



DRAW

Drug Review Analysis Work



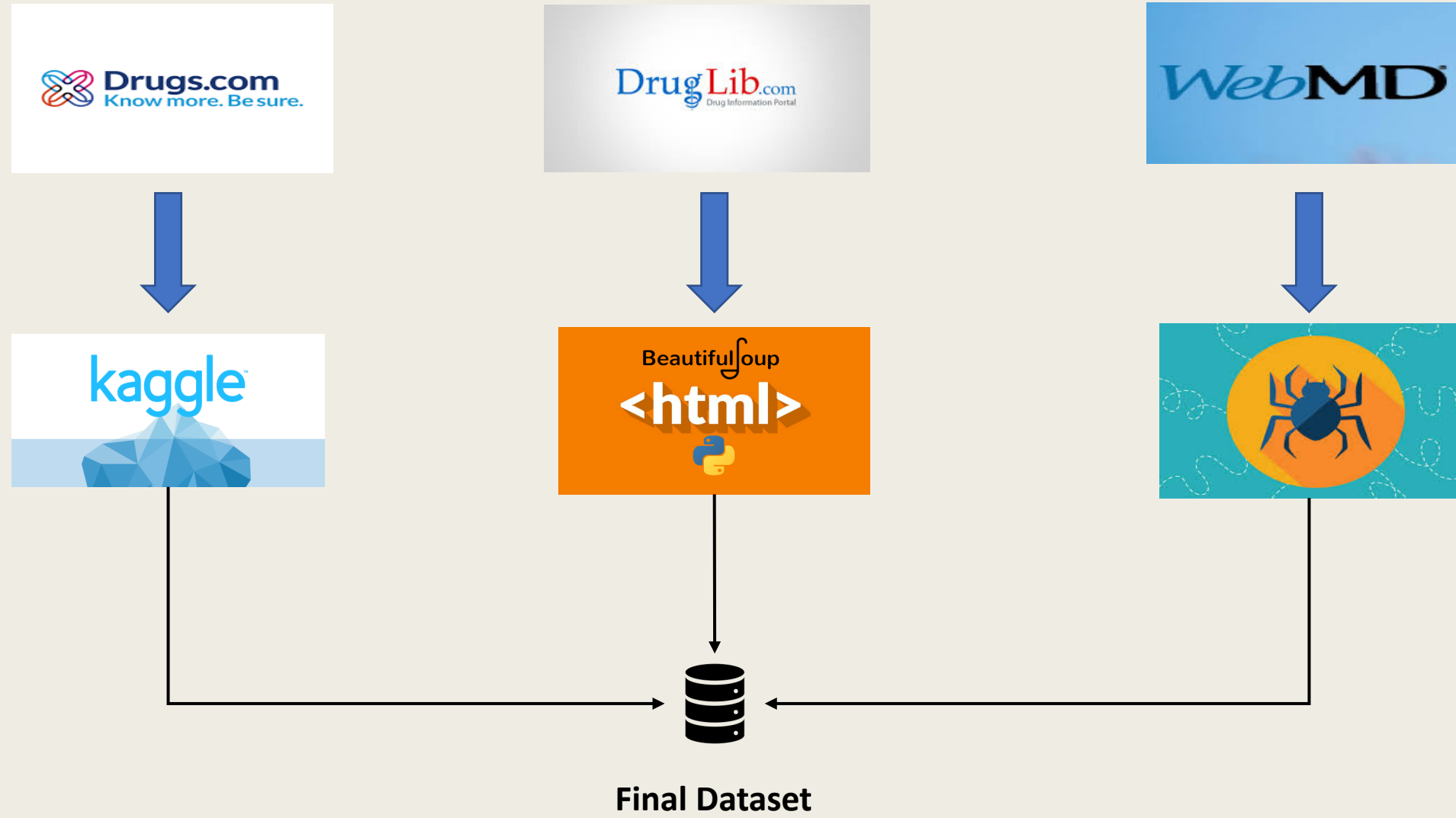
Motivation

- Adverse side effects may not be detected before the product launches due to various limitations in clinical trials.
- We can get easily overwhelmed by numerous reviews listed for symptoms and medicines available for any conditions.
- There is no tool available in the market which provides such consolidated information at a single place.

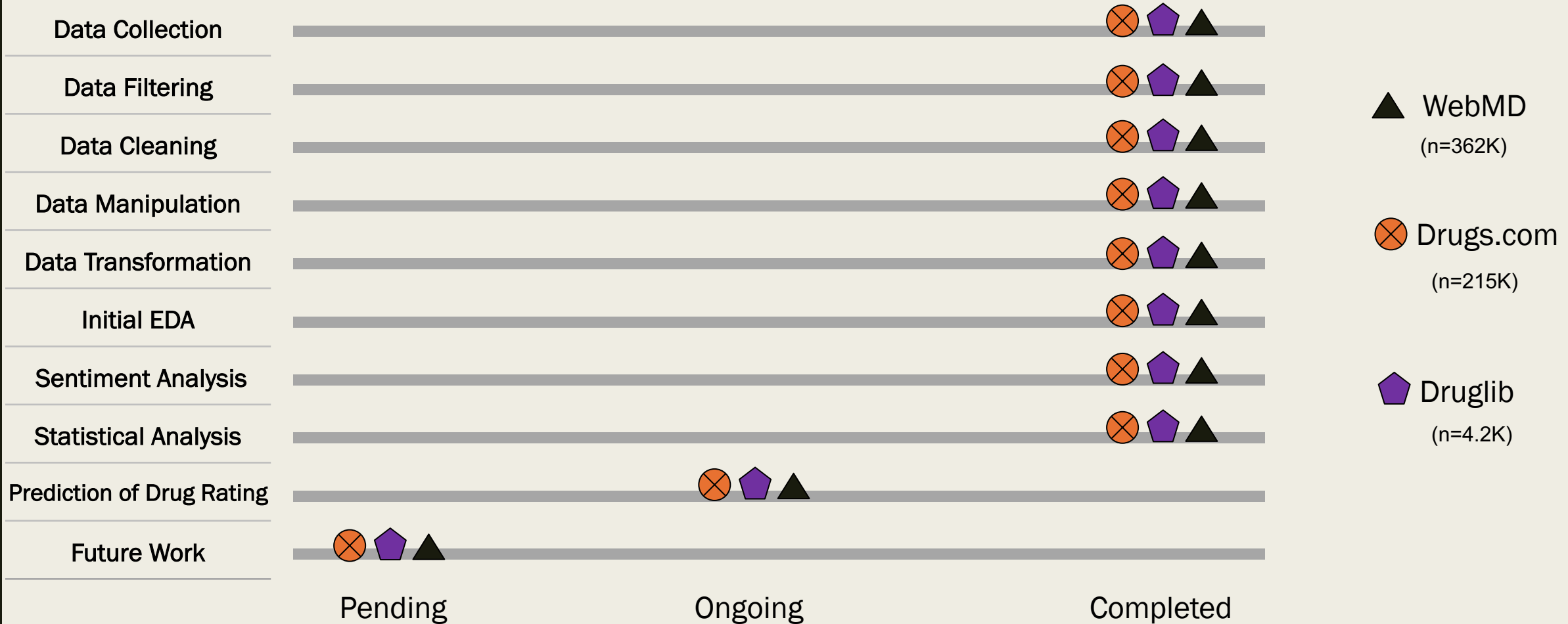
Why this project is challenging?

- Analysis of patient review data on their drug experience is a challenging problem, as it requires good understanding of medicine market, and critical brainstorming on feature engineering and data analysis.
- Reviews and opinion about drugs, treatments or side effects expressed on online media (Web 2.0 platforms like WebMD, Drugs.com, Druglib etc.) provides new opportunities for analysts to conduct aspect-based opinion mining and sentiment analysis.

Data Sources and Collection

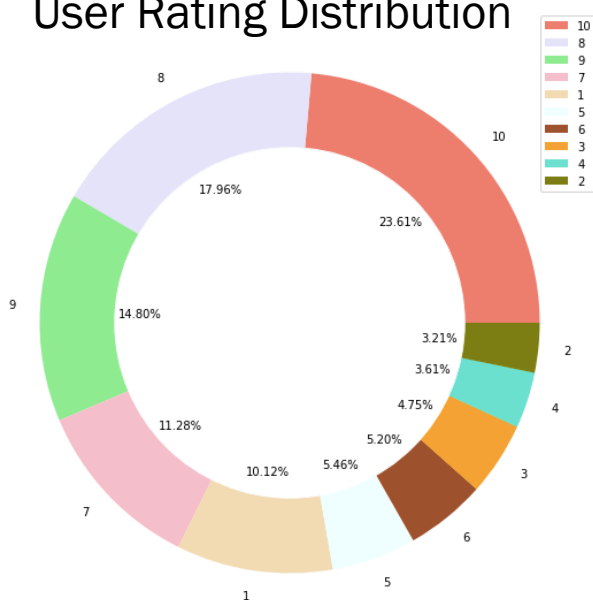


Current Status

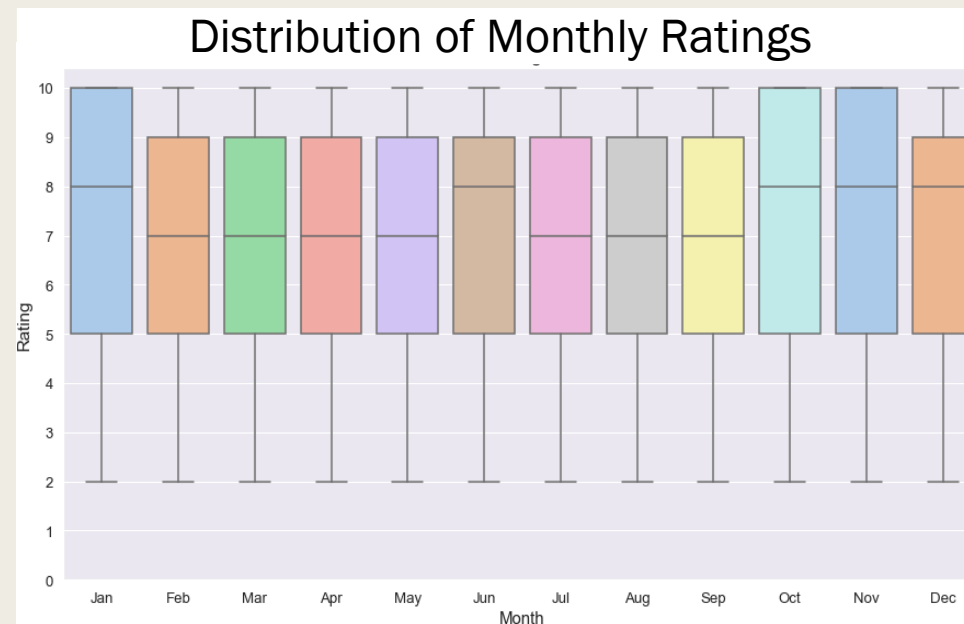


EDA Findings

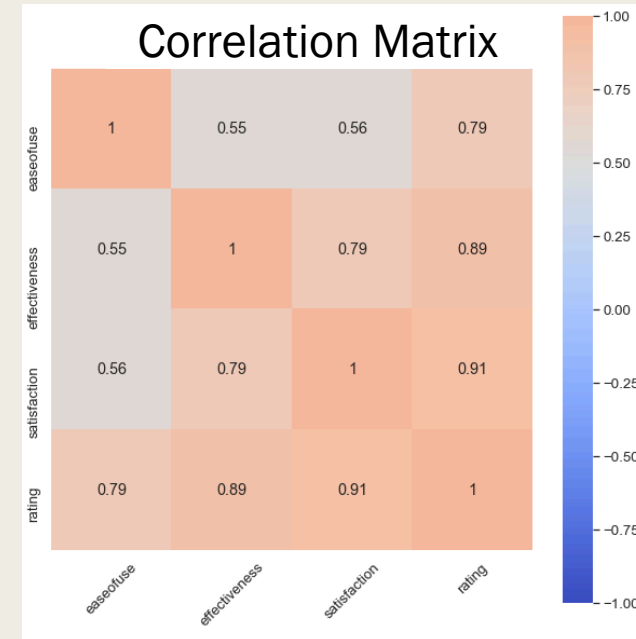
User Rating Distribution



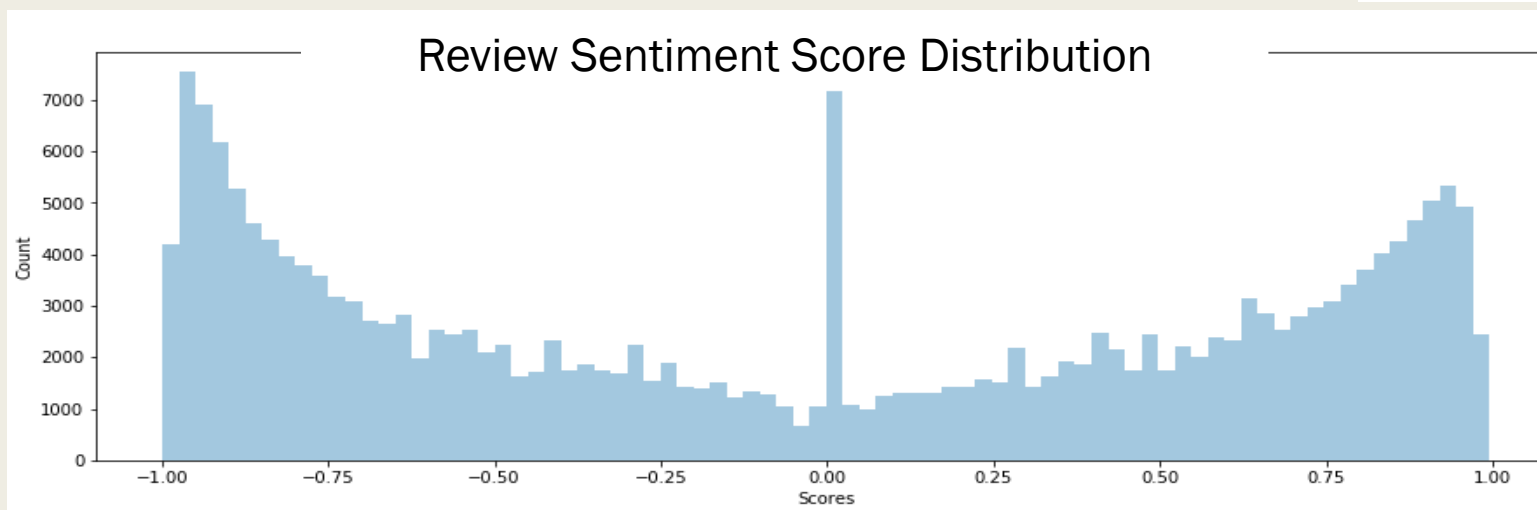
Distribution of Monthly Ratings



Correlation Matrix



Review Sentiment Score Distribution



Progress Report Timeline



We Are Here

Feb 14

Project Proposal

Finalizing project topic
Identifying Analyses
Decided Tech Stack

Feb 28

Data Pre-Processing

Data Filtering
Data Cleaning
Data Manipulation

Mar 13

EDA and Milestone

EDA on individual data sources
Database completion (merging dataset)
Statistical and Sentiment Analysis
Presentation/Video Creation

Feb 21

Data Collection

Finding Data Sources
WebScrapping Data
(WebMD, Drugs.com,
Druglib websites)

Mar 6

Data Transformation

Conditional Grouping
Jaccard Similarity
Fuzzy Matching

Till April 19

Future Work

Future Work

Mar 20

Machine Learning Techniques

Drug Rating Prediction
Classification of Sentiments using NLP
Model Evaluation

Apr 3

Data Analysis

Final Data Analysis
Data Visualizations

Apr 19

Report & Video

Video submission
Report/Medium post

Mar 27

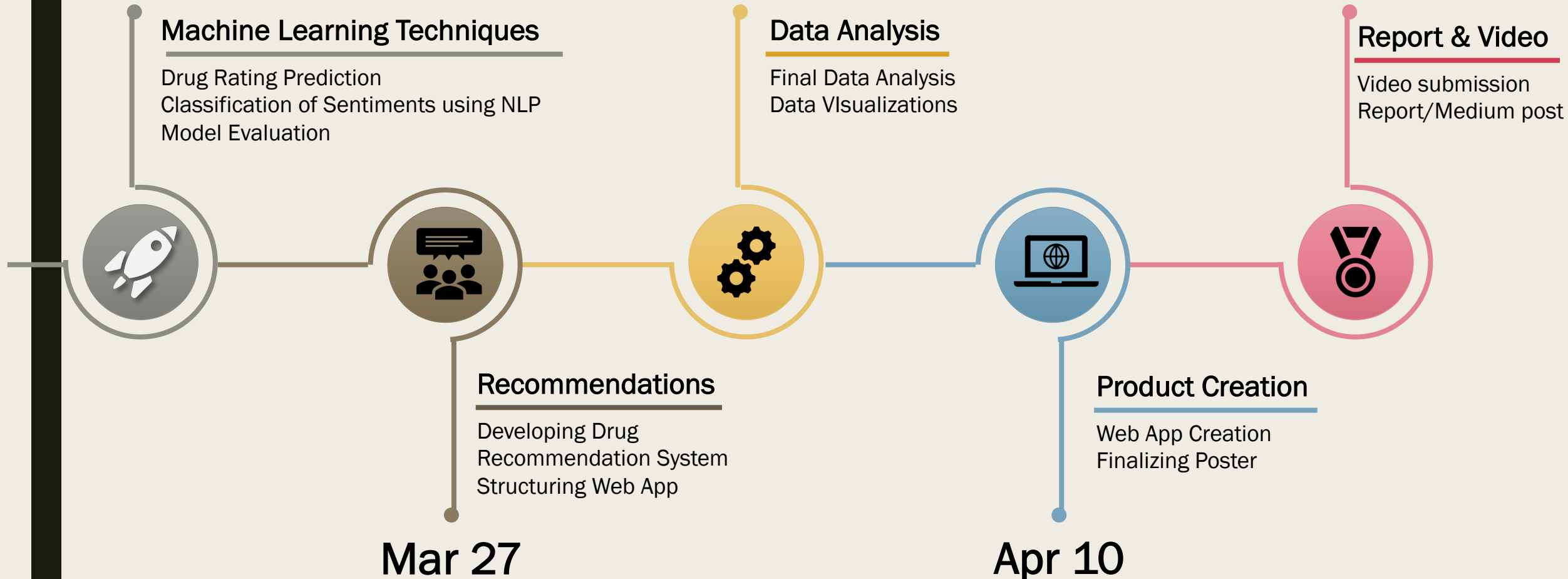
Recommendations

Developing Drug
Recommendation System
Structuring Web App

Apr 10

Product Creation

Web App Creation
Finalizing Poster



Risk Mitigation

IMBALANCED DATASET

Unstructured and imbalanced data increases analysis complexity

CONDITION MISMATCH

Drugs and conditions mismatch due to diverse data sources hinders data uniformity

Potential Risks

**ACCURACY & CORRECTNESS**

Building a drug recommendation system with higher accuracy and evaluating its correctness

LATENCY

Monitoring the latency of visualizations and results on the interactive web app

Data Pirates

- Akash Singh Kunwar
- Rohan Harode
- Shubham Malik

DATA PIRATES



Link to our repository-- <https://csil-git1.cs.surrey.sfu.ca/akunwar/drug-review-analysis-work>

Thank You!