IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104233 |

Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning

Menovi Yhoshu, Vamshi Raghuram G, Suryateja Dasari, Bauloi Ato, Prof. Naveen HM

B.E. Student, Dept. of CSE, Bapuji Institute of Engineering and Technology, Davangere, India

B.E. Student, Dept. of CSE, Bapuji Institute of Engineering and Technology, Davangere, India

B.E. Student, Dept. of CSE, Bapuji Institute of Engineering and Technology, Davangere, India

B.E. Student, Dept. of CSE, Bapuji Institute of Engineering and Technology, Davangere, India

Assistant Professor, Dept. of CSE, Bapuji Institute of Engineering and Technology, Davangere, India

ABSTRACT: Since the arrival of the coronavirus, the unavailability of legal clinical resources is at its peak, as is the shortage of healthcare professionals and workers, lack of proper equipment and medicines, etc. The entire medical community is at risk, resulting in many deaths. Due to its unavailability, people began self-medication without proper advice, which worsened their condition. Recently, machine learning has become invaluable in many applications, and the number of innovative articles on automation is increasing. The goal of this article is to present a drug recommendation system that can drastically reduce the mountain of specialists. We have created a drug recommendation system that uses patient ratings to predict sentiment using various vectorization methods such as Bow, TF-IDF, Word2Vec, and manual feature analysis, which can help recommend the best drug for a given disease according to various classification algorithms. The predicted moods were scored for accuracy, recall, f1 score, accuracy, and AUC score.

KEYWORDS: Drugs, Recommendation System, Machine Learning, NLP, Bow, TF-IDF, Word2Vec, Sentiment Analysis

I. INTRODUCTION

With the exponential increase in the number of coronavirus cases, nations are facing a shortage of doctors, particularly in rural areas where the number of specialist doctors is fewer than in urban areas. It takes 6 to 12 years for a doctor to acquire the necessary qualifications. Therefore, the number of doctors cannot be increased rapidly in a short period of time. The telemedicine framework needs to be revived as much as possible during this difficult time. Clinical errors are very common nowadays. Prescribing errors affect more than 200,000 people in China and 100,000 in the United States each year. More than 40% of specialists make prescription errors because the specialists put together the solution based on their very limited knowledge. Choosing a top-notch drug is important for patients who need specialists with in-depth knowledge of microscopic organisms, antimicrobial drugs, and patients. Every day there is a new Study with more drugs and tests available to clinical staff every day. For this reason, it turns out that choosing a doctor is becoming increasingly difficult which treatments or medications to administer to the patient based on the indications and medical history. With the exponential growth of the Internet and web-based industries, item reviews have become an essential and integral factor in purchasing items worldwide. People all over the world are getting used to analyzing reviews and websites before deciding to buy. While most previous research has focused on e-commerce expectations and propositions, the realm of medical care or clinical therapies has rarely been touched upon. The number of people concerned about their well-being and seeking a diagnosis online has increased.

A drug recommendation framework is really important as it can help professionals and patients gain knowledge about drugs for specific conditions. A frame of reference is a common system that presents an item to the user according to his or her utility and need. This framework uses customer surveys to break down their feelings and suggest a recommendation that exactly fits their needs. In the drug recommendation system, a drug in a given condition is suggested based on patient opinion using sentiment analysis and feature engineering. Sentiment analysis is an

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104233 |

advancement of strategies, methods and tools for distinguishing between and extracting emotional data such as opinionsand attitudes from language. On the other hand, inclusive engineering is the process of creating more functionality from existing ones; improves model performance. This review is divided into sections: the introductory part, which provides a brief overview of the needs of this research, the methodological part concerns the methods adopted in this research, the results section evaluates theresults of the applied model using various measures and finally the conclusions section

II. LITERATURE SURVEY

With the development of AI in leaps and bounds, efforts have been made to apply machine learning and deep learning strategies to recommendations. Today, recommendation frameworks are very common in travel industry, e-commerce, restaurants, etc. Unfortunately, in the field of drug suggestions, there are only a limited number of studies using sentiment analysis because drug ratings are much more complicated to analyze because includes clinical phrases such as names of infections, reactions, and synthetic names used in the manufacturing of drugs.

Leilei Sun [1] analyzed treatment records on a large scale to find the best treatment prescription for patients. The idea was to use an efficient semantic clustering algorithm to estimate similarities between treatment records. Similarly, the author provided a framework to assess the appropriateness of the proposed treatment. This facility can prescribe the best treatment regimens for new patients based on demographics and medical complications. Electronic Medical Records (EMR) from patients collected by many clinics for research purposes. The result shows that this structure improves the healing rate.

In this study [2], multilingual mood analysis was performed using Naive Bayes and Recurrent Neural Network (RNN).GoogleTranslate API was used to convert multilingual tweets into English. The results show that RNN at 95.34at Naive Bayes at 77.21%.

The study [3] assumes that the recommended drug should depend on the capacity of the patient.For example, if the patient's immunity is weak, then proven drugs should be prescribed at this time. He proposed a risk classification method to identify the patient's immunity. For example, more than 60 risk factors, high blood pressure, alcohol addiction, etc. are assumed that determine the patient's ability to protect themselves from infections. An online prototype of the system was also developed, which uses a decision support system to help physicians choose first-line drugs.

Xiaohong Jiang et al.[4] studied three different algorithms, a decision tree algorithm, a support vector machine (SVM) and a back-propagation neural network for data processing. The SVM was selected for the drug suggestion module because it performed very well within each of three unique constraints: model accuracy, model competency, model versatility. In addition, an error checking system is offered to ensure the analysis, accuracy and quality of the administration.

Mohammad Mehedi Hassan et al. [5] developed a proposal for drugs in the cloud (FRAME). In terms of side effects for patients, CADRE can suggest medications with high-end prescriptions. This proposed structure was originally based on collaborative filtering techniques where drugs are initially grouped into clusters as shown in the functional description data. However, after accounting for its weaknesses such as costly computations, cold start and lack of information, the model is switched to a cloud-based approach that uses the tensor distribution to improve the quality of drug recommendations.

III. PROPOSED SYSTEM

In the drug recommendation system, a drug in a given condition is suggested based on patients' opinions using sentiment analysis and feature engineering. Sentiment Analysis is an advancement of strategies, methods and tools to distinguish and extract emotional data, like opinions and attitudes, language. On the other hand, inclusive engineering is the process of creating more functionality from existing ones and also improves model performance. Due to problems in the existing system such as the system did not implement accurate sentiment analysis for large data sets and the existing system is less efficient due to not as efficient data classification techniques being implemented. The proposed system is more efficient as it introduces the proposed algorithm used in natural language processing, which is responsible for counting the number of repetitions of all tokens in a journal or document. The proposed system has accurate sentiment analysis forecasting techniques to clean and visualize data.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104233 |

IV. METHODOLOGY

The dataset used in this study is the Drug Review Dataset (Drugs.com) from the UCI ML repository [6]. This record contains six attributes, the name of the drug used (text), the patient's opinion (text), the patient's condition (text), a useful number (number) indicating the number of people from whom the assessment was taken is helpful, the date (date) of the review entry and a 10-star (numeric) patient rating for overall patient satisfaction. Contains a total of 215063 instances. Fig.1 presents the proposed model for building a drug recommendation system. It includes four stages, namely data preparation, classification, evaluation and recommendation.

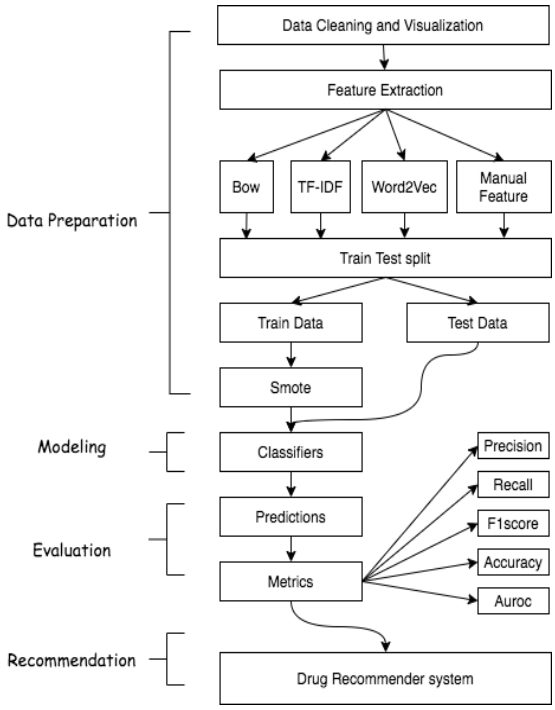

Fig. 1: Proposed Model Flowchart

A. *Data Cleaning and Visualization:*

Standard data preparation techniques were used in this study, such as checking for spaces, duplicating rows, unnecessary values and removing row text. We ensure that the unique identifier is unique to eliminate duplicates. Fig. 2 shows a display of the number of 10-star values ratingsystem. The vast majority choose four grades; 10, 9, 1, 8 and 10 are more than doublesame number. This shows that the positive level is greater thannegative and people's reactions are polar.

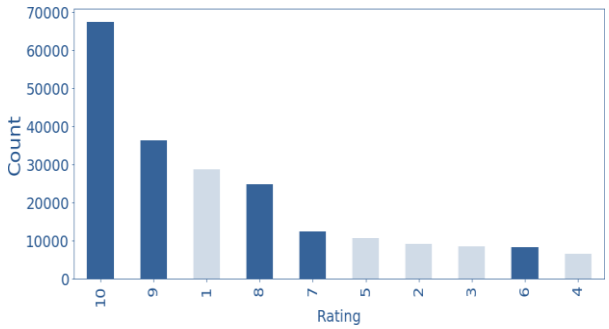

Fig. 2. Bar chart of count of values of rating versus rating

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104233 |

B. *Feature Extraction:*

After text pre-processing, an appropriate summary of the data neededto create classifiers for sentiment analysis. Machine learning algorithms cannot work directly with text; must be converted to a numerical format. A well-known and simple feature extraction strategy with textual information used in this research is Bag of Words (Bow), TF-IDF, Word2Vec.

1.Bag of Words: Bow is an algorithm used in natural language processing that is responsible for counting the number of occurrences of all tokens in a document. A term or token can be called a unigram or any subjective number of words, n grams. In this study, the (1,2) n-gram range was chosen. The Bow model has a significant disadvantage in that it accounts for all terms without considering how multiple terms in the corpus are uniquely consecutive, which in turn creates a large matrix that is computationally expensive to train.

2. TF-IDF: TF-IDF is a popular weighting strategy that weights words instead of counting them. The principle was to give low weight to common terms in the dataset, meaning that TF-IDF assesses relevance, not reproducibility. The term frequency (TF) can be defined as the probability of finding a word in the document. Inverse Document Frequency (IDF) is the inverse of the number of times a given term occurs in the corpus. Detects how the given term is specific to the document.TF-IDF is the multiplication of TF by IDF, indicating the importance and relevance of the word in the document. As with Bow, the n-gram range chosen for TF-IDF in this design is (1,2).

3. Word2Vec: Although TF and TF-IDF are well-known vectorization methods used in various natural language preparation tasks, they ignore the semantic and syntactic similarities between words. For example, in the TF and TF-IDF extraction methods, the words lovable and lovely in the TF and TF-IDF vectorization techniques are referred to as two distinct words, although they are nearly equivalent. Word2Vec is a template for embedding words. Built-in words modelled from huge corpora using various deep learning modelsWord2Vec takes a huge collection of text as input and returns a vector space typically made up of a hundred dimensions. The basic idea was to take the semantic meaning of words and arrange vectors of words in vector space, with the ultimate goal of having words of similar meaning in the dataset close to each other in vector space.

C. *Train Test Split:*

Train test split is a technique for evaluating the performance of a machine learning algorithm. The procedure is to take a dataset and divide it into two subsets. The dataset is split into 75% training and 25% testing.

D. *Classifiers:*

Different classification algorithms based on machine learning were used to create the classifier for sentiment prediction. Logistic regression, naive Bayes polynomial, stochastic gradient descent, linear support vector classifier, perceptron and peak classifier experimented with the Bow model, TF-IDF because they are very sparse matrices and using tree-based classifiers would be time consuming. We selected only machine learning classification algorithms that reduce training time and provide faster predictions.

E. *Metrics:*

Precision (Prec), recall (Rec), f1score (F1), accuracy (Acc.) and AUC score are the metrics used in the recommender system.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104233 |

F. *Recommender System:*

After evaluating the metrics, the top four predicted outcomes were selected and combined to create a combined prediction. The combined scores were then multiplied by a useful normalized count to produce a total drug score for a particular condition. The higher the score, the better is the drug being recommended by the recommender system.The goal is that the more drugs people search for, the moreindividuals read the survey regardless of their opinionpositive or negative, making the useful count high. Therefore, when creating the recommendation system, we normalized the useful count according to the conditions.

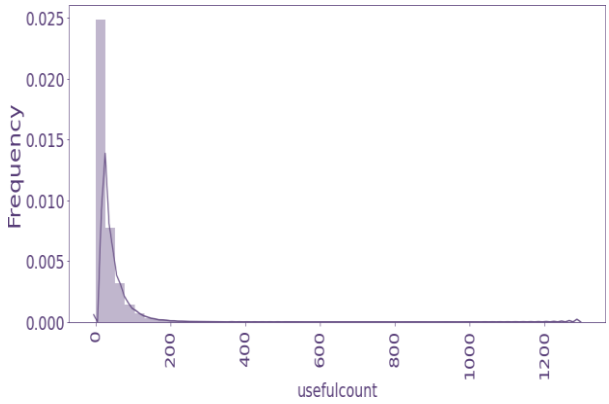

Fig. 3: Useful Count Distribution

V. CONCLUSION

Reviews have become an integral part of our daily lives, whether we go shopping, buy something online or go to a restaurant, we check the reviews first to make the right decisions. This project studied sentiment analysis in drug proving to build a recommender system with different types of machine learning classifiers. Drug recommendation frameworks are very important as they can help professionals and patients acquire knowledge about drugs for specific conditions.

REFERENCES

1. S. Leilei Sun, Chuanren Liu, Chonghui Guo, Hui Xiong, and Yanming Xie. 2016. Data-driven Automatic Treatment Regimen Development and Recommendation. Association for Computing Machinery, New York, NY, USA, 1865–1874. DOI:https://doi.org/10.1145/2939672.2939866
2. V. Goel, A. K. Gupta and N. Kumar, "Sentiment Analysis of MultilingualTwitter Data using Natural Language Processing," 20188th International Conference on Communication Systems and NetworkTechnologies (CSNT), Bhopal, India, 2018, pp. 208-212, doi:10.1109/CSNT.2018.8820254.
3. Shimada K, Takada H, Mitsuyama S, et al. Drug recommendation system for patients with infectious diseases. AMIA Annu Symp Proc.2005;2005:1112.
4. Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 1383-1388, doi: 10.1109/ICIEA.2016.7603801.
5. Zhang, Yin & Zhang, Dafang & Hassan, Mohammad & Alamri, Atif & Peng, Limei. (2014). CADRE: Cloud-Assisted Drug Recommendation Service for Online Pharmacies. Mobile Networks and Applications. 20. 348-355. 10.1007/s11036-014-0537-4.
6. DrugReview Dataset, https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529#
7. Powers, David & Ailab,. (2011). Evaluation: From precision, recall andF-measure to ROC, informedness, markedness & correlation. J. Mach.Learn. Technol. 2. 2229-3981. 10.9735/2229-3981

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details