UG Project Presentation

Efforts By:

Tushar Shandilya (18135116)

Sumit Kumar Pradhan (18135100)

Ankit Sarangi (18135117)

Vijaykumar A. Lokhande (18135119)

Anshul Shrivastava (18135014)

Supervised By:

Dr. Cherian Samuel

Mechanical Engineering Department,

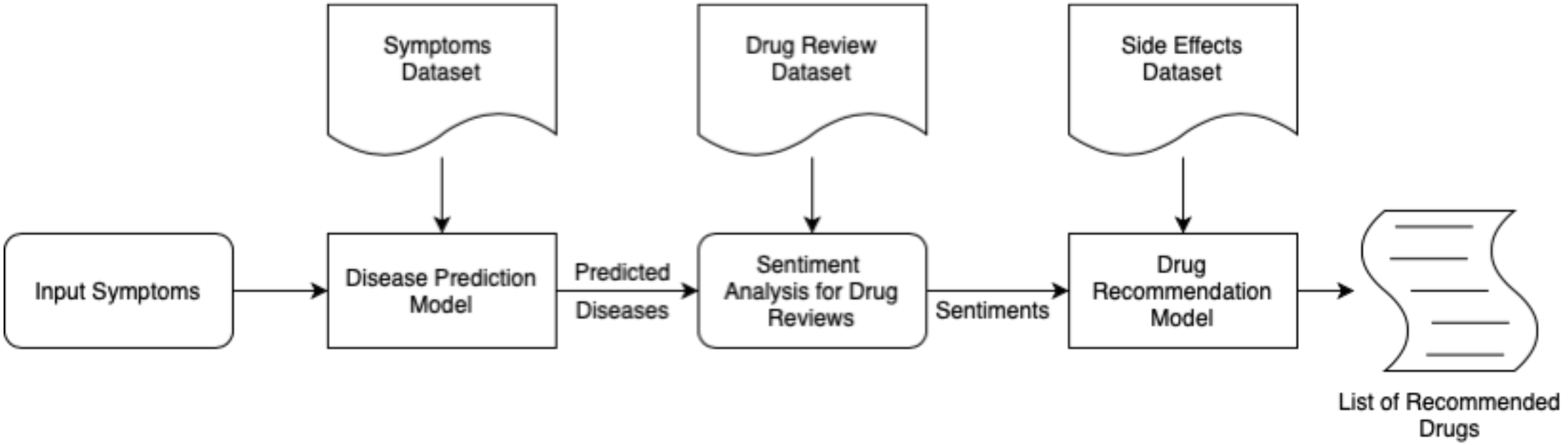
IIT BHU VARANASI

Objective

The objective of this project was to design and implement a universal Disease Prediction and Drug Recommendation System that applies various technologies to the recommendation system. By combining information from different sources we have used various prediction algorithms along with NLP for sentiment analysis and recommendation.

Methodology

The main goal of our project is to recommend a drug to a patient based on the symptoms she/he has. In accordance with our objective to implement a drug recommender system there are two main subcategories which are to be addressed i.e a disease prediction model and a recommendation model and below is the design pipeline and dataflow of our implementation.



Dataset And Preprocessing

Data Gathering

- Symptoms Dataset
- Drug Review Dataset
- Side Effects Dataset

Data Preprocessing

- The gathered datasets were checked and all the irrelevant and null values were dropped from the dataset.
- The Dataset containing symptoms was merged with the one with reviews to ne used for final drug prediction.

Disease Prediction

Approach One

- The preprocessed symptoms dataset was used to map the newly created disease prediction data frame by setting the values of symptoms present for a particular disease as 1 and 0 if not and the rows were grouped to obtain a single row for each unique disease with the corresponding symptoms column mapped to 1.
- The data was then trained using Decision Tree Classifier Model, Support Vector Machine Classifier Model and was used to predict the diseases by passing one, two, and three features to the classifier models.

Approach Two

- In this method, the order of importance of each symptom for a disease was preserved and the mapping of 0's and 1's were done in a probabilistic manner. The symptom with the highest importance was mapped to 1 for all the rows of that particular disease. The second most important symptom was mapped to 95% of all the rows and so on.
- The data was then trained using Decision Tree Classifier Model, Support Vector Machine Classifier Model and was used to predict the diseases by passing one, two, three and four features to the classifier models.

Sentiment Analysis of drug Reviews

In this step, we adopted sentiment analysis using the Natural Language Processing on the drug review data set to understand the trend in the positive and negative reviews given by the patients. The main rationale behind doing this sentiment analysis was the rating given and the review stated was seen to be inconsistent so likely if the review stated it was good and no side effects it still had a rating below 5 which made it hard to just make use of the rating based on some threshold and as NLP captures the essence of a sentence to predict a sentiment hence it proves to be more reliable.

Drug Recommendation

Features considered for model prediction:

- Reviews with positive sentiment.
- Rating given by the user in the review.
- Useful count (how many users have given that particular rating).

Weighted Average Approach:

- This approach allows us to use all the information present in the dataset according to its importance and hence as an output we get the best possible drug recommendation.
- In this approach, data elements with a high weight contribute more to the weighted mean than do elements with a low weight.

Results

Approach One:

- Decision Tree Classifier produced 89.93% accuracy.
- SVM Classifier produced 87.58% accuracy.

Approach Two:

- Decision Tree Classifier produced 86.95% accuracy.
- SVM Classifier produced 86.96% accuracy.

Conclusion and Future Scope

Key Conclusions:

- Successfully built a drug recommendation system that predicts diseases and recommends drugs along with possible side effects based on user symptoms input.
- Designed three models for this project implementation. A disease prediction model, Sentiment Analysis model and a recommendation model.
- Each of the three models gave good accuracy contributing to the overall reliability of the drug recommendation model.

Future Scope:

- Improving the accuracies of the prediction and recommender model using deep neural networks by using larger data.
- Demonstration of the project is done using the Chatbot as a UI between the code that we've written and its results for better understanding.
- Chatbot UI can be integrated to websites that cater towards online medical services.

Thank you