Sentiment Analysis Of Drug Reviews Using Machine Learning Techniques Based On Drug Recommendation

[1]Avudhurthi Shylaja , [2]B . Sharan ,

[12]Vaageswari College Of Engineering , Telangana , India

[1]shylajaavudhurthi@gmail.com [2]sharanbediga@gmail.com

ABSTRACT

Since the emergence of the corona virus, there has been a dramatic increase in the difficulty with which authorized clinical resources, such as doctors, nurses, diagnostic tools, and medications, may be obtained. Many people in the medical community perish because of the widespread sorrow. As a result of the shortage, people started medicating themselves without first consulting a professional, worsening the health crisis. Machine learning has proven useful in many areas, and new research and development in the field of automation has recently increased in pace and scope. The goal of this research is to introduce a drug recommender system that can significantly lessen specialists' workload. In this study, we developed a medicine recommendation system that predicts sentiment based on patient reviews by employing a number of vectorization processes, including Bow, TF-IDF, Word2Vec, and Manual Feature Analysis, and thus aids in the selection of the best medication for a specific illness as determined with the help of a number of various classification methods. Precision, recall, f1score, accuracy, and area under the curve (AUC) were used to rate the anticipated emotions. The findings demonstrate that the classifier Linear SVC with TF-IDF vectorization achieves the highest accuracy compared to the other models.

I. INTRODUCTION

Especially in rural areas, where there are fewer specialists than in urban areas, countries are experiencing a lack of doctors while the number of corona virus cases increases dramatically. Depending on the medical school you attend, it might take anywhere from six to twelve years to become a fully qualified doctor. Therefore, the number of medical professionals cannot be increased rapidly. In this trying time, the use of a Telemedicine framework should be promoted extensively [1]. Nowadays, medical mistakes occur frequently. Each year, medication errors harm approximately 200 thousand people in China and 100,000 people in the States. In above 40 percent of cases, doctors make mistakes when writing prescriptions because they tailor the treatment to the patient based on their own limited understanding. Patients in need of specialists who have extensive knowledge of microorganisms, antibacterial drugs, and patients are in a position to make informed decisions about which medication to use [6]. Every day, more and more research becomes available, and with it, additional treatments and tests that can be used by clinical professionals. Accordingly, it becomes increasingly difficult for clinicians to decide based on indications, previous clinical history, and other factors, which treatment or medications to recommend for a patient. The proliferation of the internet and e-commerce websites has made product reviews a vital part of the buying process everywhere. People everywhere have gotten in the habit of doing some preliminary research in the form of online reviews and shopper comparison sites before making any major purchases. Most previous research has focused about plans and rating expectations in the e-commerce sector, but the realm of medical care or clinical remedies has been rarely attended to. The number of people seeking out a diagnosis for themselves or a loved one online has increased. It was shown in a 2013 poll [5] by the Pew American Research Center that around 60% of adults looked for fitness-related topics online, and roughly 35% of users sought for health disorders. There is a critical need for a medication recommender framework that can aid doctors and patients in expanding their understanding of the effects of medications on individual diseases. A recommender framework is a commonplace application that makes product suggestions based on the user's stated preferences and needs. These models utilize customer surveys to categorize responses and make tailored recommendations. The medicine recommender system makes use of feature engineering and sentiment analysis determines under what circumstances a given medication should be prescribed. The term "sentiment analysis" [7] refers to a set of techniques for identifying and extracting linguistic expressions of emotion, such as opinion and attitude. Alternatively, Featuring engineering enhances model efficiency by creating new features from preexisting ones. There are five parts to this examination: There will be a place label "Introduction" where you can provide a brief overview of why this study Previous studies in this field are briefly discussed in the "Related Works" section, and the research techniques used herein are described in the "Methodology" section. The framework's constraints are presented in the Discussion part, and the results of the applied models are evaluated using the Evaluation section.

A Statement of the Problem:

Even with the advent of high-powered computers, medical professionals have continued to have a need for technologies such as surgical representation processes and x-ray photography. Medical records, the environment, blood pressure, and other variables all play a role in this strategy, therefore the doctor's knowledge and experience are still essential. No model has successfully assessed the enormous number of factors that are considered as whole variables necessary to understand the complete functioning process itself. Using a medical decision support system is the only way to overcome this limitation. This system can help doctors make the right choice. The term "medical decision support system" can be used to describe either the effort put forth to ascertain the possibility of a disease or ailment, or the conclusion reached after doing so.

Why It's Important:

Depending on the circumstances, such as with uncommon diseases, making a medical decision can be a highly specialized and challenging task. Stress, exhaustion, and a lack of sleep are all possible contributors, as can a lack of resources and a lack of knowledge on the part of medical professionals. It's possible for a standard algorithm to examine all of the determinants, such as the patient's current health status, previous medical history, family medical history, and other aspects relevant to the patient's medical file. If there are numerous potential explanations for anything, differential diagnosis can be used to zero in on the most likely one. This strategy calls for an elimination procedure or data collection that reduces the likelihood of potential situations to zero.

II. RELATED WORKS

2.1. An Overview of Medication Errors for Healthcare Professionals. This article was originally published in the Mayo Clinic Proceedings.

Medication errors contribute significantly to patient morbidity and mortality, however the concept is often misunderstood. This article is intended to serve as a review for practicing physicians, covering the following topics related to medication errors: (1) definitions and terminology, (2) prevalence, (3) exposure factors, (4) prevention methods, and (5) discovery and legitimate ramifications. Any mistake made while taking a drug is considered an error. An estimated 1 in 131 deaths among outpatients and 1 in 854 deaths among hospitalized patients are the result of drug errors, according to the Institute of Medicine. Medication-related factors (such as similar-sounding names or a low therapeutic index), patient-related factors (such as impaired cognition, poly pharmacy , and poor renal or hepatic function), and healthcare provider-related factors (such as the use of abbreviations in prescriptions and other communications or cognitive biases) can all contribute to medication errors. Doctors who make mistakes with their patients' prescriptions risk losing their patients' faith, potentially facing lawsuits or even criminal charges and disciplinary action from their licensing body. Medication mistake prevention strategies (such improved drug label and medication reconciliation) have had mixed results. When an error is detected, patients demand prompt, personal disclosure, accompanied by an apology and an explanation of what is being done to prevent similar mistakes in the future. Health care providers' knowledge of medication errors is likely to improve as more research is conducted on this issue.

2.2. What causes medication mistakes in hospitals, and how to stop them Clinical Medicine: A Journal of Practice.

Many preventable pharmaceutical mishaps in hospitals occur as a result of incompetent prescribing, and many of those involved are recently-minted physicians. Prescribing is a difficult skill that requires knowledge of medications, clinical pharmacology concepts, the capacity to weigh risks and benefits, and preferably, prior experience. It should come as no surprise that mistakes are made. Second, being a prescriber is probably harder now than it has ever been. Over the past two decades, there have been significant shifts in medical education in response to criticisms of a heavy workload and a dearth of social science content. As a result of these shifts, clinical pharmacology and practical prescription are no longer staples of undergraduate education and assessment in the United Kingdom. There is rising worry, not least among medical students, that they are not adequately prepared for the demands of prescribing by the time they graduate medical school. In other countries, people are voicing similar worries. Although direct proof is lacking, there is some evidence linking these alterations to actual instances of pharmaceutical mistakes in clinical use. 3 Knowledge and training are key factors in error causation, and targeted education enhances prescription performance, as shown by a systems analysis of errors. There is already enough data, in our opinion, to warrant a thorough examination of the current state of education concerning the preparation of future prescribers and the development of their abilities throughout postgraduate study. We offer a set of concepts upon which educational programmers may be built.

Using NLP, we conduct a sentiment analysis on tweets written in multiple languages.

Since the appearance of the corona virus, there has been a severe dearth of legitimate clinical resources such as specialists, healthcare workers, adequate equipment, and medicines. Numerous people in the medical community have died due to the widespread distress. Because of the lack of access to necessary medication, many began self-medicating without first consulting a doctor, further exacerbating the underlying health problem. Machine learning has proven useful in many areas, and new research and development in the field of automation has recently increased in pace and scope. The goal of this research is to introduce a drug recommender system that can significantly lessen specialists' workload. With the use of vector methods such as Bow, TFIDF, Word2Vec, and Manual Feature Analysis, we are able to anticipate the sentiment of patient evaluations, which in turn allows us to prescribe the best drug for a certain ailment using a variety of classification algorithms. Precision, recall, f1score, accuracy, and area under the curve (AUC) were used to rate the anticipated emotions. Linear SVC, a classifier that employs TFIDF vector, performed better than all other models tested (by a significant margin).

Over sampling with a Synthetic Minority:

Learning algorithms are hindered by imbalanced data sets since it is more difficult to acquire knowledge about classes with fewer data samples. In order to remedy this, synthetic oversampling methods generate artificial minority samples to incorporate into the main data set. Many of these methods, however, have the potential to generate erroneous synthetic minority samples that are located within majority areas. This work introduces a new Cluster Based Synthetic Oversampling (CBSO) approach to address this issue. CBSO takes its cue from preexisting synthetic oversampling methods and adds unsupervised clustering to the mix in order to generate its own set of synthetic data. CBSO assures that any synthetic samples generated using this technology are always located within minority regions. Simulation studies on several real-world datasets demonstrate the efficacy of CBSO, with gains in overall accuracy, F-measure, and G-mean, among other assessment metrics.

III. DATASET DESCRIPTION

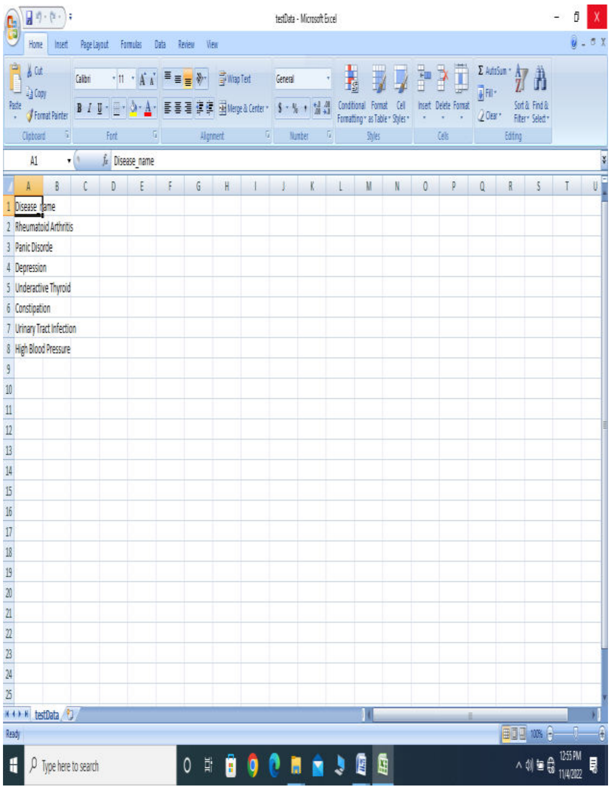

Fig 1: Dataset

IV. METHODOLOGY

In order to create a classifier that could anticipate the sentiment, various machine-learning classification techniques were applied. The four best findings were chosen and pooled to give the combined prediction after the metrics were evaluated. After that, a normalized useful count was multiplied by the combined data to produce an overall score for the medicine for a given condition. The better the medication , the higher the score. The rationale behind this is that as more people look up drugs online, more people read survey responses whether beneficial or negative raising the usefulness factor.

A.IMPLEMENTATION

Modules Used in Project:

1) Upload Drug Review Dataset: We will upload the dataset to the program using this module.

2) Read & Preprocess Dataset: Using this module, we'll read the dataset's reviews, names of drugs, and ratings in order to create a features array.

3) TF-IDF Features Extraction: The features array will be fed into the TF-IDF algorithm, which will calculate the average frequency of each word before substituting the frequency value for the word and forming a vector. If a word is absent from the sentence, 0 will be entered. All reviews will be taken into account as input features for the machine learning algorithm, and the drug name and rating will be taken into account as a class label.

4) Train Machine Learning Algorithms: Using this module, we will feed all machine learning algorithms TF-IDF features before training a model.

5) Comparison Graph: using this module we will plot accuracy graph of each algorithm.

6) Recommend Drug from Test Data: using this module we will upload disease name test data and then ML will predict drug name and ratings.

V. EXPERIMENT, RESULTS, AND ANALYSIS

To launch the project, double-click the "run.bat" file to bring up the screen below.

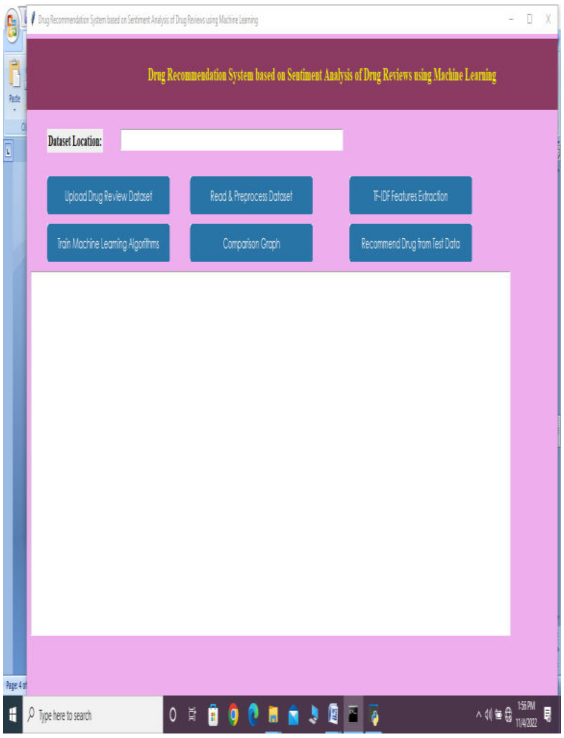

Fig 2: To upload a dataset to the application and view the screen below, click the "Upload Drug Review Dataset" button in the screen above.

Fig 3 : The dataset is loaded in the graph above, which also shows the ratings on the x-axis and the total number of records that received each rating on the y-axis. and y-axis represents total number. Close the previous graph now, then select "Read & Preprocess Dataset" to read all dataset values, then select "Preprocess" to

eliminate stop words and other special symbols, and finally form a features array.

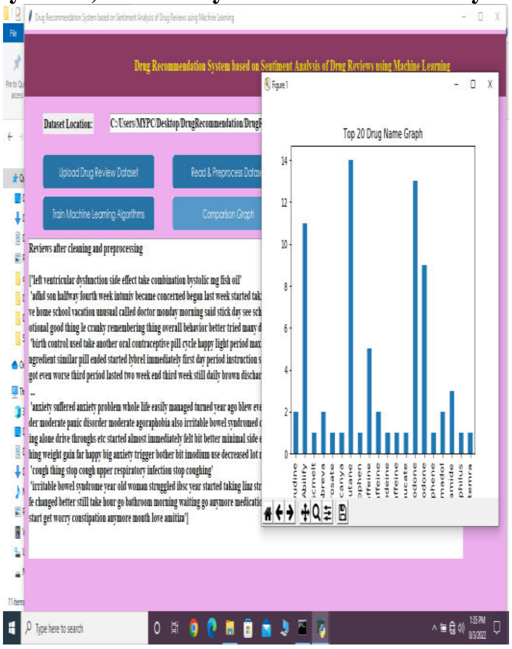

Fig 4 : In above screen we can see from all reviews stop words and special symbols are removed and in graph I am displaying TOP 20 medicines exist in dataset. In above graph x-axis represents drug name and y-axis represents its count. Now close above graph and then click on 'TF-IDF Features Extraction' button to convert all reviews in to average frequency vector.

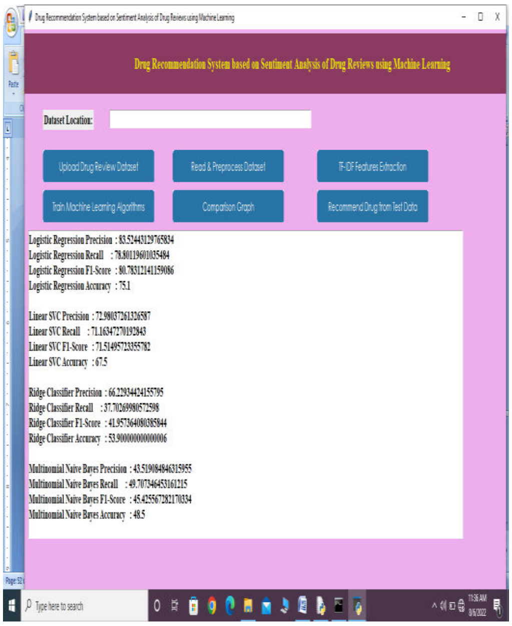

Fig 5 : Now click on 'Train Machine Learning Algorithm' button to train all algorithm and get below output.

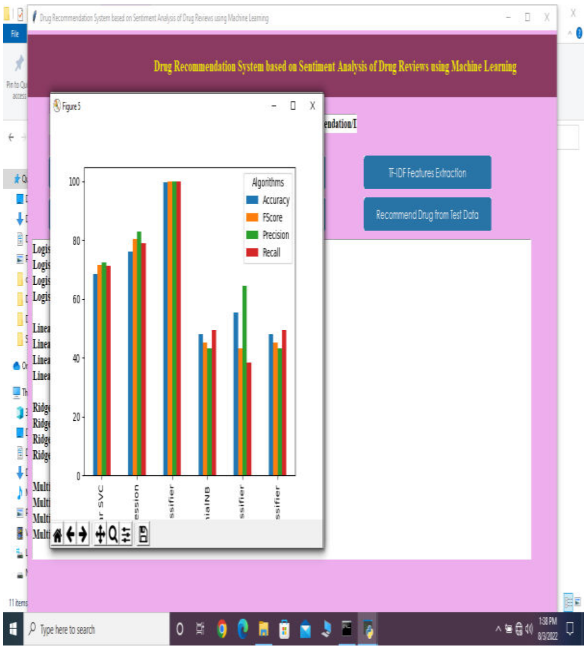

Fig 6 : In above graph x-axis represents algorithm name and y-axis represents accuracy, precision recall and FSCORE where each different colour bar will represents one metric and in above graph we can see MLP got high performance. Now close above graph and then click on 'Recommend Drug from Test Data' button to upload test data and to get predicted result as drug name and ratings.

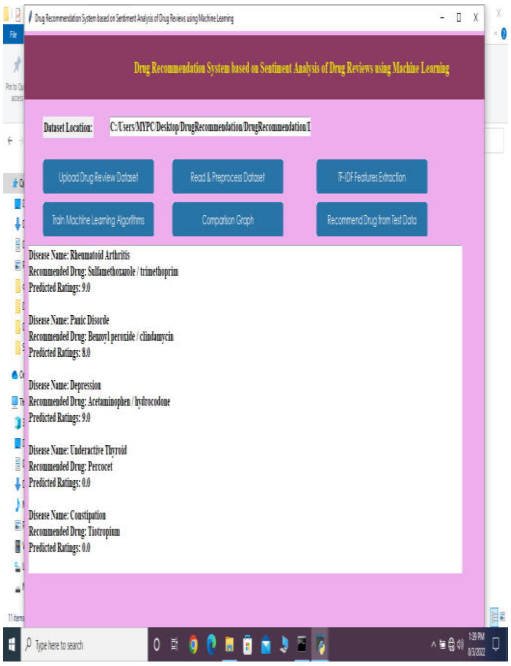

Fig 7 : Each illness name application's predicted recommended medicine name and ratings are shown in the above screen.

VI. CONCLUSION AND FUTURE WORK

Reviews are becoming an essential part of our daily lives; before going shopping, making an online purchase, or visiting a restaurant, we first read reviews to help us make the best choices. A recommender system was developed using a variety of machine learning classifiers, including Logistic Regression, Perception, Multinomial Naive Bayes , Ridge classifier, Stochastic Gradient Descent, Linear SVC, applied on Bow, TF-IDF, and classifiers like Decision Tree, Random Forest, Lgbm, and Cat boost, applied on Word2Vec and Manual Features Method. This research was motivated by this. Using five distinct metrics—precision, recall, f1score, accuracy, and AUC score—we assessed them and found that the Linear SVC on TF-IDF performs 93% better than the other models.

References:

[1]Telemedicine, https://www.mohfsw.gov.in/pdf/Telemedicine.pdf.

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

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

[4]Drug Review Dataset, https://archive.ics.uci.edu/ml/datasets/Drug%2BReview%2BDataset%2B%2528Drugs.com%2529#.

[5] Fox, Susannah, and Maeve Duggan. "Health online 2013. 2013." URL: http://pewinternet.org/Reports/2013/Health-online.aspx.

[6] 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. doi: 10.1086/313954. Epub 2000 Sep 7. PMID: 10987697; PMCID: PMC7109923.

[7] Fox, Susannah & Duggan, Maeve. (2012). Health Online 2013.Pew Research Internet Project Report.

[8] 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,2016, pp. 1471-1476, doi: 10.1109/SCOPES.2016.7955684.

[9] Doulaverakis, C., Nikolaidis,G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Semant 3, 14 (2012).https://doi.org/10.1186/2041-1480-3-14.

[10] Leilei Sun, Chuanren Liu, ChonghuiGuo, Hui Xiong, and YanmingXie. 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.DOI:https://doi.org/10.1145/2939672.2939866.