IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DRUG RECOMMENDATION SYSTEM

[1]Yash Ritesh Tanna, [2]Vaibhav Avinash Parmar, [3]Surakshit Shivaji Bhalkeshware, [4]Abhinav Maindre,

[5]Pragati Malusare

[1]Student, [2]Student, [3]Student, [4]Student, [5]Professor
[1]Computer Engineering Department,
[1]Indira College of Engineering and Management, Pune, Maharashtra, India

ABSTRACT: As a result of the coronavirus, access to legitimate clinical resources has worsened significantly, including shortages of specialists and health workers, insufficient equipment, and drug shortages. Due to the emergency of the entire medical community, many individuals died. Individuals started taking medication themselves without proper consultation due to unavailability, which made their health condition more serious than usual. A growing number of applications are using machine learning and innovative work is being done in automation. In this project, a drug recommendation system is presented with the aim of significantly reducing the burden on specialists. Using patient reviews, we developed a drug recommendation system that uses various vectorization processes such as Bow, TF-IDF, Word2Vec, and Manual Feature Analysis to predict sentiment, which can be used to recommend the most appropriate drug for a given disease based on different classification algorithms. AUC, precision, F1 score and accuracy were used to evaluate the predicted feelings. In emergency situations such as pandemics, floods or cyclones, a medical referral system can help. In the era of machine learning (ML), recommender systems produce more accurate, faster, and more reliable clinical predictions at minimal cost. As a result, these systems maintain better performance, integrity and privacy of patient data in the decision-making process and provide accurate information at all times. Therefore, we present drug recommendation systems to improve the equity and safety of infectious disease treatment. To reduce side effects, medications are recommended based on the patient's previous health profile, lifestyle and habits. A system like this could be useful in recommending safe drugs to patients, especially during medical emergencies.

Keywords: TF-IDF, Manual Feature, Machine Learning (ML)

I. INTRODUCTION

With the exponential increase in the number of coronavirus cases, nations are facing a shortage of doctors, especially in rural areas where the number of specialists is less compared to urban areas. It takes a doctor approximately 6 to 12 years to obtain the necessary qualifications. Therefore, the number of doctors cannot be rapidly expanded in a short period of time. The framework of telemedicine should be strengthened as much as possible in this difficult time [1].

With the exponential development of the web and the web based business industry, item reviews have become a necessary and integral factor in acquiring items worldwide. Individuals all over the world adapt to analyze reviews and websites first before deciding to buy a thing. While most past surveys have focused on evaluation expectations and proposals in the field of e-commerce, the field of medical care or clinical therapies has rarely been addressed.

There has been an increase in the number of individuals concerned about their health and seeking diagnosis online. According to a 2013 Pew American Research Center survey [5], roughly 60% of adults have searched for health-related topics online, and about 35% of users have searched for diagnoses of health conditions on the Web. A Medicines Recommendation Framework is really essential with the goal that it can help specialists and help patients build their knowledge of medicines for specific medical conditions. A recommender framework is a conventional system that suggests an item to the user depending on its benefit and necessity. These frameworks use customer surveys to dissect their sentiment and design recommendations for their exact need. In a drug recommendation system, medicine is offered under specific conditions dependent on patient reviews using sentiment analysis and feature engineering. Sentiment analysis is the development of strategies, methods, and tools for discerning and extracting emotional data, such as opinions and attitudes, from language [7].

On the other hand, Featuring engineering is the process of creating more features from existing ones; improves the performance of models. This thesis is divided into five segments: Introduction section, which provides a brief overview of the need for this research, Related papers segment provides a brief overview of previous examinations in this field of study, Methodology section includes the methods used in this research, Results segment evaluates the results of the applied model using different metrics, the Discussion section contains the limitations of the framework and finally the conclusion section.

II. Problem Definition

The world is experiencing a shortage of doctors due to the exponential increase in coronavirus cases, especially in rural areas where there are fewer specialists than in urban areas. A doctor must complete his education between six and twelve years. As a result, it is not possible to add more doctors in a short time. The infrastructure for telemedicine needs to be upgraded as soon as possible at this difficult time.

III. Objectives

The sector is experiencing a shortage of doctors because of the exponential growth in coronavirus instances, mainly in rural areas where there are fewer specialists than in urban regions. A health practitioner should complete his training between six and twelve years. As a end result, it is not viable to feature extra doctors in a quick time. The infrastructure for telemedicine desires to be upgraded as quickly as possible at this tough time.

IV. Literature Survey

Witch CM et al. [1] The work in this article focuses on pharmaceutical errors, which are reviewed for the general practitioner, with an emphasis on terminology, definitions, incidence, risk factors, disclosure and legal implications. A number of variables can contribute to medication errors, including those related to the drug, the patient, and the health care provider. One or more of the outcomes that doctors may face after making medication errors include losing the trust of their patients, civil lawsuits, criminal charges, and medical board discipline. Various approaches have been tried in the prevention of pharmaceutical errors with varying degrees of success. Physicians' ability to provide safe care to their patients can be improved by learning more about medication errors. Bartlett JG et al.

[2] In the more than 10 years since the last Community-Acquired Pneumonia (CAP) proposal from the American Thoracic Society (ATS) / Infectious Diseases Society of America, the guideline development process has changed and new clinical data (IDSA) have been created. Given the proliferation of information on diagnostic, treatment, and management decisions for the care of patients with CAP, we intentionally limited the scope of this framework to cover judgments ranging from the medical diagnosis of pneumonia to the discontinuation of antibiotic therapy and the wearing of chest imaging. T. N. Tekade et al.

[3]This article offers a brief summary of facet mining methods as they are used in the search for new drugs. It is essential for the pharmaceutical industry to carry out research aimed at detecting adverse drug reactions as quickly as possible. It is a difficult task to identify important themes from short and noisy reviews. As a solution to this problem, a Probabilistic Aspect Mining Model (PAMM) is proposed to find aspects and objects related to class labels. Due to the special characteristic of PAMM, it focuses on discovering features specific to one class rather than simultaneously discovering 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 vast number of drugs already on the market and ongoing pharmaceutical research. Although international standards have been created to facilitate efficient information exchange, such as ICD-10 classification and UNII registration, healthcare personnel still need to be regularly informed to effectively identify drug interactions prior to prescribing. In previous publications, the use of Semantic Web technology was proposed as a solution to this problem. Gao, Xiaoyan et al.

[5] The work in this paper focuses on drug recommendation using a Graph Convolution Network, which mainly uses the mechanism of information propagation and embedding propagation layers to model high-order connectivity and elaborate representation learning. The proposed system includes three key components, namely, an embedding layer, an information propagation layer, and a prediction layer. The work focuses on accuracy rather than evaluation of the recommendation system. Li-Chih Wang et al.

[6] The proposed system in this paper focused on recommending a parameter that is effective when using a curing parameter recommendation system. The proposed system includes a voting method that is developed by seven machine learning algorithms. These ML models are trained as classifiers primarily to recommend a candidate representative medical data set. The file with the highest frequency is selected as the recommended representative data set. Long short-term memory networks are used for heating curve prediction presets. Susannah et al.

[7] In this research, a deep learning approach is proposed for health-based healthcare datasets. This approach automatically identifies what food should be served to which person based on condition and other parameters such as age, race, body weight, calories, fat, sodium, protein, fiber and cholesterol. The main focus of this study framework is the integration of deep learning and machine learning methods such as regression analysis, naive baye, recurrent neural networks, multi-level perceptrons, gated recurrent units, and long-short-term memory (LSTM). The properties of these IoMT samples were evaluated and further processed before applying machine learning, deep learning and other learning-based methods.

V. Proposed Methodology

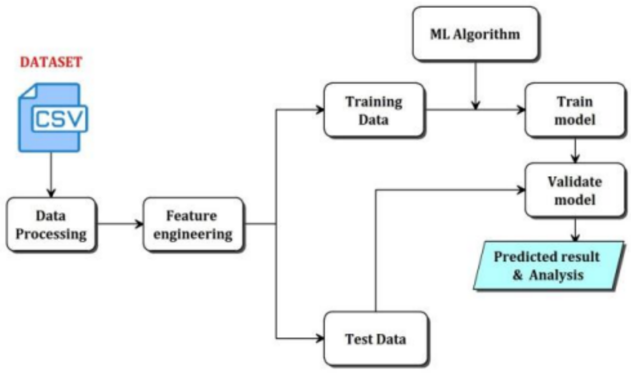

Figure 1: Representation Of System Architecture

The above discern determines the device structure of the proposed machine. The system architecture involves following steps:

A. Data Collection and Preprocessing Machine

Learning needs models and lots of data to work. The process of collecting signals that monitor actual physical situations and converting the obtained results into electronic integer values that can be manipulated by a computer is known as data acquisition. The processing of primary data includes subsequent procedures. In order to compare the details of individual responses, it is necessary to combine a huge amount of raw data obtained from field investigations. The method for transforming dirty data into clean datasets is known as data preprocessing. Real-world information is consistently inaccurate and lacks specific behaviors or patterns. It is also often inconsistent and incomplete.

B. Function Selection and Data Preparation

In order to create attributes for machine learning algorithms, it is necessary to use domain information from the data. The method used right here is referred to as feature engineering. By generating features from input data that aid in a machine learning model, feature extraction can improve the predictive capacity of machine learning algorithms. In machine learning, feature engineering is a core skill that significantly differentiates a successful model from a bad model. The concept of "feature engineering" involves taking raw data and turning it into features that predictive models can use to more accurately represent the underlying problem. The practice of grouping and categorizing data based on specific characteristics is known as data classification. It can be done either by numerical characteristics or by attributes.

C. Model Building and Model Training

The act of training an ML model involves providing a learning algorithm with a training set that can be used as a learning resource. The model artifact created during training is recognized as a "machine learning model". The correct solution, sometimes referred to as a target or target attribute, must be incorporated into the training data. The learning method constructs an ML model that represents these patterns by looking for patterns in the training data that associate characteristics of the input data with the target.

D. Validation of the Model and Evaluation of the Result

The model is used as a new input during the testing phase. There are two distinct samples for training and test data. Designing a machine learning technique with the intention of implementing it effectively. Generalize well to fresh data in both the test set and the training set. Real-time data will be passed for prediction after evaluating the built model. After creating a forecast, the result will be examined for the most important data.

The system architecture involves following steps:

1. Data Collection and Preprocessing Machine

Learning needs models and lots of data to work. The process of collecting signals that monitor actual physical situations and converting the obtained results into electronic integer values that can be manipulated by a computer is known as data acquisition. The processing of primary data includes subsequent procedures. In order to compare the details of individual responses, it is necessary to combine a huge amount of raw data obtained from field investigations. The method for transforming dirty data into clean datasets is known as data preprocessing. Real-world information is consistently inaccurate and lacks specific behaviors or patterns. It is also often inconsistent and incomplete.

2. Function Selection and Data Preparation

In order to create attributes for machine learning algorithms, it is necessary to use domain information from the data. The technique used here is called feature engineering. By generating features from input data that aid in a machine learning model, feature extraction can improve the predictive capacity of machine learning algorithms. In machine learning, feature engineering is a core skill that significantly differentiates a successful model from a bad model. The concept of "feature engineering" involves taking raw data and turning it into features that predictive models can use to more accurately represent the underlying problem. The practice of grouping and categorizing data based on specific characteristics is known as data classification. It can be done either by numerical characteristics or by attributes.

3. Model Building and Model Training

The act of training an ML model involves providing a learning algorithm with a training set that can be used as a learning resource. The model artifact created during training is recognized as a "machine learning model". The correct solution, sometimes referred to as a target or target attribute, must be incorporated into the training data. The learning method constructs an ML model that represents these patterns by looking for patterns in the training data that associate characteristics of the input data with the target.

4. Validation of the Model and Evaluation of the Result

The model is used as a new input during the testing phase. There are two distinct samples for training and test data. Designing a machine learning technique with the intention of implementing it effectively. Generalize well to fresh data in both the test set and the training set. Real-time data will be passed for prediction after evaluating the built model. After creating a forecast, the result will be examined for the most important data.

VI. Conclusion

Whether we are shopping, buying products online or eating out, reviews are gradually becoming part of our daily routine. We use reviews to help us make the best decisions. Multiple machine learning techniques were used to construct the recommender system, which includes Perceptron, Multinomial Naive Bayes, Logistic Regression, Ridge classifier, and Linear SVC implemented on TF-IDF, Bow, and classifiers such as LGBM, Decision Tree, and Random Forest. Our examination of the models using five main metrics: f1score, validity, recall, precision, and AUC score shows that linear SVC using TF-IDF outperforms all other models with 93 percent accuracy. On the other hand, the Word2Vec decision tree algorithm fared the worst, achieving only 78% accuracy.

We integrated the highest expected sentiment values from each strategy LGM on Word2Vec (91%) Perceptron on Bow (91%) Random Forest on manual features (88%) Linear SVC on TF-IDF (93%) and combined them according to the standardized number of useful to create referral system. This gave us a total drug score for each condition. To increase the effectiveness of the recommender system, future work will evaluate different resampling techniques, use alternative n-gram values, and simplify the algorithms.

VII. Future Scope

The unborn work involves comparing different slice shapes using different values of n-grams and optimizing algorithms to improve the performance of the recommender system.

VIII. Result

The drug review sample used in this study was obtained from the UCI ML resource. This data consists of six components: the name of the drug used, the patient's rating, the patient's condition, a valuable number that indicates the number of people who experienced the benefit of the rating, the date the review was written, and a 10-star patient rating that indicates how the patient is doing overall satisfactory. According to the user's star rating, each review in this work has been categorized as positive or negative. Positive reviews are reviews with five or more stars, while negative reviews vary from one to five stars. In Figure 2, we can see the top medical conditions with the largest number of treatment options. One factor to observe in this figure is the fact that there are two green bars that indicate criteria that are of little importance.

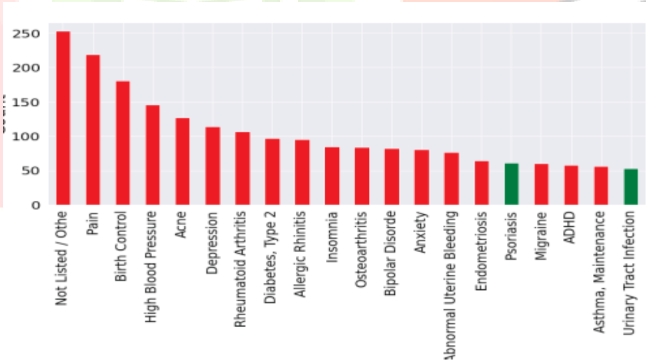

Figure 2: Most Recommended Drugs Per Conditions

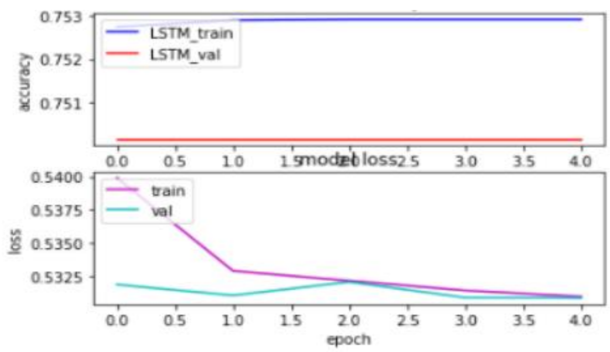

Figure 3: Demonstrates That The Features Used Are Effective Pattern Predictors With High Accuracy And Little Error

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: Displays The Top Four Medications That Our Algorithm Recommends for The Five Top Medical Issues Including Acne, Contraception, High Blood Pressure, Anxiety and Depression.

IX. 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.