Evaluating Patient Satisfaction Through Drug Reviews

Vidhi Patel

Objective

- ➔ Analyze drug reviews to understand sentiment and satisfaction levels of patients
- ➔ Develop sentiment analysis algorithms to classify reviews as positive, neutral, or negative
- ➔ Explore correlations between drug characteristics, such as dosage and side effects, and patient satisfaction
- ➔ Provide insights to healthcare providers and pharmaceutical companies for improving patient care

Motivation

- ➔ There is a limited understanding of patient satisfaction levels with prescribed medications
- ➔ Healthcare providers need the insights into patient feedbacks to tailor treatment plans effectively
- ➔ Implementing this system can help analyze the sentiment and predict patient needs easily
- ➔ Utilizing machine learning models can empower healthcare providers with data-driven insights

Goal

- ➔ Collect a diverse dataset of drug reviews and preprocess the data by cleaning and tokenizing, removing stopwords, etc.
- ➔ Conduct sentiment analysis to classify reviews into positive, neutral, and negative sentiments
- ➔ Develop ML models such as SVM, Naive Bayes, and Random Forest Classifier to predict the satisfaction levels
- ➔ Evaluate model performance using metrics and generate actionable insights to enhance patient care

Dataset Description

- ➔ Dataset Size: 169MB
- ➔ Data Types: int64, string, and date
- ➔ Target: review column
- ➔ Features: There are in total 11 different features in this dataset

- ➔ drugName: name of the drug
- ➔ condition: patient suffering from
- ➔ Review: patients' review
- ➔ Rating: 1-10 star patient rating
- ➔ sideEffects: patient issues
- ➔ effectiveness: patient rates if the product is effective or not

Steps Performed

- ➔ Descriptive Statistic
- ➔ Data Preprocessing
- ➔ Feature Engineering
- ➔ Data Visualization
- ➔ Model Training & Building

Models Used

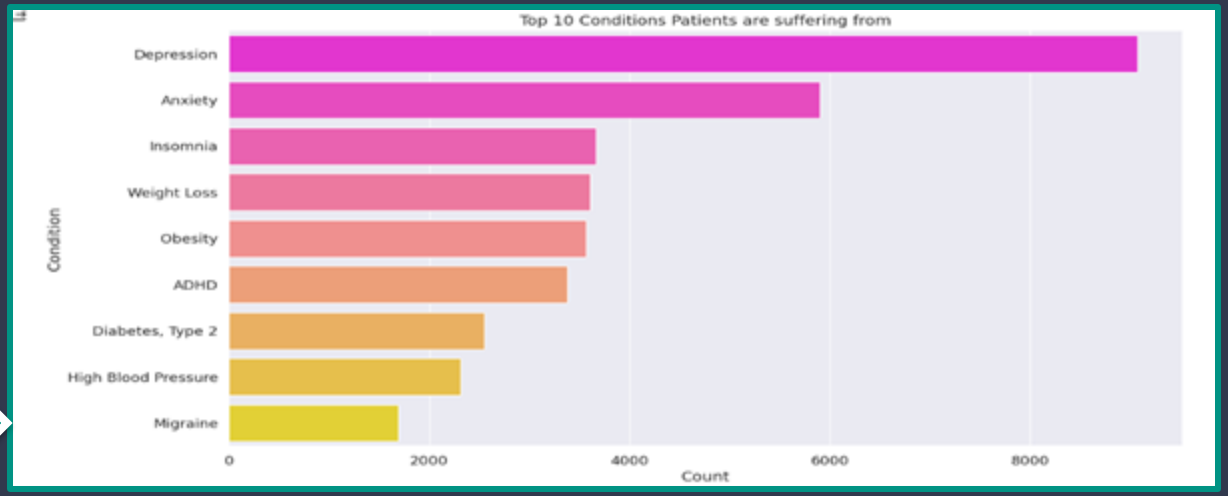
- ➔ Random Forest: Utilizing for its ability to handle large datasets with high dimensionality, providing accurate predictions
- ➔ Support Vector Machine: To classify drug reviews into satisfaction categories, works with linear and non-linear data
- ➔ Decision Tree: Easy to interpret and visualize, allowing identification of important features

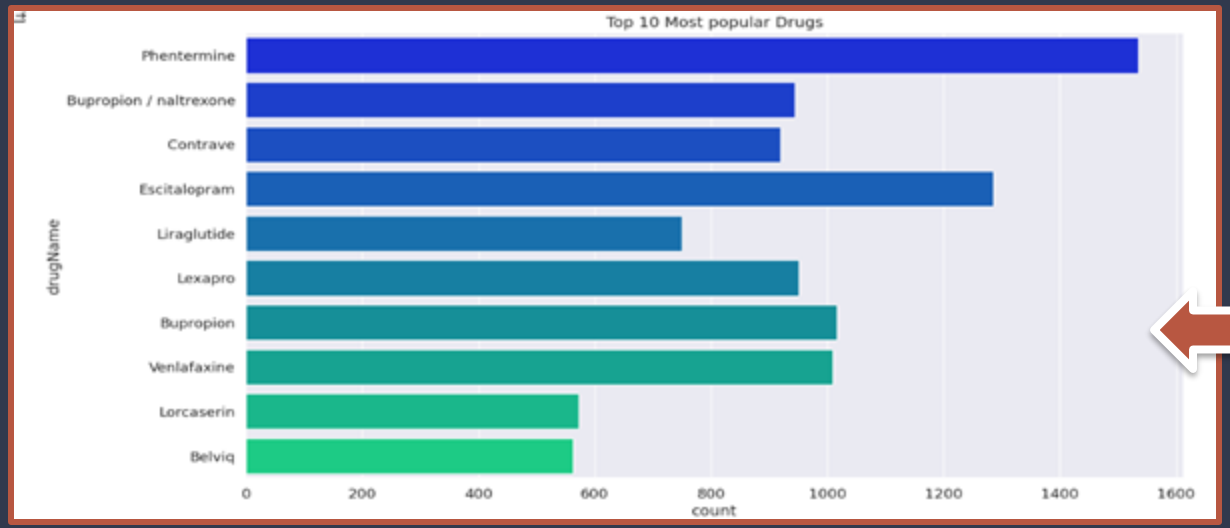
Models Used

- ➔ Naive Bayes: It is simple and efficient in analyzing text data, well suited for large datasets and also it can indicate positive or negative sentiments
- ➔ Logistic Regression: good for knowing the probability of patient satisfaction with a drug based on different factors

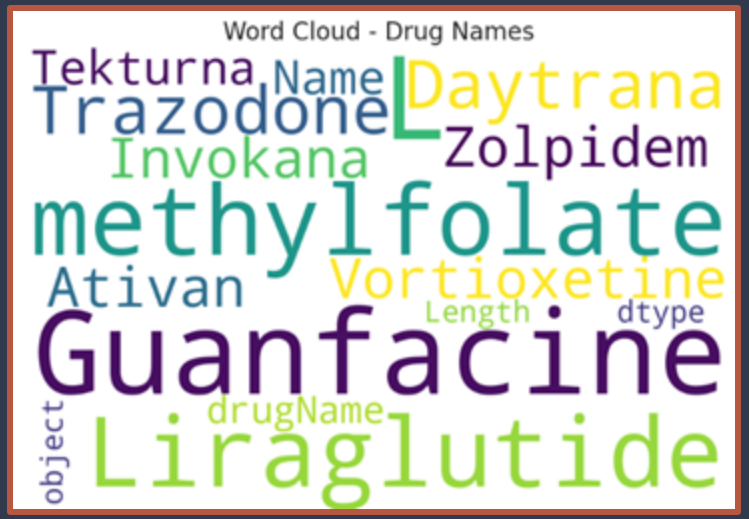
Results

Top 10 conditions that patients are suffering from



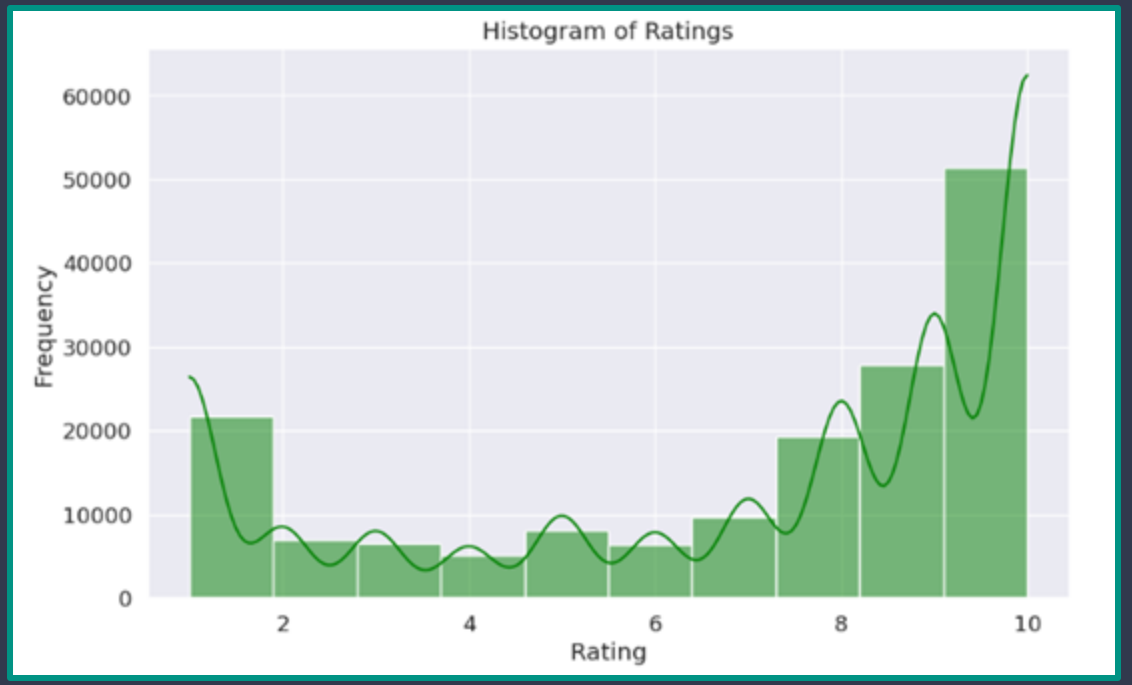


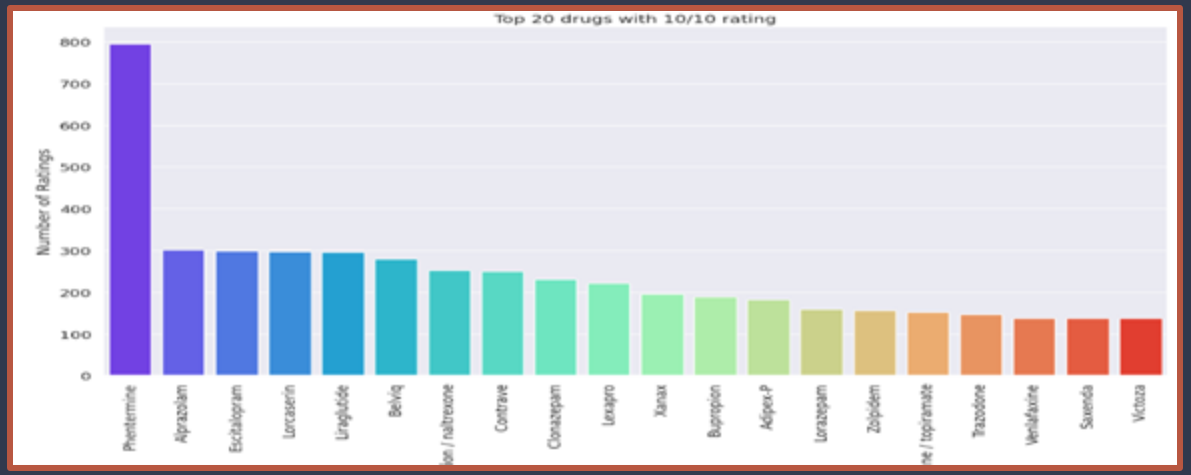
Top 10 most popular drugs that patients rated



WordCloud of most frequent Drugs

Histogram to view the frequency of ratings



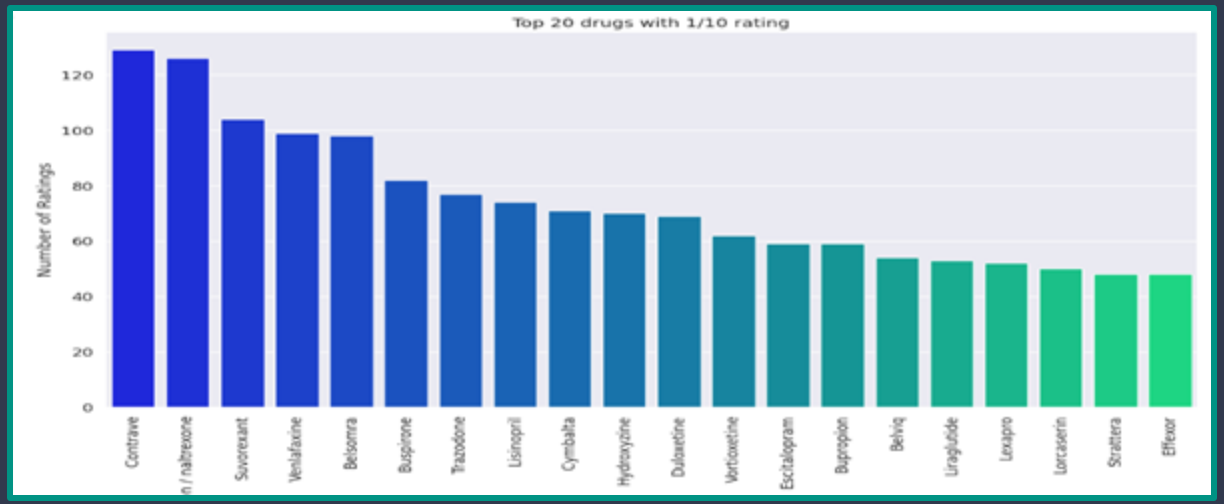


Top 20 drugs with

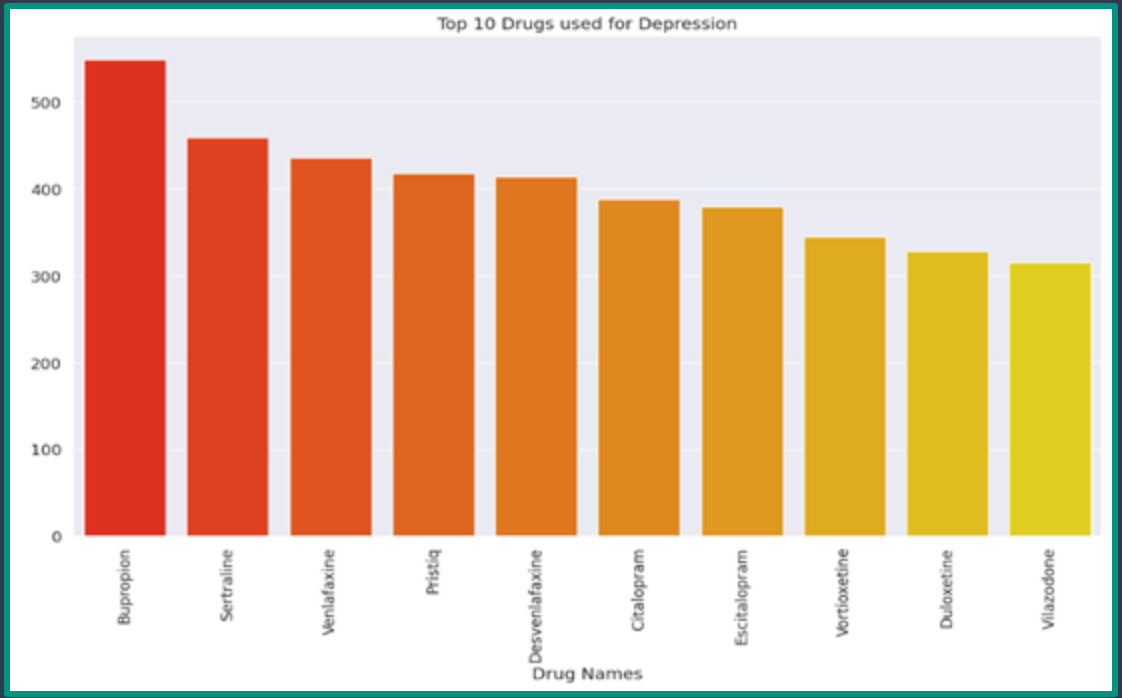
10 star ratings

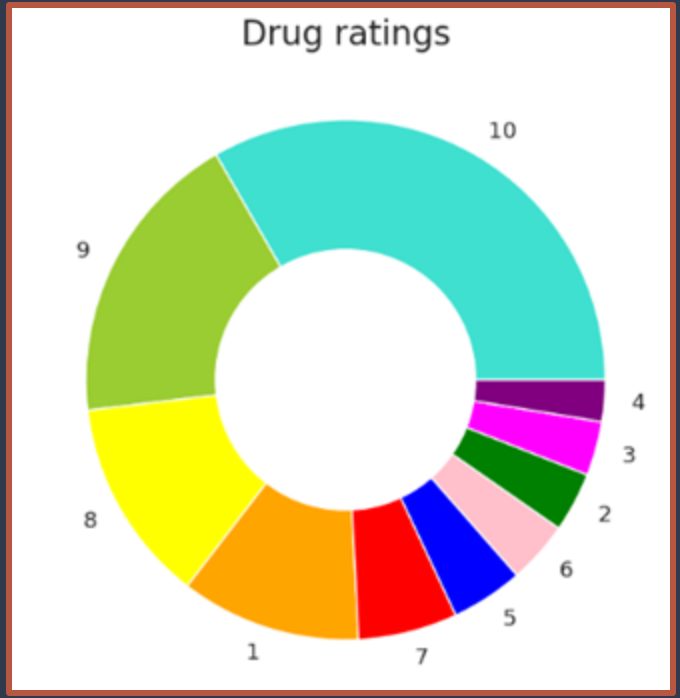
Top 20 drugs with

1 star ratings



10 drugs that are used for Depression condition





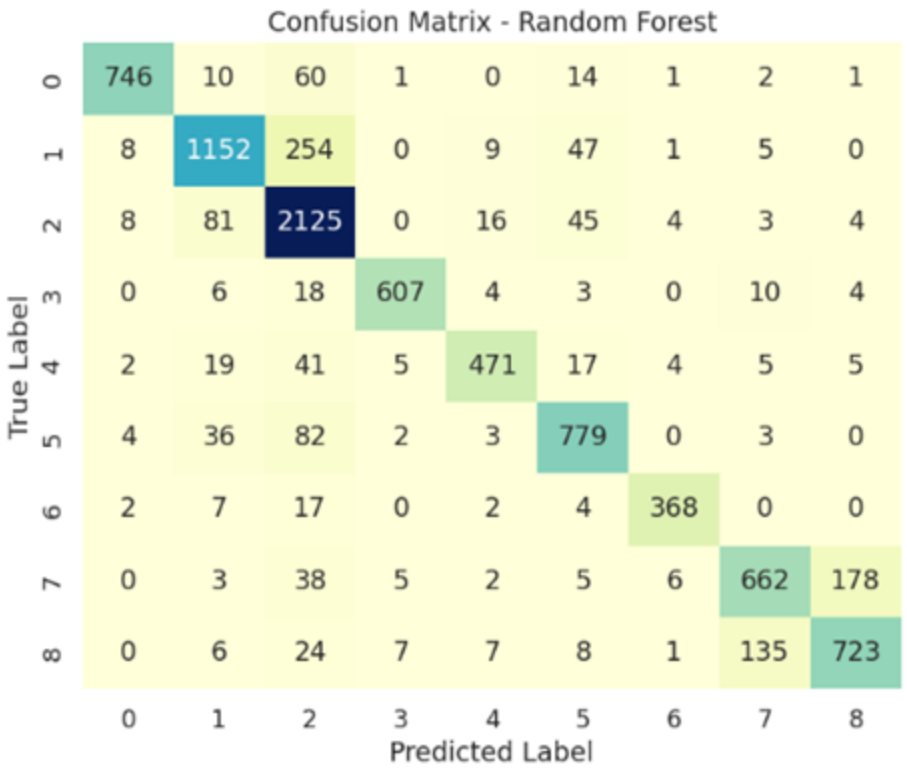
Overall ratings for drugs

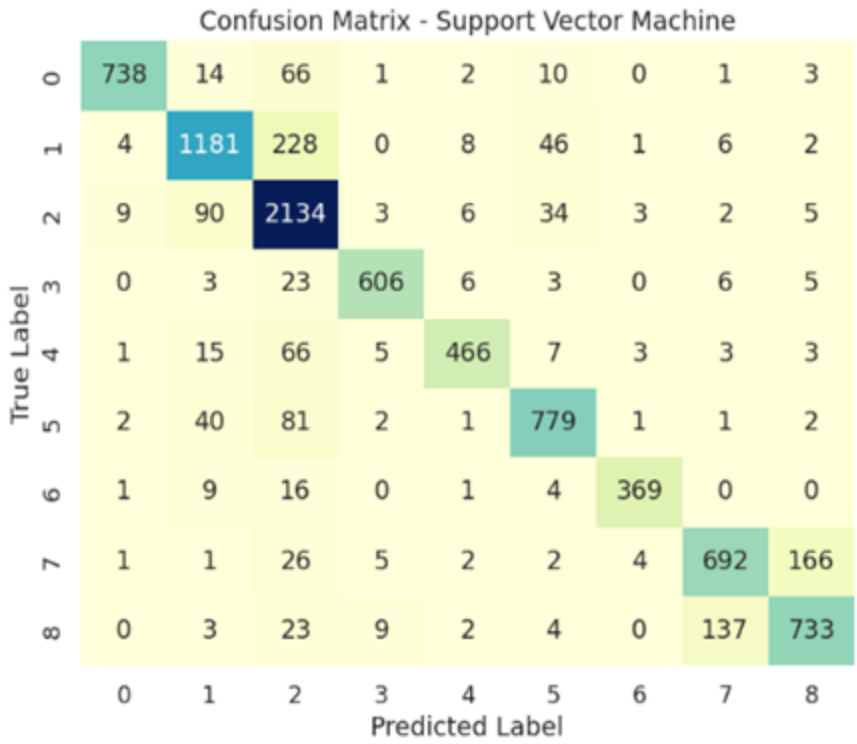
WordCloud for Sentiments

Pie chart for polarity of sentiments

Random Forest

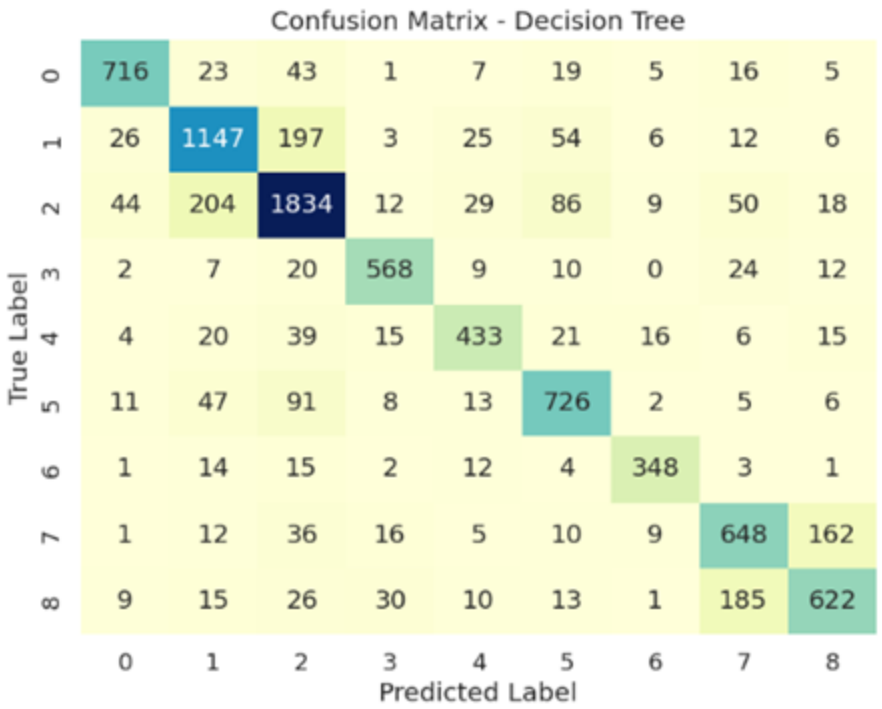
SVM

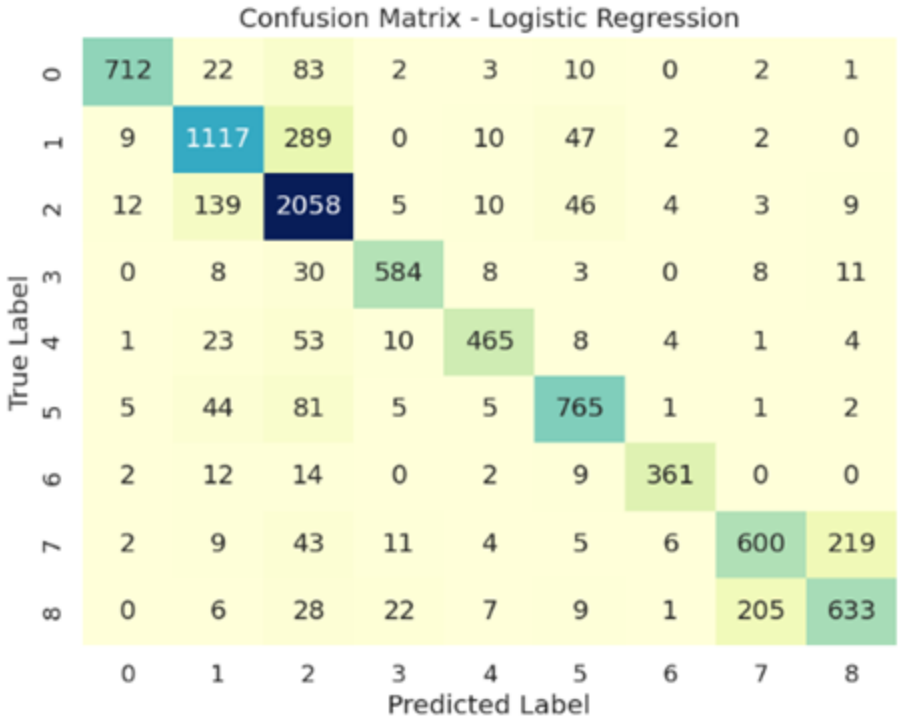




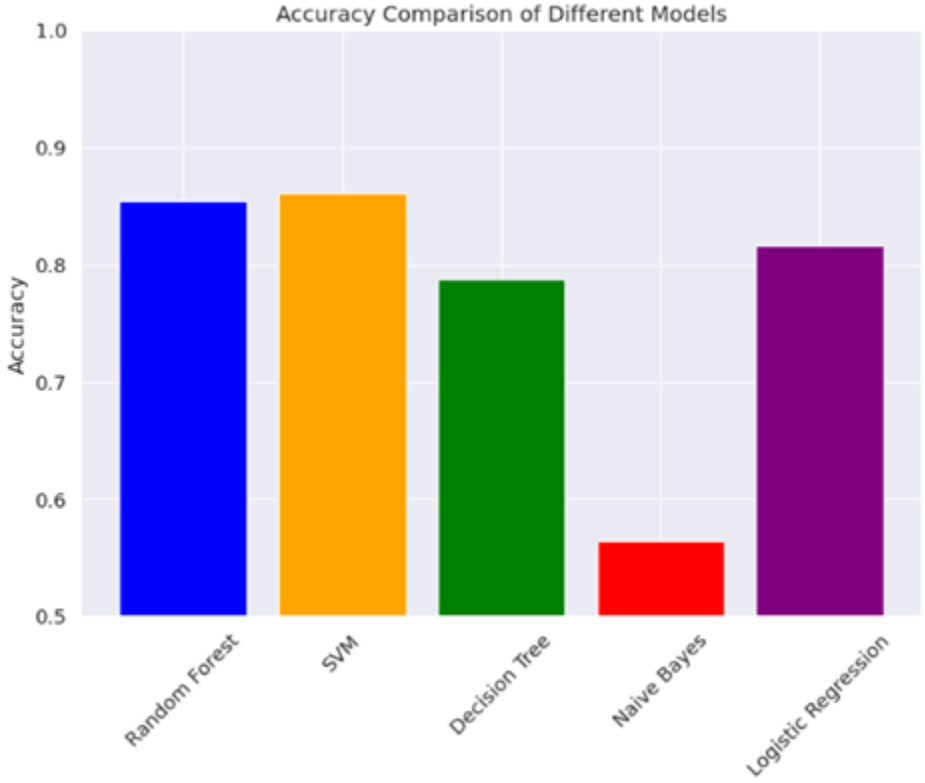
Decision Tree

Logistic Regression





Overall comparison of Accuracies



Conclusion

- ➔ Random Forest and SVM models were the top models giving the highest accuracy than other models
- ➔ Naive Bayes did not perform well due to its limitation capturing sentiment patterns within the reviews
- ➔ Overall, this project highlights the potential of ML techniques in analyzing large-scale patient feedback data to enhance healthcare decision making and improve patient outcomes

References

1. Sentiment Analysis in Drug Reviews using Machine Learning and Deep Learning Techniques.
2. Exploring Drug Sentiment Analysis with Machine Learning Techniques.
3. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis.
4. Drug review sentimental analysis based on modular lexicon generation and a fusion of bidirectional threshold weighted mapping CNN-RNN

Thank You