DRAW

Drug Review Analysis Work

What Inspires us

- We can get easily overwhelmed by numerous online reviews listed for medicines available for any conditions
- Due to various limitations in clinical trials, adverse side effects may not be detected before the drug is launched in the market
- Text analysis on drug reviews could be useful for patients, pharmacy companies and doctors to improve consumer safety by assisting in the reduction of medication errors
- There is no tool available in the market to suffice these issues. Therefore, we will create a web application where patients and doctors can search by symptoms and get drug recommendations, side effects of drugs and obtain insights into patients' portfolio.

Problem Statement

- We aim to answer the following questions through our platform:
  - How to use sentiment analysis on reviews and recommend the most effective drugs for the given condition?
  - To know the side effects and exact polarity of drug reviews to help pharmacy companies take required action to improve the drug
  - Help doctors and pharma identify the age group and gender of target audience for a chosen drug
  - Provide consolidated information for side effects, emotional inclination towards drugs and classification of drug reviews

Methodology

RAW DATA

PROCESSED DATA

Entity Resolution + Fuzzy Matching

Uniform Patient Conditions

Drug Review + NLTK Vader

Sentiment Analysis

Review + Rating + LSTM

Rating Prediction

Real time review + Deep NN (Sequential Model)

Sentiment Classification

Review + NRC Emotion Lexicon

Emotion Classification

Drug Recommendation based on Condition

Data Sources and Collection

Drugs.com Professional Edition.

WebMD Better information. Better health.

DrugLib.com Drug Information Portal

kaggle

Scrapy

Beautiful Soup <html>

SCRAPPED DATA

Filtering

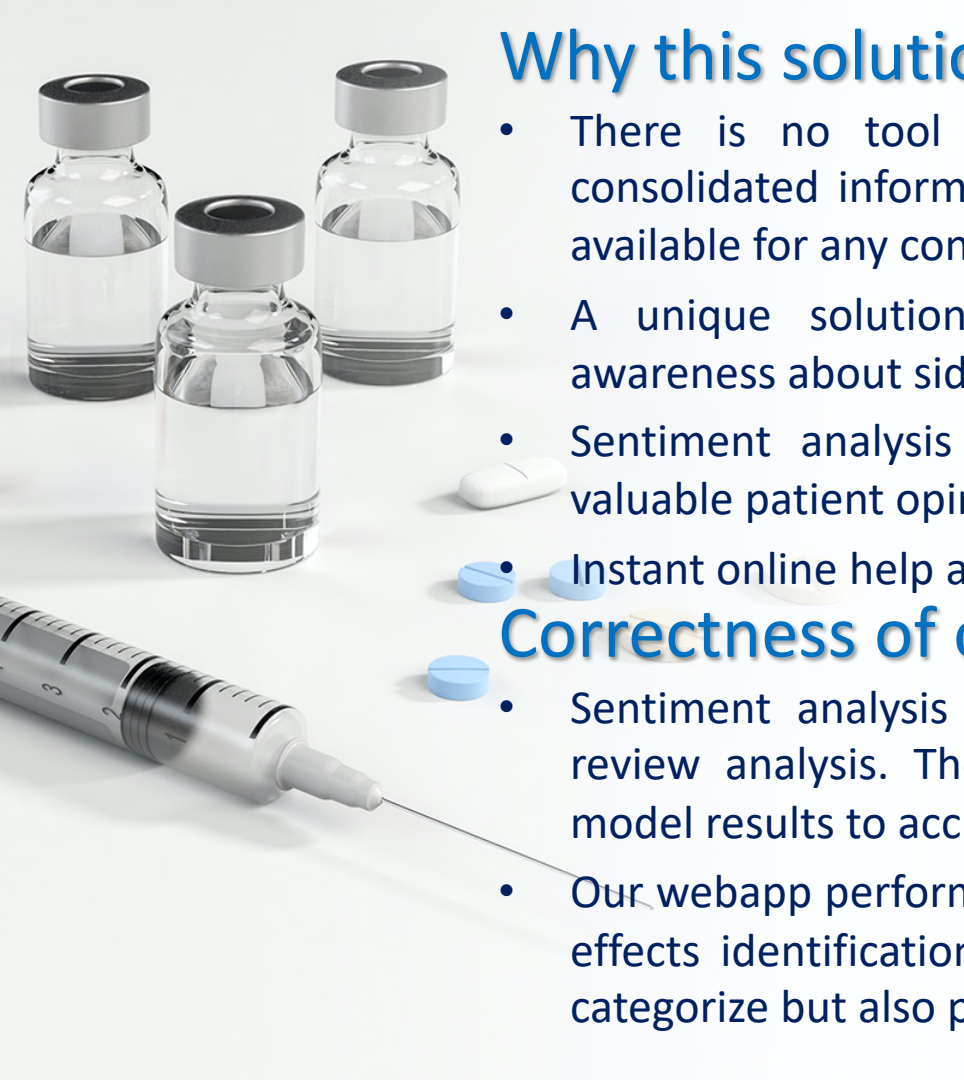
Wrangling

Pre-Processing

PostgreSQL Database

Data Science Pipeline

Data Collection	Scrapped drugs data from WebMD and DrugLib using Beautiful soup and Scrapy and collected Drugs.com data from Kaggle
Data Cleaning	Used Entity Resolution (Jaccard Similarity) and Fuzzy Matching to create uniform patient conditions, filtered data and performed data grouping(age)
Preprocessing & EDA	Visualized most common drugs, conditions, rating distribution, sentiment score vs rating and review useful count vs rating
Data Aggregation	Aggregated matching drugs and condition from 3 websites into one dataset, removing large scale shifts
Feature Engineering	Quantify review words using Bag of Words models like TF-IDF, Hashing vectorizer and normalized sentiment rating to review rating scale
Predictive Modelling	Used user review to predict rating and classify sentiment using ML models (LR, RF, XGBoost, SVM) & Deep Learning models (LSTM, GRU, Sequential)
Model Selection	Selected the best model using multiple criteria like confusion matrix, model accuracy, early stopping, model checkpoint & classification report
Web Application	Interactive webapp using Plotly and Dash, built Drug Recommendation System, implemented real time sentiment classification, rating vs age distribution, emotion analysis of reviews, TSNE plots (reviews, side effects), and side effects wordcloud



Why this solution?

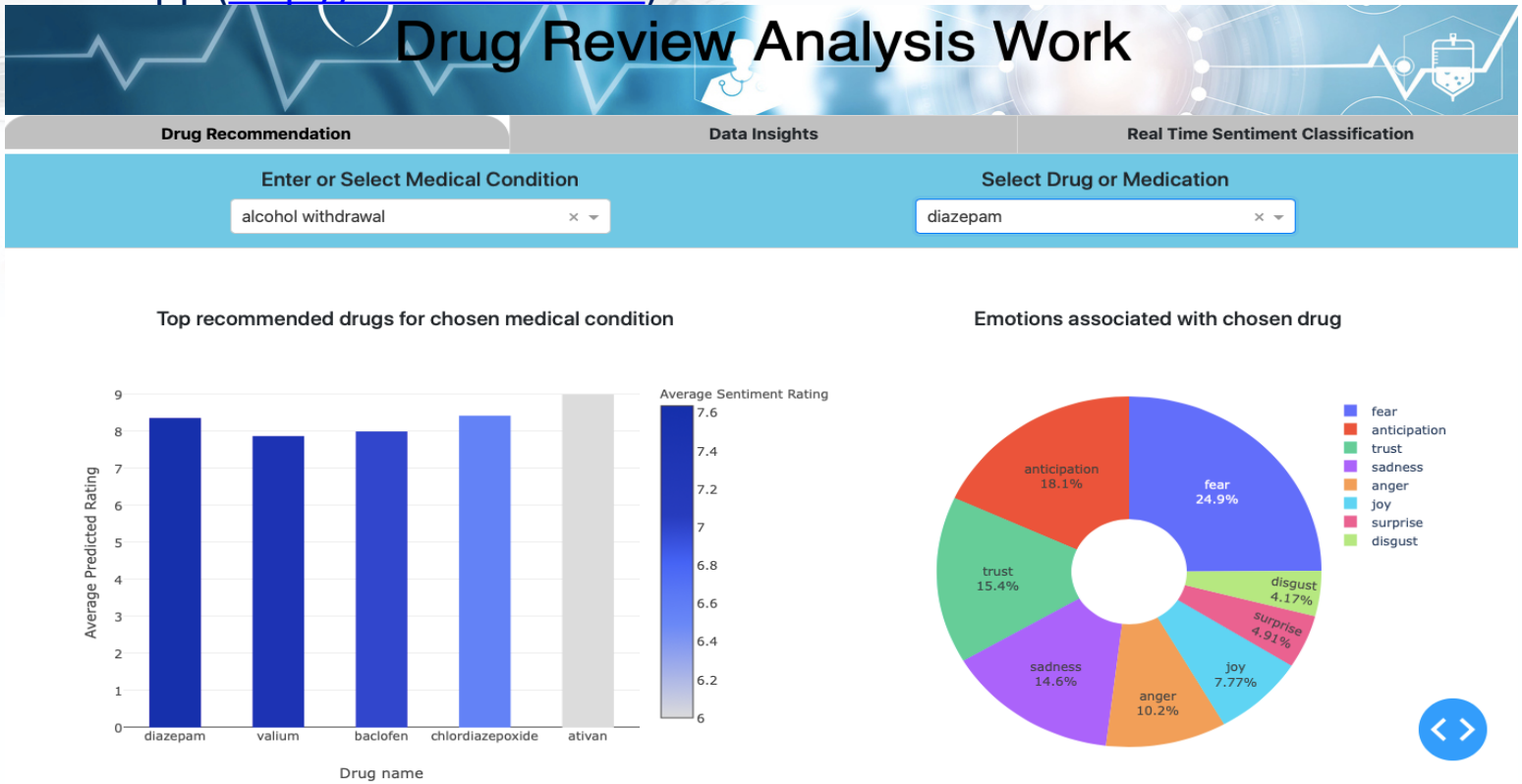
- There is no tool available in the market which provides consolidated information at a single place regarding medicines available for any conditions
- A unique solution for drug recommendation as well as awareness about side effects
- Sentiment analysis can be used by drug makers to obtain valuable patient opinion
- Instant online help about medications and symptoms.

Correctness of our results

- Sentiment analysis alone is not sufficient to provide correct review analysis. Therefore, we have used multiple predictive model results to accurately provide recommendation.
- Our webapp performs precise classification of emotions and side effects identification, which will not only help patients to self categorize but also pharma industry to inherit patient feedback

Data Product

- Web app (http://127.0.0.1:8050)



Learnings

- Developed various web scraping scripts: Scrapy (Spider), Beautiful Soup for data acquisition
- PostgreSQL: Storage and retrieval of data
- Text mining using NLP (Stopwords Removal, Tokenizing, Stemming)
- Sentiment and Emotion classification using NLTK libraries
- Training different ML and Deep Learning models and their deployment for real time analysis
- Web data visualization tools (Plotly, Dash, Wordclouds, Confusion Matrix, Bar Polar charts, t-SNE plots)
- App deployment on Heroku

Future Work

- A lot of information like cost, dosage, duration of condition and past medical history are hidden under reviews. We can build a robust Text Mining system to make personalized recommendations to users and provide in-depth analysis to pharmacies.
- Healthcare Chatbot
  - Online assistance/recommendation based on patient symptoms
  - Monitor emotional heath of patients through conversations
- Connecting to a medical information platform through API for real time data and visualizations
- Add a scheduled retraining of models to website for incorporating latest changes

Team: Data Pirates

- Akash Singh Kunwar: 301401851
- Rohan Harode: 301406504
- Shubham Malik: 301403562

thank you