e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707151 |

Drug Recommendation System Based on Sentiment Analysis

Anjana H N*[1]. Shwetha Shri K*[2]

Student, Master of Computer Applications. East West Institute of Technology, Bengaluru, Karnataka, India. *[1]

Associate Professor, Master of Computer Applications. East West Institute of Technology, Bengaluru, Karnataka, India. *[2]

ABSTRACT: Since coronavirus has shown up, inaccessibility of legitimate clinical resources is at its peak, like the shortage of specialists and healthcare workers, lack of proper equipment and medicines etc. The entire medical fraternity is in distress, which results in numerous individual's demise. Due to unavailability, individuals started taking medication independently without appropriate consultation, making the health condition worse than usual. As of late, machine learning has been valuable in numerous applications, and there is an increase in innovative work for automation. This paper intends to present a drug recommender system that can drastically reduce specialist's heap. In this research, we build a medicine recommendation system that uses patient reviews to predict the sentiment using various vectorization processes like Bow, TF-IDF, Word2Vec, and Manual Feature Analysis, which can help recommend the top drug for a given disease by different classification algorithms. The predicted sentiments were evaluated by precision, recall, f1score, accuracy, and AUC score. The results show that classifier LinearSVC using TF-IDF vectorization outperforms all other models with 93% accuracy

KEYWORDS: Machine learning, , LinearSVC, TF-IDF

I. INTRODUCTION

With the number of coronavirus cases rising at an exponential rate, countries are experiencing a doctor crisis, particularly in rural areas where the number of experts is lower than in urban areas. Obtaining the requisite qualifications for a doctor takes between 6 to 12 years. As a result, the number of doctors cannot be increased rapidly in a short period of time. In this tough moment, a Telemedicine framework should be energised as much as feasible. Clinical errors are all too common these days. Every year, around 200 thousand people in China and 100 thousand in the United States are harmed by medication errors. Over 40% of doctors make mistakes while prescribing because they create the answer based on their limited understanding. For patients who require doctors with broad knowledge of microscopic organisms, antibacterial drugs, and patients, selecting the highest-level medication is critical. Every day, a new study is published, along with more medications and diagnostics that are made available to healthcare professionals. As a result, choosing which treatment or drugs to offer a patient based on indications and past clinical history is becoming increasingly difficult for clinicians.With the rapid growth of the internet and the web-based commercial industry, item reviews have become an essential and vital part of purchasing products all over the world. People all across the world have become accustomed to reading reviews and visiting websites before making a purchase decision. While most previous research focused on rating expectations and suggestions in the E-Commerce area, medical care or clinical therapies have received little attention. There has been an increase in the amount of people concerned about their health and seeking a diagnosis online. According to a Pew American Research Center poll conducted in 2013, around 60% of adults searched online for health-related topics, and approximately 35% of users searched for diagnosing health disorders. A medication recommender framework is critical in order for doctors and patients to gain a better understanding of medications used to treat various health

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707151 |

II. LITRATURE SURVEY

2.1 EXISTING SYSTEM:

On treatment data, Xiaohong Jiang et al. looked at three different algorithms: decision tree algorithm, support vector machine (SVM), and backpropagation neural network. SVM was chosen for the medicine suggestion module because it excelled in all three of the criteria: model exactness, model proficiency, and model adaptability. In addition, a mistake-checking system was proposed to verify the quality of analysis, precision, and administration.

Mohammad Mehedi Hassan and colleagues created a cloud-based medication proposal (CADRE). CADRE can recommend medications with top-N related prescriptions based on the side effects of patients. This suggested approach is based on collaborative filtering procedures, in which drugs are first grouped into clusters based on the functional description data. However, the model is transferred to a cloud-aided strategy employing tensor decomposition for increasing the quality of experience of medication suggestion after considering its shortcomings such as computationally costly, cold start, and information sparsity

2.2 PROPOSED SYSTEM:

A recommender framework is a standard system that suggests an item to a user based on their benefit and requirement. These frameworks use customer surveys to break down their feelings and make recommendations tailored to their specific needs. The drug recommender system uses sentiment analysis and feature engineering to offer drugs based on patient feedback for a specific disease. Sentiment analysis is a set of tactics, methods, and tools for separating and extracting emotional data from language, such as opinions and attitudes.

The Drug Review Dataset (Drugs.com) from the UCI Machine Learning repository was used in this study. The following attributes are included in this dataset: name of drug (text), patient review (text), patient condition (text), useful count (numerical) indicating the number of people who found the review helpful, date (date) of review entry, and a 10-star patient rating (numerical) indicating overall patient satisfaction.

Each review was categorized as favorable or negative in this study based on the user's star rating. Positive ratings range from one to five stars, while negative ratings range from one to five stars. LinearSVC was chosen as the best algorithm because the accuracy we attained was Train Accuracy: 0.903 and Test Accuracy: 0.8369, which was higher than all other systems.

III. SYSTEM REQUIREMENT SPECIFICATION

3.1 ACCESSORS

Modules are used:

Data Collection
Dataset
Data Preparation
Model Selection
Analyze and Prediction
Accuracy on test set
Saving the Trained Model
Database connecting using MySQL

MODULES DESCSRIPTION:

Data Collection:

This is the first step in collecting data for the building of a machine learning model. This is an important stage that will affect how good the model is; the more and better data we have, chevalier, the better our model will perform.

Data can be collected via a variety of methods, including online scraping, manual interventions, and so on.Machine Learning-based Drug Recommendation System based on Sentiment Analysis of Drug Reviews.

Data set Link: https://www.kaggle.com/jessicali9530/kuc-hackathon-winter-2018

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707151 |

Dataset:

The train dataset consists of 161297 and test dataset consists of 53766 There are 3 columns in the dataset, which are described below

Index: unique id
Drug Name: Name of drug used
Condition: Condition of a patient
Review: Review of a patient
Rating: 1 to 10
Date: The day, month, or year
useful Count: review count

Software Requirements

The program needs the following software requirements to be developed:
Python
Django
MySQL
MySQL client
Wamp Server 2.4
Datasets

Operating Systems supported

Windows 10
Windows 11

Technologies and Languages used to Develop

Python

Debugger and Emulator

Any Internet Explorer (Particularly Chrome)

Hardware Requirements

Moniter:15 LED
Processor:i3
RAM: 4 GB
Space on Hard Disk: minimum 512MB

IV. PROBLEM STATEMENT AND OBJECTIVES

4.1 PROBLEM STATEMENT:

The healthcare system has faced unprecedented difficulties ever since the coronavirus pandemic began, such as a lack of specialists and healthcare workers, inadequate medical equipment, and limited access to essential medicines. Numerous deaths have occurred as a result of the situation, which has caused severe distress within the medical community. The detachment of expert clinical counsel has constrained people to self-cure, frequently worsening their ailments. Utilizing technology for healthcare automation presents an opportunity thanks to advances in machine learning. This venture expects to resolve these issues by fostering a medication recommender framework that uses patient surveys to foresee feelings and suggest fitting drugs, subsequently diminishing the weight on medical care experts and working on persistent results.

4.2 OBJECTIVES:

Develop a Drug Recommender System: Create a system that recommends suitable drugs for various diseases based on patient reviews.
Implement Sentiment Analysis: Use machine learning techniques to analyze patient reviews and predict their sentiments regarding different medications.
Utilize Various Vectorization Methods: Apply different vectorization techniques such as Bag of Words (BoW), TF-IDF, Word2Vec, and Manual Feature Analysis to process the text data.Employ
Classification Algorithms: Test multiple classification algorithms to determine the most accurate model for sentiment prediction.
Evaluate Model Performance: Assess the performance of the models using precision, recall, F1 score, accuracy, and AUC score.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707151 |

Achieve High Accuracy: Aim to achieve a high accuracy rate for the sentiment prediction, with the goal of exceeding 90%.

V. SYSTEM DESIGN

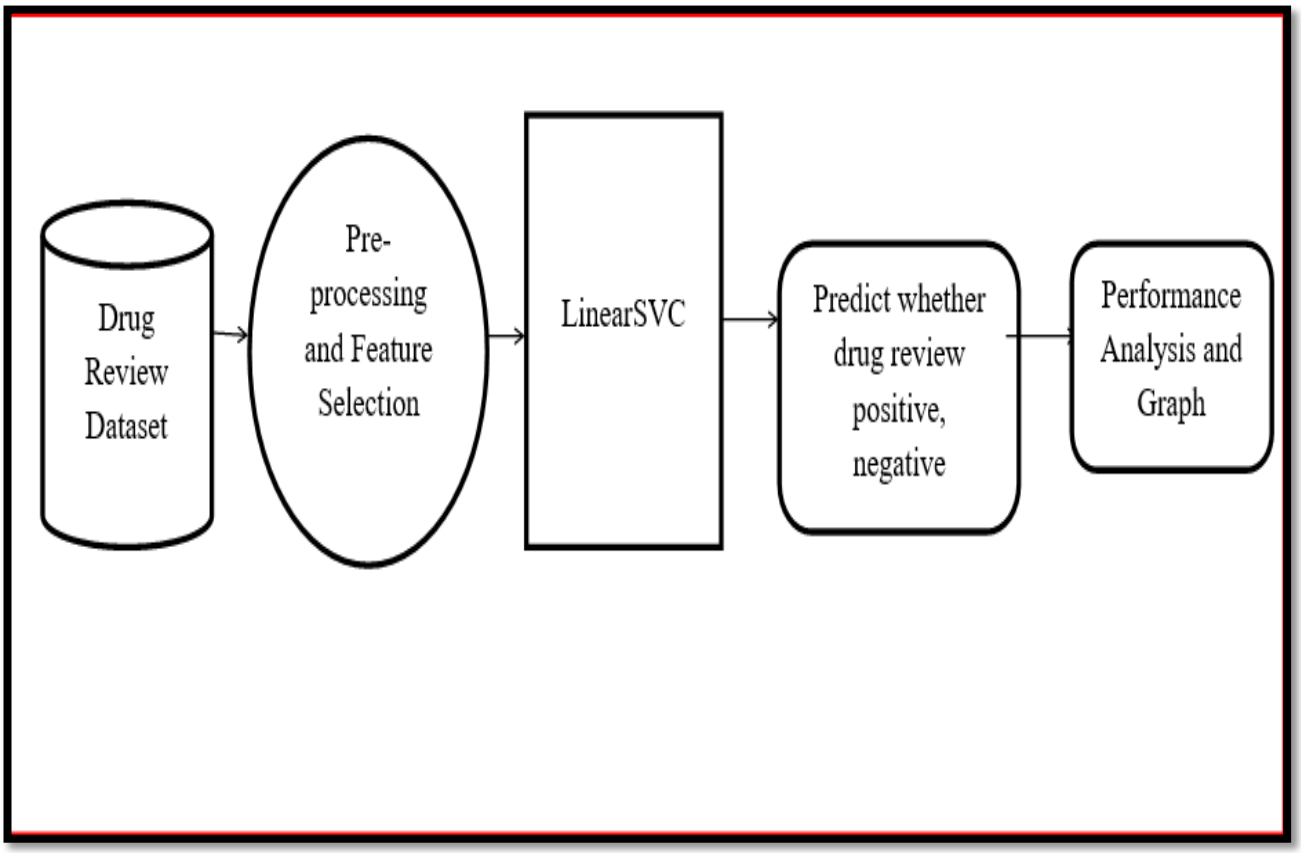
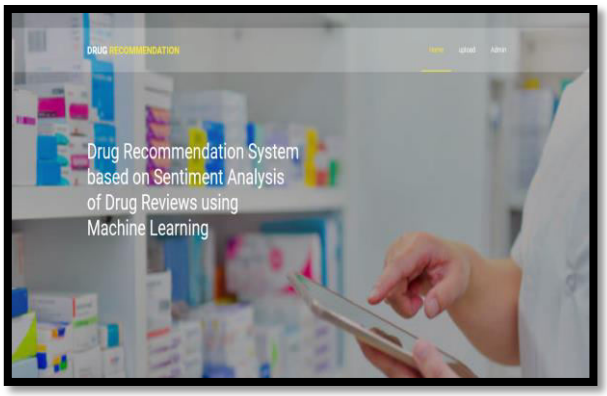


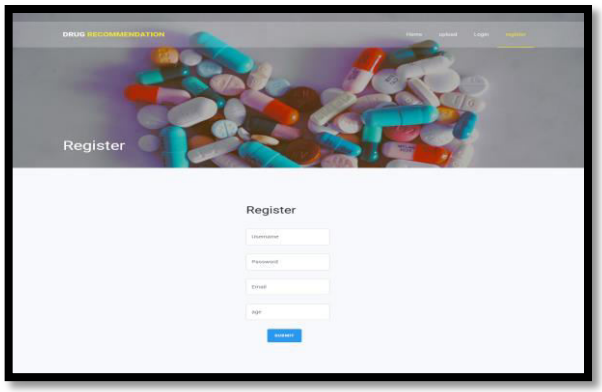
Fig: Architecture

VI. IMPLEMENTATION

6.1 SCREENSHORTS:



6.1Home page

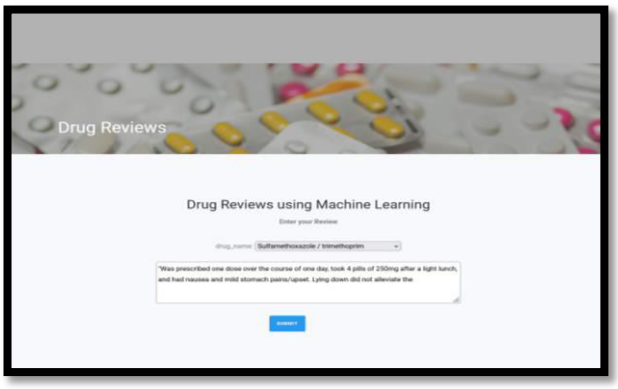


6.2Register Page

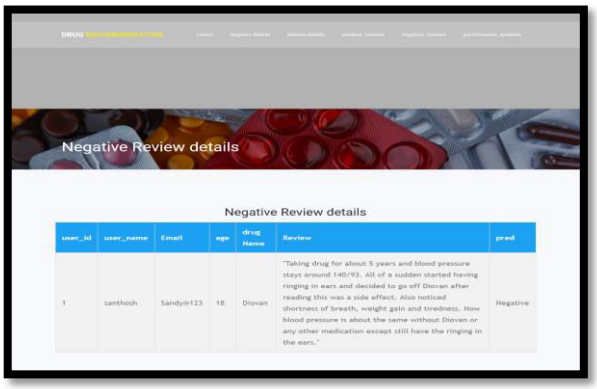
| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707151 |



6.3 Review Detailes



6.4 Negative Review details

VII. TESTING

The goal of testing is to find mistakes. Testing is the practise of attempting to find all possible flaws or weaknesses in a work product. It allows you to test the functionality of individual components, sub-assemblies, assemblies, and/or a whole product. It is the process of testing software to ensure that it meets its requirements and meets user expectations, and that it does not fail in an unacceptable way. There are various types of tests. Each test type addresses a specific testing requirement.

VIII. CONCLUSION

Reviews have become an important part of our daily lives; whether we go shopping, buy something online, or eat at a restaurant, we always read the reviews beforehand to make the best decision. Motivated by this, sentiment analysis of drug reviews was investigated in order to develop a recommender system using a variety of machine learning classifiers, including Logistic Regression, Perceptron, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, LinearSVC, applied to Bow, TF-IDF, and classifiers such as Decision Tree, Random Forest, Lgbm, and Catboost applied to Word2Vec and Manual features methods.We used five different measures to evaluate them: precision, recall, f1score, accuracy, and AUC score, and found that the Linear SVC on TF-IDF surpasses all other models by 93 percent. The Decision tree classifier on Word2Vec, on the other hand, had the poorest performance, with only 78 percent accuracy. To create a recommender system, we added the best-predicted emotion values from each approach, Perceptron on Bow (91 percent), LinearSVC on TF-IDF (93 percent), LGBM on Word2Vec (91 percent), and Random Forest on manual features (88 percent), and multiplied them by the normalised usable Count.

IX. FUTURE ENHANCEMENT

We intend to further develop the medication recommender framework's highlights and helpfulness later on by investigating various choices. In the first place, we will thoroughly analyze different oversampling techniques to eliminate any potential dataset biases. The system will be better able to handle a wider variety of reviews. Additionally, during the vectorization process, experimenting with various n-gram values can assist in determining the most ideal way to portray composed information, which can work on the exactness of mind-set investigation and proposals much more. Making the will play a big role in making the recommender system faster and more useful. Better algorithms work. We need to work on the framework's capacity to suppose and make great ideas by adjusting the parameters of the model and investigating more advanced approaches to optimization. Finally, these Future efforts will contribute to the development of a stronger and more trustworthy drug recommendation system, providing healthcare people and workers need useful information to help them make good decisions.

REFERENCES

[1] Telemedicine, https://www.mohfw.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,

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707151 |

[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.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com