Sentiment-Driven Medication Guidance: A Machine Learning Approach for Drug Recommendations

Mr. R. Bhargava[1] ,K. Divya Harshika[2],K. Vinay Kumar[3], S. Nithin Ashok[4] , Sai Vamsi Punith[5]

#1Assistant Professor in Department of CSE,in PBR Visvodaya Institute of Technology and Science, Kavali.

#2#3#4#5 B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology and Science, Kavali.

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 drug for a given disease as determined by a number of different classification algorithms. 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.

INDEX TERMS Drugs, Sentiment analysis , Recommender system, Natural language processing, Health care .

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 around 200 thousand people in China and 100,000 people in the United States. In over 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 which treatment or medications to prescribe a patient based on indications, past clinical history, and other factors. 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 on rating expectation and proposals in the E-Commerce industry, 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 health-related topics online, and roughly 35% of users sought for diagnosing 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. Using sentiment analysis and feature engineering, the drug recommender system 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) risk factors, (4) prevention methods, and (5) disclosure and legal 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.

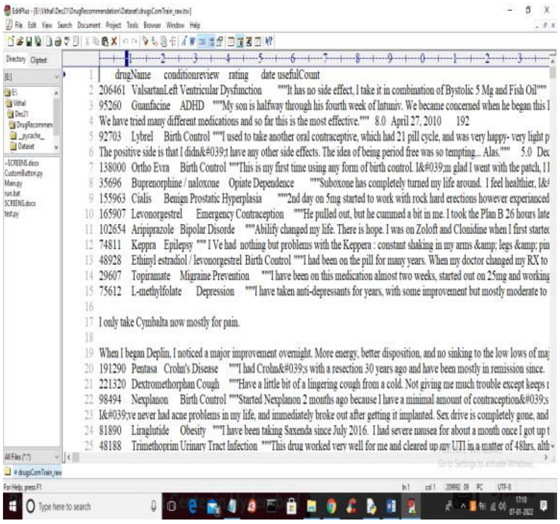
Using NLP, we conduct a sentiment analysis on tweets written in multiple languages.

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



```
drugName	conditionreview	rating	date usefulCount
206461	ValsartanLeft Ventricular Dysfunction	"""It has no side effect, I take it in combination of Bystolic 5 Mg and Fish Oil"""
95260	Guanfacine	ADHD	"""My son is halfway through his fourth week of Intuniv. We became concerned when he began this l
We have tried many different medications and so far this is the most effective."""	8.0	April 27, 2010	192
92703	Lybrel	Birth Control """I used to take another oral contraceptive, which had 21 pill cycle, and was very happy- very light p
The positive side is that I didn&#039;t have any other side effects. The idea of being period free was so tempting... Alas."""	5.0	Dec
138000	Ortho Evra	Birth Control """This is my first time using any form of birth control. I&#039;m glad I went with the patch, I l
35696	Buprenorphine / naloxone	Opiate Dependence	"""Suboxone has completely turned my life around. I feel healthier, I&#
155963	Cialis	Benign Prostatic Hyperplasia	"""2nd day on 5mg started to work with rock hard erections however experianced
165907	Levonorgestrel	Emergency Contraception	"""He pulled out, but he cummed a bit in me. I took the Plan B 26 hours late
102654	Aripiprazole	Bipolar Disorde	"""Abilify changed my life. There is hope. I was on Zoloft and Clonidine when I first startec
74811	Keppra	Epilepsy """ I Ve had nothing but problems with the Keppera : constant shaking in my arms &amp; legs &amp; pin
48928	Ethinyl estradiol / levonorgestrel Birth Control """I had been on the pill for many years. When my doctor changed my RX to
29607	Topiramate	Migraine Prevention	"""I have been on this medication almost two weeks, started out on 25mg and working
75612	L-methylfolate	Depression	"""I have taken anti-depressants for years, with some improvement but mostly moderate to

I only take Cymbalta now mostly for pain.

When I began Deplin, I noticed a major improvement overnight. More energy, better disposition, and no sinking to the low lows of ma
191290	Pentasa	Crohn's Disease	"""I had Crohn&#039;s with a resection 30 years ago and have been mostly in remission since.
221320	Dextromethorphan Cough	"""Have a little bit of a lingering cough from a cold. Not giving me much trouble except keeps r
98494	Nexplanon	Birth Control """Started Nexplanon 2 months ago because I have a minimal amount of contraception&#039;s
I&#039;ve never had acne problems in my life, and immediately broke out after getting it implanted. Sex drive is completely gone, and
81890	Liraglutide	Obesity	"""I have been taking Saxenda since July 2016. I had severe nausea for about a month once I got up t
48188	Trimethoprim Urinary Tract Infection	"""This drug worked very well for me and cleared up my UTI in a matter of 48hrs. alth
```

In above screen first row represents dataset column names such as drug name, condition, review and rating and remaining rows contains dataset values and we will used above REVIEWS and RATINGS to trained machine learning models.

Fig 1: Dataset

IV. METHODOLOGY

Distinct machine-learning classification algorithms were used to build a classifier to predict the sentiment. After assessing the metrics, all four best-predicted results were picked and joined together to produce the combined prediction. The merged results were then multiplied with normalized useful count to generate an overall score of drug of a particular condition. The higher the score, the better is the drug. The purpose behind is that the more medications individuals search for, the more individuals read the survey regardless of their review is positive or negative, which makes the useful count high.

A.IMPLEMENTATION

Modules Used in Project:

1) Upload Drug Review Dataset: using this module we will upload dataset to application.

2) Read & Preprocess Dataset: using this module we will read all reviews, drug name and ratings from dataset and form a features array.

3) TF-IDF Features Extraction: features array will be input to TF-IDF algorithm which will find average frequency of each word and then replace that word with frequency value and form a vector. If word not appear in sentence then 0 will be put. All reviews will be consider as input features to machine learning algorithm and RATINGS and Drug Name will be consider as class label.

4) Train Machine Learning Algorithms: using this module we will input TF-IDF features to all machine learning algorithms and then trained a model and this model will be applied on test data to calculate prediction accuracy of the algorithm.

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

RUN ML ALGORITHMS:

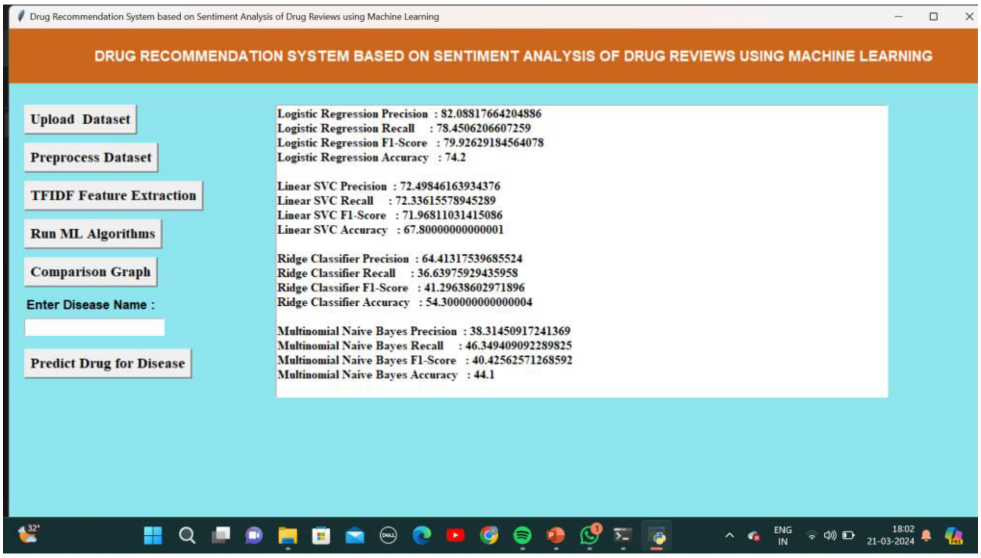


Fig 2: Running ML Algorithms

Description: After clicking "Run ML Algorithms" button the evaluation metrics of ML algorithms are displayed.

COMPARISON GRAPH:

Fig 3:Comparison Graph

Description: After clicking "Comparison Graph" button the graph between algorithms is displayed.

DRUG RECOMMENDATION:

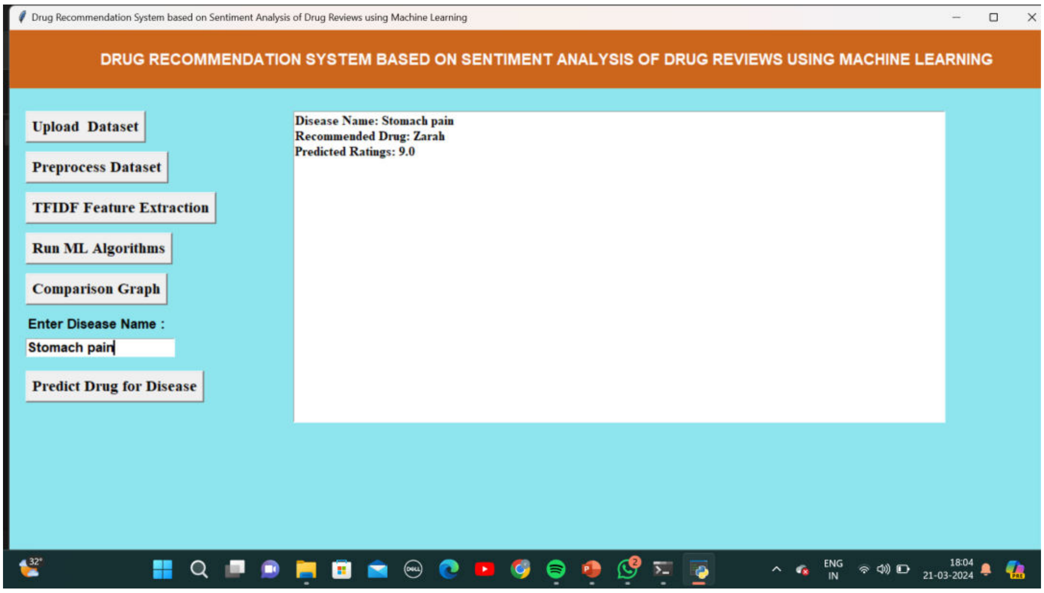


Fig 4: Drug Recommendation

Description: After entering disease name by clicking on "Predict Drug for Disease" button the drug name that is recommended based on reviews will be displayed.

VI. CONCLUSION AND FUTURE WORK

Reviews are becoming an integral part of our daily lives; whether go for shopping, purchase something online or go to some restaurant, we first check the reviews to make the right decisions. Motivated by this, in this research sentiment analysis of drug reviews was studied to build a recommender system using different types of machine learning classifiers, such as

Logistic Regression, Perception, Multinomial Naive Bayes , Ridge classifier, Stochastic gradient descent, Linear SVC, applied on Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Cat boost were applied on Word2Vec and Manual features method. We evaluated them using five different metrics, precision, recall, f1score, accuracy, and AUC score, which reveal that the Linear SVC on TF-IDF outperforms all other models with 93% accuracy. On the other hand, the Decision tree classifier on Word2Vec showed the worst performance by achieving only 78% accuracy. We added best-predicted emotion values from each method, Perception on Bow (91%), Linear SVC on TF-IDF (93%), LGBM on Word2Vec (91%), Random Forest on manual features (88%), and multiply them by the normalized useful Count to get the overall score of the drug by condition to build a recommender system. Future work involves comparison of different oversampling techniques, using different values of n-grams, and optimization of algorithms to improve the performance of the recommender system.

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/10.1145/293967 2.2939866.

Author’s Profiles

Mr.R.Bhargava working as Assistant Professor in Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali.

K. Divya Harshika B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology and Science, Kavali.

K. Vinay Kumar B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology and Science, Kavali.

S. Nitin Ashok B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology and Science, Kavali.

Sai Vamsi Punith B.Tech with Specialization of Computer Science and Engineering in PBR Visvodaya Institute of Technology and Science, Kavali.