

Empowering Medical Practitioners: A Clinical Decision Support Tool Utilizing Vader Sentiment Analysis of Drug Reviews

Hemashree P

Department of Computing
Coimbatore Institute of Technology
India

Dhanush Kannan A

Department of Computing
Coimbatore Institute of Technology
India

Mohindran S R

Department of Computing
Coimbatore Institute of Technology
India

ABSTRACT - With growing concerns over health and medical diagnostic issues, the impact of medication errors by new doctors cannot be ignored, as government reports indicate over 1 million deaths annually attributed to such errors. A significant proportion (42%) of these errors arise from prescriptions written by doctors with limited experience. To address this issue, the integration of data mining and recommendation systems has emerged as a promising avenue to leverage valuable insights from diagnosis history, aiding physicians in prescribing medicines accurately and reducing medication errors. In the era of machine learning, recommender systems have demonstrated their potential to generate more precise, efficient, and reliable clinical predictions, while minimizing costs. By enhancing the performance, integrity, and confidentiality of patient data in the decision-making process, these systems contribute to improved accuracy and timeliness of information provided. In this study, we propose a drug recommendation system that utilizes data mining techniques within a comprehensive framework. This system suggests a curated list of drugs to physicians based on the medical condition and patient reviews, ensuring a more informed prescription decision-making process. By harnessing the power of data mining and recommendation systems, our framework aims to empower physicians with evidence-based recommendations, enabling them to make more accurate and appropriate medication choices. Ultimately, this approach seeks to enhance patient safety, optimize treatment outcomes, and mitigate the risks associated with medication errors, thereby contributing to the advancement of clinical practice.

Keywords-Clinical Decision, Recommendation System, Medical Practitioners, Sentiment Analysis, Drug Reviews.

I. INTRODUCTION

Health-related information has emerged as one of the most critical and intricate topics on the Internet. In light of the current global healthcare landscape, individuals have heightened concerns regarding their health and medical diagnoses. Hospitals and healthcare institutions possess vast amounts of patient data, necessitating the need for effective utilization of this information by medical professionals.

Specifically, there is a growing demand for convenient access to aggregated data from existing databases, enabling medical practitioners to make informed decisions at the point of care. Furthermore, the expanding repertoire of medications, tests, and treatment recommendations presents a daunting challenge for medical staff in determining appropriate courses of action based on patient symptoms, test results, and medical history. Regrettably, prescription errors occur in over 40% of cases, primarily attributable to the limited knowledge available to experts when making crucial decisions. Particularly in the fields of microbial infections, antiviral treatments, and drug efficacy, selecting the most suitable medication assumes paramount importance. With new research conducted daily and a constant influx of medicines and tests, clinicians face mounting difficulty in choosing an optimal treatment plan or prescription for individual patients based on rationale and historical medical data. To bridge this knowledge gap and support clinical decision-making during treatment, the development of a medicine recommendation engine holds immense potential. In this research endeavor, we present a novel recommendation system that leverages patient input regarding specific diseases to suggest appropriate drugs. The recommendations are based on scores calculated through an analysis of patient reviews pertaining to the effectiveness and experiences associated with prescribed drugs.

By harnessing the power of this recommendation system, clinicians can benefit from evidence-based insights that guide their medication selection process. This approach aims to enhance treatment outcomes by aligning prescriptions with patient-specific requirements, enabling healthcare professionals to navigate the complexities of the ever-expanding medical landscape. Ultimately, the proposed system aims to augment clinical practice and improve patient care by providing medical practitioners with a valuable tool for informed decision-making in medication prescription.

II. REVIEW OF LITERATURE

Several studies have been conducted to explore the development and application of drug recommendation systems and clinical decision support tools in the field of healthcare. Usharani Bhimavarapu et al. [1] introduced a drug recommender system utilizing a Stacked Artificial Neural Network (ANN) model, aiming to improve the fairness and safety of treatment for infectious diseases. Their proposed system showcased the potential of recommending safe medicines, particularly during health emergencies.

Lennert Verboven et al. [2] presented a hybrid knowledge and data-driven treatment recommender Clinical Decision Support System (CDSS). Their methodology was applied to develop a CDSS for individualized treatment of drug-resistant tuberculosis, serving as a proof of concept. The novel hybrid approach demonstrated promising results, emphasizing its potential in automating individualized treatment for personalized medicine. Further research is suggested to assess its applicability in diverse medical domains, establish robust statistical approaches to evaluate model performance, and validate its accuracy in real-life clinical settings.

Sarvik Garg [3] developed a medicine recommendation system that employed patient reviews to predict sentiment using various vectorization processes such as Bag-of-Words (BoW), TF-IDF, Word2Vec, and Manual Feature Analysis. The system utilized classification algorithms to recommend the top drug for a given disease. Performance evaluation metrics, including precision, recall, F1 score, accuracy, and AUC score, were employed to assess the predicted sentiments. The findings highlighted the superiority of the LinearSVC classifier with TF-IDF vectorization, achieving an accuracy of 93%.

Baha Ihnaini et al. [4] proposed a smart healthcare recommendation system specifically designed for diabetes disease, leveraging deep machine learning and data fusion perspectives. Their intelligent recommendation system was evaluated using a well-known diabetes dataset and exhibited improved disease diagnosis performance compared to existing deep machine learning methods. The proposed system's enhanced accuracy in disease diagnosis suggests its potential utility in automated diagnostic and recommendation systems for diabetic patients.

Juan G. Diaz Ochoa et al. [5] reported the development of a novel recommender system based on predicted outcomes using continuous-valued logic and multi-criteria decision operators. Their approach involved data synthesis to achieve an error rate of approximately 1% for relevant parameters. The accuracy of the implemented Logic of Natural Numbers (LONN) models was found to be around 75%, albeit lower

than conventional deep learning models. A comparative analysis with these models revealed the LONN models' relatively lower accuracy.

Deloar Hossain et al. [6] proposed a recommendation system that incorporated sentiment analysis of social users' reviews. They utilized Decision Tree, K-Nearest Neighbors (KNN), and Linear SVC machine learning algorithms to predict drug ratings. Emotional analysis was performed using an emotional word dictionary to overcome limitations associated with pre-existing packages developed for movie data. The study demonstrated the significant contribution of sentimental attributes to drug rating prediction and recommendations.

Thi Ngoc Trang Tran et al. [7] presented a comprehensive overview of existing research on Healthcare Recommender Systems (HRS). The article provided insights into different recommendation scenarios and approaches. The authors highlighted the benefits of HRS in terms of health-related improvements while acknowledging the challenges that need to be addressed for their future development.

Venkat Narayana Rao et al. [8] proposed a medicine recommendation system that incorporated patient review data and conducted sentiment analysis using the N-Gram model. To enhance accuracy, a Lightgbm model was employed for medication analysis. The proposed system served as a supportive tool for doctors in disease diagnosis.

Benjamin Stark et al. [9] conducted a literature review focusing on existing solutions for medicine recommender systems. They described and compared these systems based on various features, including diseases, data storage, interface, data collection, data preparation, platform/technology, algorithm, and future work. The review concluded that limited information was available for certain aspects, such as data storage, interface, data collection, data preparation, platforms and technology, and customized algorithms. Future research directions were suggested to address these gaps.

Catherine Henshall et al. [10] developed a web-based computerized Clinical Decision Support Tool (CDST) aimed at providing continuous, personalized information about the efficacy and tolerability of different interventions to patients and clinicians. A focus group study was conducted to assess the feasibility and acceptability of the CDST, revealing its usefulness in supporting clinical decision-making, fostering clinician-patient collaboration, and contributing to the advancement of personalized medicine.

These related studies provide valuable insights into the development and utilization of drug recommendation

systems, clinical decision support tools, and their potential impact on healthcare practice.

III. METHODOLOGY

Medicine recommendation is one of the most important and challenging tasks in the modern world. In the course of time many new diseases are discovered by the doctors. Sometimes, a medicine for one disease can lead to the side-effects which can further lead to the discovery of new diseases. Our goal is to build a recommendation model which will help the Medical Practitioners to prescribe a medicine to the patient even if they are not familiar with the drugs and its effects. The doctor has to open the framework and has to search about the diseases of the patient. It will also help inexperienced doctors and patients to use the right drugs with high accuracy and efficiency. The objective of this paper is to develop a drug recommendation system for medical practitioners to get information on the popular drugs in the market at any point of time. The main functionality of this tool is to recommend drugs available in the market to doctors based on a recommended rating score provided by patient’s reviews and medical conditions. We intend to create this tool to help Medical Practitioners get an idea about different drugs, with respect to the condition, based on reviews and ratings given by the patients. This would help them to shortlist the best drugs for each ailment and prescribe them accordingly.

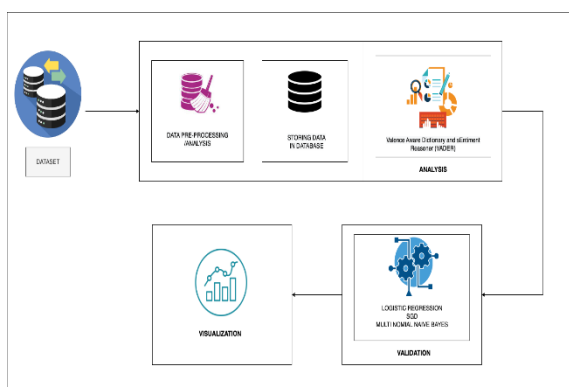


Fig. 1. Architecture Flow of Drug Recommendation System

DATA CLEANING AND VISUALIZATION

Standard data preparation techniques were applied in this study, including checking for null values, duplicating rows, and removing unnecessary values and text from rows. Then all 1200 rows with null values in the condition column were deleted. Unique IDs must be ensured to be unique to remove duplicates.

Feature Extraction

After preprocessing the text, the data must be properly set up to create classifiers for sentiment analysis. Machine learning algorithms cannot process text directly. It needs to be converted to numeric format, especially numeric vectors. Word Bag (Bow), TF-IDF and Word2Vec are well-known simple feature extraction strategies using textual information used in this study. Several feature engineering techniques were also used to manually extract features from the review columns to create another model called Manual Feature alongside Bow, TF-IDF and Word2Vec.

Vader Sentiment Analysis

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a dictionary and rule-based sentiment analysis tool. VADER is used for text sentiment analysis that responds to both the polarity (positive/negative) and intensity (strength) of the sentiment. VADER's sentiment analysis relies on dictionaries that map lexical features to emotional intensity called sentiment scores. The sentiment value of a text can be determined by summing the strength of each word in the text. Vader Sentiment Analysis was performed on the preprocessed reviews and the compound score of the reviews are calculated. The compound score is in the range of [-1, +1]. The compound scores were rescaled into the range of 1 to 10. Finally, the mean normalized score of review text is calculated as the average of original rating and normal vader score.

Classifiers

Distinct machine learning classification algorithms were used to build a classifier to predict the sentiment such as Logistic Regression, Multinomial Naive Bayes and Stochastic gradient descent. These models are performed to cross-evaluate the score found using Vader Analyzer.

- *Logistic Regression:*

Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

- *SGD:*

Stochastic Gradient Descent (SGD) is a simple yet very efficient approach to fitting linear classifiers and regressors under convex loss functions such as (linear) Support Vector Machines and Logistic Regression. SGD has been successfully applied to large-scale and sparse machine learning problems often encountered in text classification and natural language processing.

• *Multinomial Naive Bayes:*

The Multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). It is a Bayesian learning approach popular in Natural Language Processing (NLP). The program guesses the tag of a text, such as an email or a newspaper story, using the Bayes theorem.

HYPERPARAMETER TUNING

• *GridSearchCV:*

Hyper parameters are parameters that are explicitly specified and control the training process. Model performance is highly dependent on hyper parameter values. Hyper parameter tuning is a technique for choosing the optimal set of hyper parameters for a learning algorithm. GridSearchCV is a process of hyper parameter tuning to determine the optimal values for a given model. GridSearchCV tries all combinations of values passed in the dictionary and evaluates the model for each combination using cross-validation.

IV. EXPLORATORY DATA ANALYSIS

DATASET

The dataset provides patient reviews on specific medicines along with affiliated conditions and a 10- star patient rating reflecting overall case satisfaction. The data was attained by crawling online medicinal review sites. The intention was to study sentiment analysis of medicine experience over multiple aspects, i.e., sentiments learned on specific aspects similar as effectiveness and side effects. The following is the table which contains the variables and its description included in the dataset.

VARIABLES	DESCRIPTION
drugName	name of drug
condition	name of condition/disease
review	patient review feedback
ratings	10-star patient rating
date	data of review entry
usefulCount	number of users who found review useful

Table. 1. Dataset description

EXPLORATORY DATA ANALYSIS

From the drug review dataset, the important findings from exploratory data analysis are as the following,

- Levonorgestrel is the drug with highest number of ratings
- There are 12 drugs which has more than 500 count of ratings
- ‘No Side effect’ is the most repetitive term in positive reviews of the given feedback

- ‘took’, ‘pill’, ‘day’, ‘side effect’ are the most repetitive terms found in negative reviews of the feedback

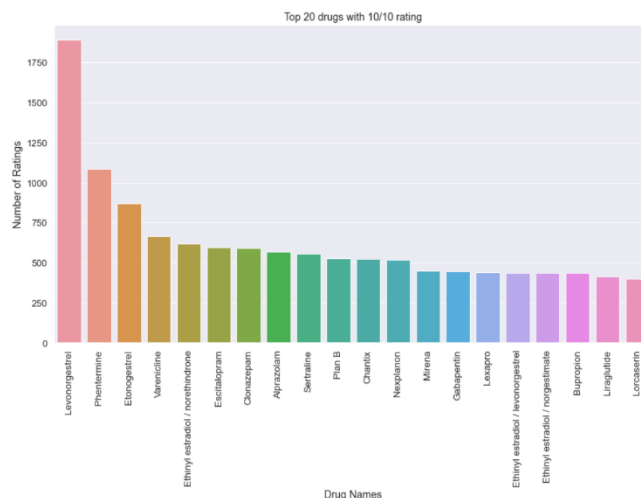


Fig. 2. Top 20 drugs with 10/10 rating

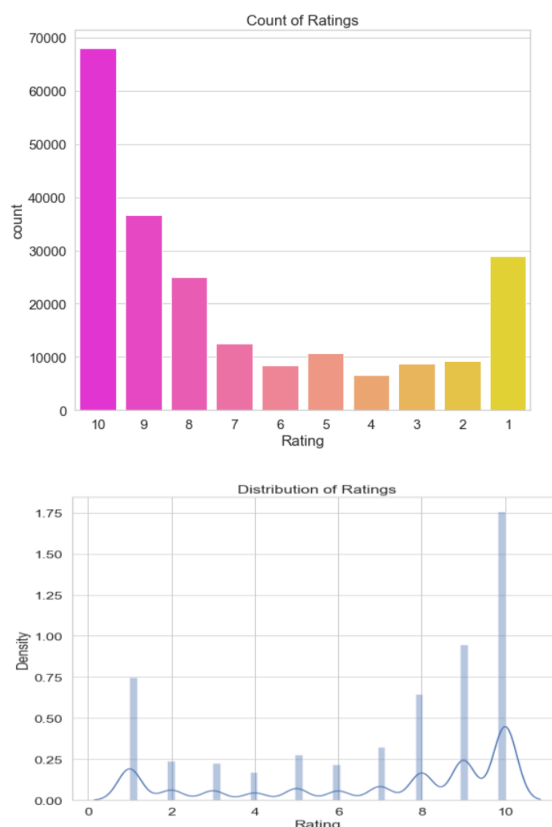


Fig. 3. Distribution and Count of Ratings

Figures 2 & 3 are snapshots of visuals supporting the above insights.

- [8] Venkat, T., Rao, N., Unnisa, A., & Sreni, K. (2020). Medicine Recommendation System based on Patient Reviews. *International journal of Scientific & Technology research*, 9 (2), 3308-3312.
- [9] Stark, B., Knahl, C., Aydin, M., & Elish, K. (2019). A literature review on medicine recommender systems. *International journal of advanced computer science and applications*, 10(8).
- [10] Henshall, C., Marzano, L., Smith, K., Attenburrow, M. J., Puntis, S., Zlodre, J., ... & Cipriani, A. (2017). A web-based clinical decision tool to support treatment decision-making in psychiatry: a pilot focus group study with clinicians, patients and carers. *BMC psychiatry*, 17, 1-10.