Medicament Guidance for Patients Based on the Drug Reviews

Tanwi Mahaboob[1] Dr. Sasi Kiran Jangala[2]

[1]Research Scholar, Dept. of Computer Science and Engineering, Lords Institute of Engineering & Technology, Hyderabad, Telangana

[2]Professor, Dept. of Computer Science and Engineering, Lords Institute of Engineering & Technology, Hyderabad, Telangana

***Abstract* - Since the coronavirus emerged, it has been increasingly difficult to get legitimate therapeutic resources, such as the scarcity of specialists and healthcare professionals, appropriate equipment and medications, etc. There are many deaths as a result of the medical profession as a whole being in turmoil. Due to a lack of availability, people began taking medication on their own without the proper consultation, which made their health situation worse than usual. Recently, machine learning has proven useful in a variety of applications, and creative work automation is on the rise. This research article aims to propose a system for prescribing medications that can significantly reduce the workload of specialists. In this research project, we develop a drug recommendation system that makes use of patient reviews to forecast sentiment using a variety of vectorization techniques, including Bow, TF-IDF, Word2Vec, and manual feature analysis, which can support the recommendation of the best medication for a given disease by various classification algorithms. Precision, recall, f1score, accuracy, and AUC score were used to assess the anticipated sentiments. The Sequential Model and XGBoost classifier surpass all other models with roughly 95% accuracy, according to the data. We implemented this model in a real-world setting in which users can log in, submit their symptoms, and receive a list of medicines that are recommended for their condition.**

***Index Terms* — Drugs, Sentiment analysis, Stochastic processes, Static VAr compensators, Manuals, Recommender systems, Random forests**

I. INTRODUCTION

Health information is one of the most contentious topics on the Internet. A Pew Internet and American Life Project survey found that 59% of Americans have looked online for health-related information, with 35% of respondents focusing on online medical problem diagnosis. The results show that more people are becoming worried about issues related to health and medical diagnosis. However, drug-related mishaps continue to take numerous lives. The administration estimates that prescription errors result in over 200,000 deaths each year in China and even 100,000 deaths in the US. Because they write prescriptions based on their limited experience, doctors are accountable for more than 42% of drug mistakes.

The following details could be a factor in these issues: For critical illnesses, many hospitals are short on doctors or medical specialists, and (ii) expert diagnosis mostly rests on the expertise of the expert, especially for those inexperienced novices who find it challenging to avoid mistakes. It is now hard to determine the value of the data because the majority of diagnosis case data in hospitals is still being maintained and has not been mined. Hospital information systems (HIS) generate enormous volumes of data, making it challenging to glean potentially

insightful information from diagnosis case data. Using data mining and recommender systems is a practical way to tackle these challenging issues.

Since the middle of the 1990s, many recommender system concepts have been proposed, and in recent years, numerous types of recommender system software have been developed for a wide range of applications. Traditional recommender systems include collaborative filtering, content-based, knowledge-based, and hybrid recommendation systems. All of them have disadvantages. The concerns in CF are cold-start, sparseness, and scalability, while the solutions in CB are extremely specific. To solve these problems, we design our own architecture for a medical recommender system using data mining and machine learning methods.

II. SYSTEM ANALYSIS

Problem Statement:

Due to the exponential rise in coronavirus cases, there is a global doctor shortage, particularly in rural areas where there are less medical professionals than in urban areas. A doctor needs spend six to twelve years finishing their education. As a result, adding more doctors in a short amount of time is not conceivable. Clinical blunders happen frequently today. Over 200 000 people in China and 100,000 persons in the USA are affected by drug errors each year. Specialists commonly (more than 40% of the time) prescribe the incorrect prescription since they only have access to a restricted amount of information. For those who require medical specialists with comprehensive understanding of microscopic disorders, selecting the correct prescription is crucial. This essay aims to propose a system for prescribing medications that can significantly lessen the workload of specialists. In this study, we develop a drug recommendation system that makes use of patient reviews to forecast sentiment using a variety of vectorization techniques, including Bow, TF-IDF, Word2Vec, and manual feature analysis, which can support the recommendation of the best medication for a given disease by various classification algorithms.

Objective:

A technique for prescription pharmaceuticals that can greatly reduce the workload of specialists is what this research effort intends to suggest. The term "vectorization" refers to a traditional technique for transforming input data from its original text-based format into real-number vectors, which is the format that is supported by ML models. This method has been around since the invention of computers, it has proven extremely successful across numerous disciplines, and it is now utilised in NLP.Vectorization is a phase in the feature extraction process in machine learning. By translating text to numerical vectors, the goal is to extract some distinguishing features from the text for the model to train on. Using a variety of vectorization techniques, such as Bow, TF-IDF, Word2Vec, and manual feature analysis, we develop a drug recommendation system in this study that uses patient reviews to forecast sentiment. This system can support the recommendation of the best medication for a given disease by a variety of classification algorithms.

Aim of the project

In order to help doctors and patients learn more about medications for specific medical situations, a medication recommender system is unavoidably required. A common system called a "recommender framework" suggests products to users based on their needs and benefits. These frameworks examine the replies to consumer surveys and make recommendations based on the specific requirements of the respondents. Sentiment analysis and feature engineering are used by the medicine recommender system to conditionally offer medications in response to customer reviews.

Scope of the project

Sentiment analysis is a set of techniques, procedures, and instruments for identifying and separating emotional information, such as attitudes and opinions, from linguistic expression. In contrast, "feathering engineering" entails enhancing model performance by adding new features to the ones that currently exist.

Algorithmic process

The technical method for the Drug Recommendation System using the Drug Review Dataset is shown below.

1. Cleaning and displaying data

2. Data Exploratory and Feature Extraction
3. Data Preparation Using TF-IDF, Word2Vec, and BOW
4. Dividing the data set into training and test data
5. Data modelling with the XGBoost classifier and sequential data
6. Evaluation and Prediction of Data
7. Developing a mechanism for drug recommendations

In order to perform Common Conditions Analysis, Common Drugs Analysis, Common Reviews Analysis, Available Drugs per Condition Analysis, Drugs Used Over For Conditions Analysis, and Condition Count Analysis, exploratory data analysis is performed on the dataset in order to clean the dataset for any missing data, identify patterns, and identify the relationships of various output parameters with the help of graphs, statistics, etc.

Machine Learning Algorithms Evaluation

- K-Nearest Neighbour:
- Decision Trees:
- Support Vector Machine:
- Naive Bayes:
- Stochastic Gradient Descent:

Proposed System:

In order to help doctors and patients learn more about medications for specific medical situations, a medication recommender system is unavoidably required. A common system called a "recommender framework" suggests products to users based on their needs and benefits. These frameworks use consumer surveys to analyse the replies and provide recommendations tailored to the specific requirements of the respondents. Sentiment analysis and feature engineering are used by the medicine recommender system to conditionally offer medications in response to customer reviews. Sentiment analysis is a collection of methods, procedures, and tools for locating and disentangling emotional content in language, such as attitudes and opinions. In contrast, "feathering engineering" entails enhancing model performance by adding new features to the ones that currently exist.
The findings show that the Sequential Model and XGBoost classifier outperform all other models with approximately 95% accuracy. Users can log in, enter their symptoms, and obtain a list of medicines that are suggested for their condition. We applied this model in a real-world situation.

Advantages:

• Drug recommender systems that can effectively predict outcomes use sentiment analysis and feature engineering.
• By receiving a recommended medication list for their illness. maximise profits while minimising losses.

ALGORITHM/ TECHNIQUE USED

After Feature Transformation and Data Pre-processing, the dataset is fitted to a model. The algorithm receives the training set in order to learn how to forecast values. After creating a target variable to predict, testing data is provided as input. The models are created using a technical implementation for the Drug Review dataset for the Drug Recommendation System.

- Splitting the data set into test and training data;
- Data cleaning and visualisation;
- Feature Extraction;
- Data preparation using BOW, TF-IDF, and Word2Vec;
- Data modelling using Sequential and XGBoost classifier;
- Data evaluation and prediction;
- Building the Drug recommendation system.

The technical method for the Drug Recommendation System using the Drug Review Dataset is shown below.
1. Cleaning and displaying data
2. Secondly Feature Extraction
3. Data Preparation Using TF-IDF, Word2Vec, and BOW
4. Dividing the data set into training and test data
5. Data modelling with the XGBoost classifier and sequential LSTM
6. Evaluation and Prediction of Data
7. Developing a mechanism for drug recommendations

III. PROPOSED MODULAR IMPLEMENTATION

Data Preparation for Final Classifier

1. Examining the data
2. Changing the date's format to datetime
3. Taking the year from the date
4. Taking the month out of the date
5. Taking the days out of a date
6. Remove the rows to prevent overfitting the data.
7. Eliminating rows with just one medication condition
8. Eliminating the conditions containing "span"
9. Putting words in a sack
10. Divide the Dataset into Test and Training Subsets.
11. Provide the dataset to the final machine learning classifier algorithms.

Create the final model using Sequential model and XGBoost classifier

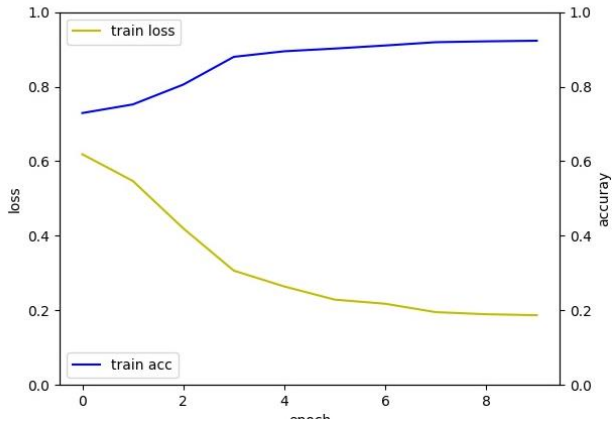
```
# Model Structure
model = keras.models.Sequential()

model.add(keras.layers.Dense(200, input_shape=(20000,)))
model.add(keras.layers.BatchNormalization())
model.add(keras.layers.Activation('relu'))
model.add(keras.layers.Dropout(0.5))

model.add(keras.layers.Dense(300))
model.add(keras.layers.BatchNormalization())
model.add(keras.layers.Activation('relu'))
model.add(keras.layers.Dropout(0.5))

model.add(keras.layers.Dense(100, activation='relu'))
model.add(keras.layers.Dense(1, activation='sigmoid'))
```

```
clf = XGBClassifier(
    n_estimators=2000,
    learning_rate=0.05,
    num_leaves=30,
    # colsample_bytree=.9,
    subsample=.9,
    max_depth=7,
    reg_alpha=.1,
    reg_lambda=.1,
    min_split_gain=.01,
    min_child_weight=2,
    silent=-1,
    verbose=-1,
```

Find sentiment of the reviews

Generate recommendations of drugs and save it in excel format.

Below is the accuracy graph of the final classifier.



Accuracy of the final classifier is about 93%

Algorithm	**Accuracy in %**
SVM	75
Logisic Regression	76
KNN	72
Naïve Bayes	71
Decision Trees	77
Stochastic Gradient Descent	73

The project's planned modular implementation is shown below. There are two modules in it:

1. Admin
2. User

Admin Module:

The system administrator is in charge of tasks like:

1. Check out google finance for stock info
2. The dataset's data analysis
3. Divvying up the dataset into training and test halves
4. Developing the model with several algorithms

5. Examine how well the algorithms performed on the provided dataset.
6. Use the LSTM technique to build the model.

User Module:

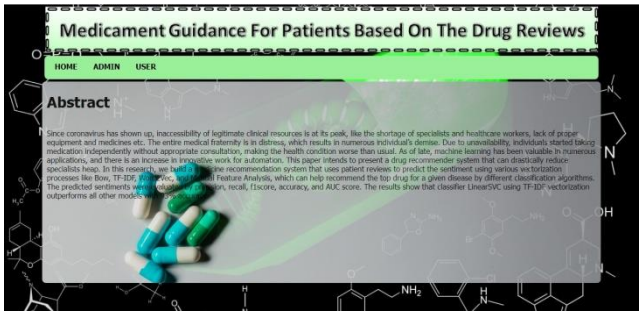
The system's user may take advantage of the following available machine learning services:

1. logging in and entering stock information to forecast future trends
2. Get future trend predictions

IV. PROJECT EXECUTION

Home page:

This is the application's home page when PyCharm is used to run the programme. The programme is hosted on a web server, and a URL is produced to access it. When a user clicks on the URL, the page shown below opens in a browser.

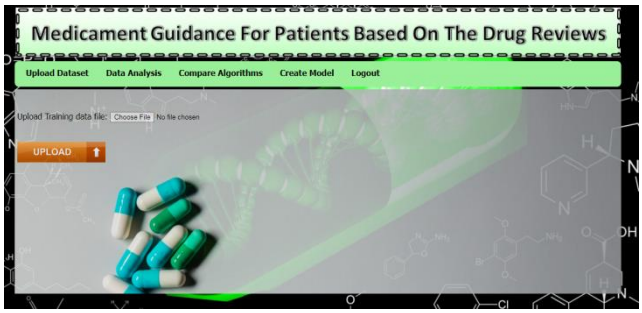


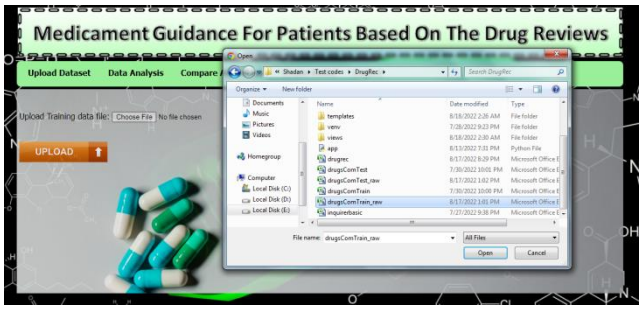
Admin Login:

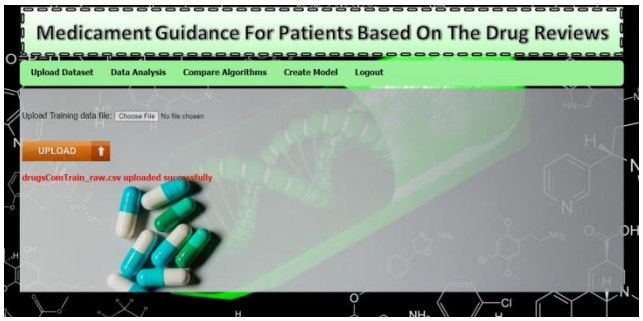
The admin module's login page is located here. In order to carry out actions like uploading the dataset, the administrator must first log in with his credentials. During dataset training, a dataset's exploratory data analysis, using a dataset to train many machine learning algorithms to determine the optimal algorithm for accuracy and Create a model that can be utilised by users and hosted on the Flask application.

Upload Dataset:

The system administrator can upload datasets that are used to train machine learning models on this page. To upload a file to a server, an administrator must first choose the file by clicking the Choose file button, then click the Upload button. A success message indicating that the file was successfully uploaded would be shown once the upload was finished. As a dataset for this research, we are utilising Train_3.csv.







Exploratory Data Analysis:

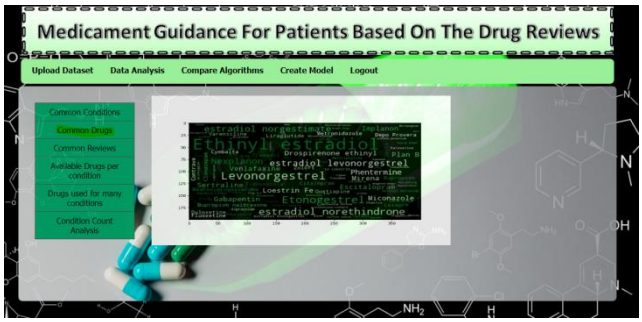
In order to perform Common Conditions Analysis, Common Drugs Analysis, Common Reviews Analysis, Available Drugs per Condition Analysis, Drugs Used Over For Conditions Analysis, and Condition Count Analysis, exploratory data analysis is performed on the dataset in order to clean the dataset for any missing data, identify patterns, and identify the relationships of various output parameters with the help of graphs, statistics, etc.

Common Conditions Analysis:

The graph below displays the most frequent crime types from the training dataset drugsComTrain_raw.csv file.

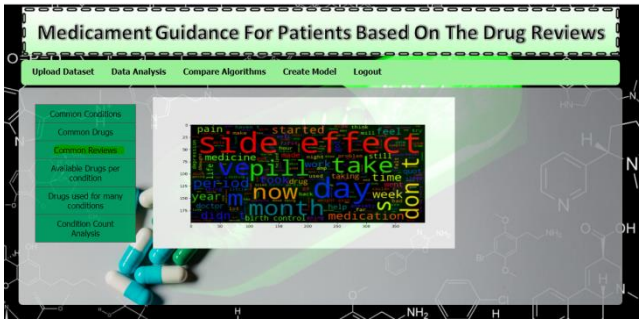
Common Drugs Analysis:

The graph below, which is taken from the training dataset drugsComTrain_raw.csv file, displays the common drugs analysis graph for the last year's worth of stock data for Microsoft.



Common Reviews Analysis:

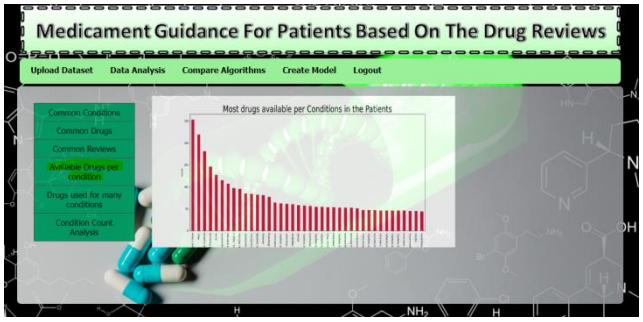
The graph below displays the Common Reviews Analysis graph for Microsoft Company's Stock Data during the course of the previous year, taken from the training dataset drugsComTrain_raw.csv file.



Available Drugs per Condition Analysis:

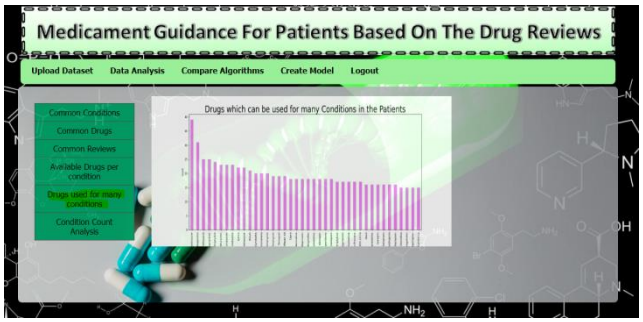
The training dataset drugsComTrain_raw.csv file's Available Drugs per Condition Analysis graph is shown below. It

displays various crimes that have been reported over time.



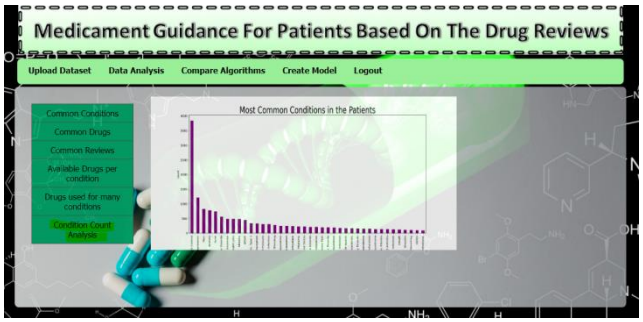
Drugs Used Over For Conditions Analysis:

The drugs used to treat various conditions are depicted in the graph below. Analysis graph for the training dataset drugsComTrain_raw.csv file's proportions of crime on ordinary days versus on holidays.



Condition Count Analysis:

From the training dataset drugsComTrain_raw.csv File, the Condition Count Analysis graph for Proportions of Crime During Regular Days Vs. Holidays is displayed below.

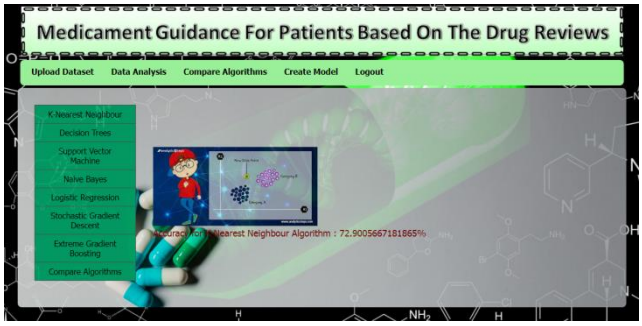


Compare Algorithms:

The administrator can use this page to train several algorithms on a dataset and determine each algorithm's test accuracy.

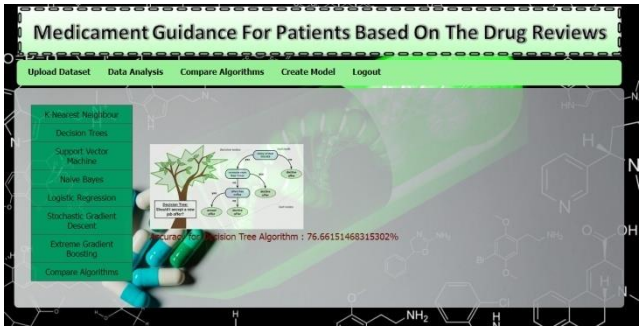
K-Nearest Neighbour:

The test accuracy is 72.90% when the dataset is fed to the K-Nearest Neighbor algorithm, as we can see.



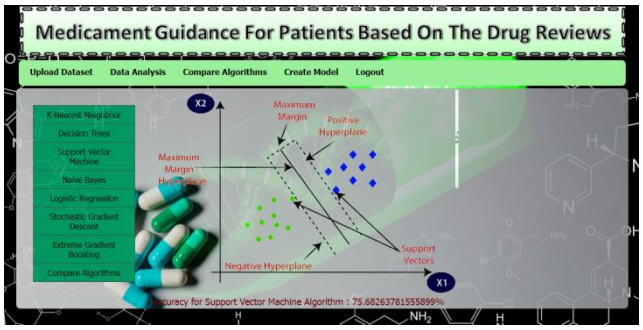
Decision Trees:

The test accuracy is 76.66% when the dataset is fed into the decision tree method, as we can see.



Support Vector Machine:

The test accuracy is 75.68% when the dataset is fed into the Random Forest algorithm, as we can see.

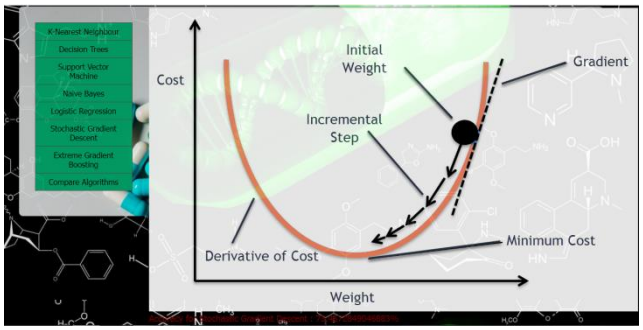


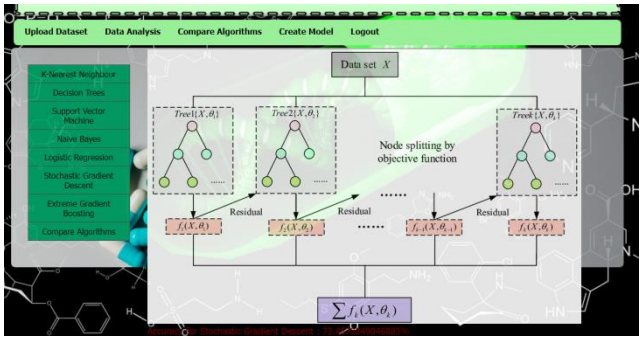
Naive Bayes:

The test accuracy is 72.43% when the dataset is fed into the Naive Bayes algorithm, as we can see.

Stochastic Gradient Descent:

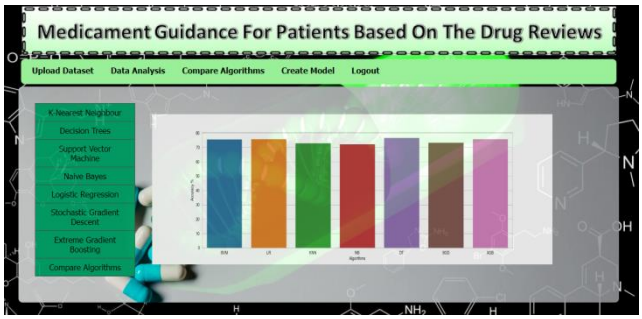
When the stochastic gradient descent algorithm is used with the dataset





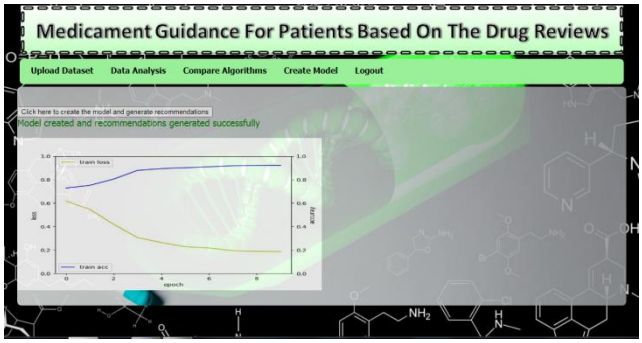
Compare Algorithms:

This screen displays a comparison of the algorithms' various test accuracies.



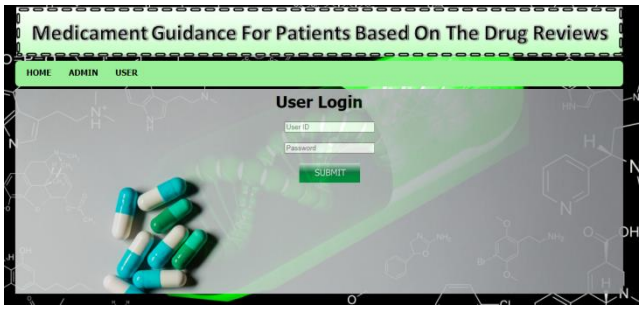
Create Model:

This screen displays the development of a model for system optimisation.



User Home Page:

This is the user module's user home page. For the purpose of facilitating data analysis and prediction over a set of drug data from the company, the user must log into the system using his credentials.



Prediction

This is on the user module's User Home Page. To enable Drug prediction over Drug dataSet, the user must log in to the system using his or her credentials.

CONCLUSION

Using the system created for Medication Guidance For Patients Based On The Drug Reviews, we correctly forecasted the medications in this project. Our study article's goal is to offer a system for prescribing pharmaceuticals that can greatly lessen the strain of specialists in order to achieve this. Using a variety of vectorization techniques, such as Bow, TF-IDF, Word2Vec, and manual feature analysis, we develop a drug recommendation system in this research project that uses patient reviews to forecast sentiment. This system can support the recommendation of the best medication for a given disease by a variety of classification algorithms. To evaluate the expected sentiments, precision, recall, f1score, accuracy, and AUC score were utilised. The findings show that the Sequential Model and XGBoost classifier outperform all other models with approximately 95% accuracy. Users can log in, enter their symptoms, and obtain a list of medicines that are suggested for their condition. We applied this model in a real-world situation.

Future Aims: In the future, we can improve the use of medication guidance for patients based on drug reviews, which identifies the main elements of health problems and can then be further divided into specialised areas and combinational categories. By attempting these improvements, the system is given the ability to offer tailored prescription suggestions and gets rid of redundant drug recommendations for those with various health conditions.

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, 4.

[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, Chonghui Guo, Hui Xiong, and Yanming Xie. 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