errors. More than 40% of the time specialists make errors while writing prescriptions, because specialists only have a limited amount of knowledge to base their decisions on. Choosing the best medication is essential for people who need specialists through in-depth understanding of microscopic organisms, antiviral medications, and consumers. Every day, new research is produced, and more medicines and testing equipment are made available to medical professionals. However, selecting a course of treatment or prescription for a particular patient based on reasons and early clinical history proves to be ever more difficult for clinicians. Item reviews have evolved into a crucial and essential aspect for purchasing things globally due to the internet's rapid expansion and the growth of the web-based company sector. Before making a purchase, people all over the world have grown conventional to reading comments and searching internet. Although the majority of prior research focused on evaluating expectations and proposals for the area of health care or scientific therapies, E-Commerce sector has been infrequently examined. The amount of people searching online for a diagnosis because they are concerned about their health has increased. A Pew American Research Center survey taken in 2013 found that around 35% of individuals checked for digital clinical diagnosis, whereas about 60% of people searched for health-related issues. In order to help physicians and patients learn more about medications for specific medical situations, a medication recommender system is obviously essential. A common method called a "recommender platform" suggests items to users based on their needs and benefits. These approaches use customer's reviews to analyze client emotion and offer a suggestion for their specific requirements. The drug recommender system uses sentiment classification and feature extraction to conditionally provide medications depending on patient feedback. Sentiment analysis is a development of approaches, techniques, and instruments for identifying and separating sentimental information from language, such as opinions and thoughts. On the other hand, Featuring engineering requires adding new features to the ones that already exist in order to enhance model performance.

2. PROBLEM STATEMENT

The world is experiencing a doctor shortage due to the exponential increase in coronavirus cases, particularly in rural areas where there are fewer specialists than in urban areas. A doctor must complete their education between six and twelve years. As a result, it is impossible to add more doctors in a short amount of time. An infrastructure for Telemedicine needs to be pushed up as soon as possible in this difficult moment.

3. LITERATURE SURVEY

Wittich CM et al. [1] The work in this paper is focusing on the pharmaceutical errors which are reviewed for practicing physicians with an emphasis on terminology, definitions, incidence, risk factors, disclosure and legal consequences. Numerous variables can contribute to medication errors, including those related to the drug, the patient, and the healthcare provider. One or more of the outcomes that doctors may encounter after making drug prescription errors includes losing the faith of their patients, civil

proceedings, criminal charges, and health board discipline. Various approaches have been tried with varying levels of success in preventing pharmaceutical errors. The ability of medical professionals to give their patients safe care may be improved by learning more about drug errors.

Bartlett JG et al. [2] In more than 10 years since the last Community-Acquired Pneumonia (CAP) proposal from the American Thoracic Society (ATS) / Infectious Diseases Society of America, the process for creating guidelines has altered, and new clinical data have been created (IDSA). Due to the expansion of information regarding the diagnostic, treatment, and managerial decisions for the patient care with CAP, we purposefully limited the extent of this framework to cover judgments from the point of medical diagnosis of pneumonia to the end of antibiotic treatment and carry chest image processing.

T. N. Tekade et al. [3] This article offers a brief summary of aspect mining methods as they apply to the search for new drugs. For the pharmaceutical industry, it is crucial to conduct research on the earliest possible detection of adverse drug reactions. A difficult task is identifying important topics from brief and noisy reviews. The probabilistic aspect mining model (PAMM) is suggested as a solution to this issue in order to find the aspects and subjects related to class labels. Due to a special characteristic of PAMM, it concentrates on discovering features specific to a single class rather than simultaneously detecting features for all categories during each operation.

Doulaverakis et al. [4] Drug-drug and drug-disease interactions can be difficult to identify, and finding the necessary information can be challenging due to the enormous number of medications that are already on the market and the ongoing pharmaceutical research. Although international standards have been created to facilitate effective information exchange, such as the ICD-10 classification and the UNII registration, medical staff still has to be regularly informed in order to efficiently identify drug interactions prior to prescription. In prior publications, the usage of Semantic Web technology has been suggested as a solution to this issue.

Gao, Xiaoyan et al. [5] The work in this paper is focusing on the recommendation of drugs with Graph Convolution Network, which mainly employs the mechanism of information propagation and embedding propagation layers to model high-order connectivity and elaborate the representation learning. The proposed system involves three key components namely the embedding layer, information propagation, and prediction layer. The work is mainly focuses on the accuracy rather than the evaluation of the recommendation system.

Li-Chih Wang et al. [6] The proposed system in this paper focused on recommending a parameter that is efficient by using a curing parameter recommendation system. The proposed system involves a voting method that is developed by seven Machine learning algorithms. These ML models are trained as the classifiers mainly to recommend a candidate representative medical datasets. The dataset with the highest frequency is chosen to be the recommended representative dataset. Long short-term memory networks are used for pre-adjusting to predict the heating curve.

Susannah et al. [7] In this research, a deep learning approach for health-based medical datasets is proposed. This approach automatically identifies what meal should be supplied to which person depending on the condition and other parameters like age, race, body weight, calories, fat, sodium, protein, fiber and cholesterol. The integration of deep learning and machine learning methods such as regression analysis, naive bayes, recurrent neural networks, multilevel perceptrons, gated recurrent units, and long short-term memory (LSTM) is the main goal of this study framework. The characteristics of these IoMT samples were evaluated and further processed prior to using machine learning, deep learning, and other learning-based methods.

4. OBJECTIVES

1. Analyzing the internal workings of our proposed system using individual and group machine learning techniques like regression analysis and naive bayes, as well as deep learning algorithms like GRU, RNN and LSTM.

2. Giving a thorough explanation of how our system functions in accordance with the product and patient disease specifications.

3. Examining how our AI and deep learning systems behave in order to better grasp the nature of the patient's problems and the medications they should take at the right time.

4. We demonstrated using a study of our machine learning and deep learning that various patient conditions have various recommender proofs, which may call for various treatments and particular care.

5. SYSTEM ARCHITECTURE

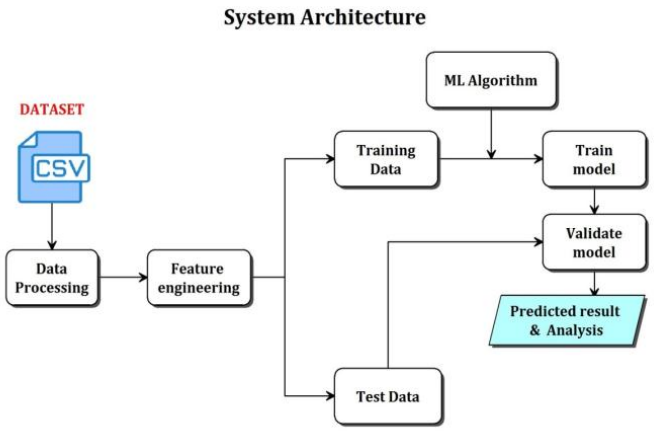

Figure 1: Representation of System Architecture

The above figure determines the system architecture of the proposed system. The system architecture involves following steps:

A. Data Gathering and Preprocessing

Machine Learning needs models and a lot of data in order to work. The procedure of gathering signals that monitor actual physical situations and converting the obtained results into electronic integer values which a computer can manipulate is known as Data Gathering. Processing primary data involves the subsequent procedures. It is necessary to combine a huge amount of raw data obtained from field surveys in order to compare the details of individual responses. A method for transforming unclean data into clean data sets is known as Data Preprocessing. Real-world information is consistently inaccurate and missing of particular behaviors or patterns. It is also frequently inconsistent and incomplete.

B. Feature Selection and Data Preparation

In order to create attributes for machine learning algorithms, one must use domain information from the data. The technique used here is called feature engineering. By generating features from input data that assists in the machine learning model, feature extraction can improve the prediction capacity of machine learning algorithms. In machine learning, feature engineering is the essential skill that distinguishes significantly among a successful model and a poor model. The concept of "feature engineering" involves taking raw data and turning it into features that the predictive models can use to more accurately depict the underlying issue. The practice of grouping and categorizing data based on particular characteristics is known as Data Classification. It may be done either in accordance with numerical characteristics or in accordance with attributes.

C. Model Construction and Model Training

The act of training an ML model involves providing the learning algorithm with training set to use as a learning resource. The model artifact produced during training is recognized as a "Machine-Learning model". The correct solution sometimes referred to as a goal or target attribute, needs to be incorporated in the training data. The learning method constructs an ML model that represents these patterns by looking for patterns in the training data that relate the characteristics of the input data to the target.

D. Model Verification and Outcome Evaluation

The model is employed to fresh input during the testing phase. There are two distinct samples for the training and test data. Designing a machine learning technique with the intention of performing it effectively. Generalize well to fresh data in the test set as well as the training set. Real-time data will be passed for the prediction when the built model has been evaluated. Once a forecast has been made, the result will be examined for the most important data.

6. RESULTS

The drug review sample utilized in this study was obtained from the UCI ML resource. This data comprises six components: the name of the drug taken, the review of the patient, the patient's status, the valuable count, which indicates the amount of people who encountered the review beneficial, the date of the review entry, and a 10-star patient rating that indicates how satisfactory the patient is overall. According to the user's star rating, each review in this work was categorized as either positive or negative. Positive ratings are those with five or more stars, whereas negative ratings vary from one to five stars.

In figure 2 we can see top 20 medical conditions with the greatest number of treatment options. One factor to observe in this figure is that there are two green bars, which indicate the criteria that have little significance.

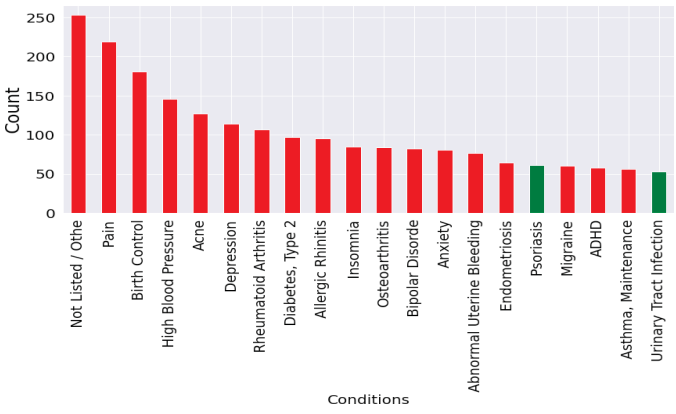

Figure 2: Most recommended drugs per conditions

Figure 3 demonstrates that the features used are effective pattern predictors with high accuracy and little error.

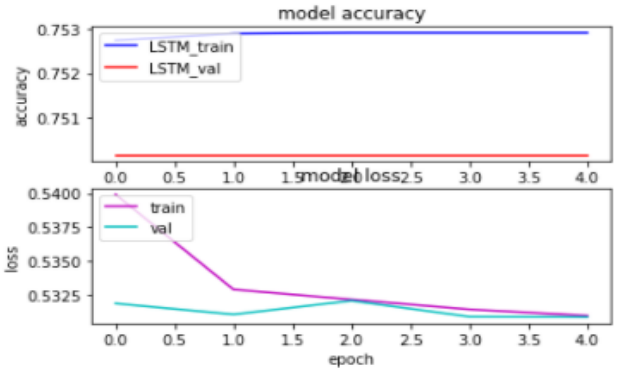

Figure 3: Model accuracy vs loss

Figure 4 displays the top four medications that our algorithm recommends for the five top medical issues including acne, contraception, high blood pressure, anxiety and depression.

condition	drugName	Score
Acne	Retin-A	0.069334
Acne	Atralin	0.088545
Acne	Magnesium hydroxide	0.088545
Acne	Retin A Micro	0.097399
Birth Control	Mono-Linyah	0.005448
Birth Control	Gildess Fe 1.5 / 30	0.005987
Birth Control	Ortho Micronor	0.006149
Birth Control	Lybrel	0.027766
High Blood Pressure	Adalat CC	0.303191
High Blood Pressure	Zestril	0.305851
High Blood Pressure	Toprol-XL	0.362589
High Blood Pressure	Labetalol	0.367021
Pain	Neurontin	0.158466
Pain	Nortriptyline	0.171771
Pain	Pamelor	0.231829
Pain	Elavil	0.304513
Depression	Remeron	0.124601
Depression	Sinequan	0.146486
Depression	Provigil	0.240185
Depression	Methylin ER	0.328604

Figure 4: Top four medications proposed for the five major conditions

7. CONCLUSION

Whether we are purchasing, buying products online or eating out, reviews are gradually becoming a part of our daily routine. We use reviews to help us make the best choices. Multiple Machine Learning techniques were used to construct a recommender system which includes Perceptron, Multinomial Naive Bayes, Logistic Regression, Ridge classifier and LinearSVC implemented on TF-IDF, Bow and classifiers like LGBM, Decision Tree, and Random Forest. Our examination of models using five main metrics: f1score, validity, recall, precision and AUC score reveals that the Linear SVC using TF-IDF outperforms all other models with 93 percent accuracy. On the other hand, the Word2Vec Decision tree algorithm scored the worst, reaching only 78% accuracy. We integrated the top expected sentiment values from each strategy LGBM on Word2Vec (91%) Perceptron on Bow (91%) Random Forest on manual features (88%) LinearSVC on TF-IDF (93%) and combined them by the standardized useful count to establish a recommender system. This provided us with the drug's total score for each condition. In order to enhance the effectiveness of the recommender system, future work will evaluate various oversampling techniques, use alternative n-gram values, and simplify algorithms.

REFERENCES

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

[2] 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.

[3] 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.

[4] Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Semant 3, 14 (2012).

[5] Gao, Xiaoyan, Fuli Feng, Heyan Huang, Xian-Ling Mao, Tian Lan, and Zewen Chi. "Food recommendation with graph convolutional network." Information Sciences 584 (2022): 170-183.

[6] Chen, Yu-Xiu, Li-Chih Wang, and Pei-Chun Chu. "A medical dataset parameter recommendation system for an autoclave process and an empirical study." Procedia Manufacturing 51 (2020): 1046-1053.

[7] Fox, Susannah and Duggan, Maeve. (2012). Implementing a Machine Learning Model to Realize an Effective IOMT-Assisted Client Nutrition Recommender System. Pew Research Internet Project Report.

[8] J. Ramos et al., "Using tf-idf to determine word relevance in document queries," in Proceedings of the first instructional conference on machine learning, vol. 242, pp. 133-142, Piscataway, NJ, 2013.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique, 2011, Journal of Artificial Intelligence Research, Volume 16, 2020.

[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.

[11] V. Goel, A. K. Gupta and N. Kumar, "Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing," 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2018.

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