**CLASSIFYING OUTCOMES BASED ON DRUG-REACTION COMBINATION**

**Project Proposal By: Team 5**

**Ankit Bhayani | Rajat Agrawal | Vishakha Sawant**

**INTRODUCTION**

"The FDA Adverse Event Reporting System (FAERS) is a database that contains information on adverse event and medication error reports submitted to FDA."

The FDA produces FDA Adverse Event Reporting System (FAERS) quarterly data. The FAERS include the following:

* demographic and administrative information and the initial report image ID number (if available);
* drug information from the case reports;
* reaction information from the reports;
* patient outcome information from the reports;
* information on the source of the reports;
* a "README" file containing a description of the files.

**Problem Addressed by the Project**

The problem addressed by this project is to classify the outcome of the drug reaction combination supplied to a patient based on the reported event. Since, we have different reported event by various organization we need to check whether the drug and reaction combination falls under following categories:

* Death
* Life-Threatening
* Hospitalization
* Initial or Prolonged
* Disability
* Congenital Anomaly
* Required Intervention
* Prevent Permanent Impairment/Damage
* Other Serious (Important Medical Event)

**Part 1: Parse files & Handling Missing data values**

**Link to Download data files:**

<http://www.nber.org/data/fda-adverse-event-reporting-system-faers-data.html>

The goal of this exercise is to extract “all” csv files present on the faers website save them without any manual intervention.

**Next step :**

• Handle missing data

• Compute summary metrics

• Check for any observable anomalies and outliers.

## **Part 2: Machine Learning: Multi-class Classification**

A Classification script in a Jupiter notebook in R/Python that builds a classification model for drug information from the case report / reaction information & patient outcome from the reports using downloaded dataset. Following steps would be covered:

* Variable selection
* Application of various ML Classification model
* Deploy the best algorithm/algorithms on Azure ML studio
* Create API(s)
* Deploy the model using web based front end
* Design of a pipeline and system to implement this approach and discussion on the system’s capabilities

## **Part 3: Summarization**

* Summarize the key insights related to different variables
* Present the results using python notebook and Tableau/Power BI

**Tools**:

* Python scikit-learn library for applying various classification model
* Jupyter Notebook
* Azure ML
* Tableau for Data Visualization
* Matplotlib library

Data sources: Publicly Available Datasets for Your Research

The data are provided in two distinct formats in the extract:

ASCII files, in which data elements are separated from each other by a '$' sign ("$ delimited"). Please refer to the ASC\_NTS.DOC file for additional information on this format.

XML files conforming to the guidelines of the International Conference on Harmonization (ICH) concerning transmission of Individual Case Safety Reports (ICSR). Please refer to the XML\_NTS.DOC file for additional information on this format.

<http://www.nber.org/data/fda-adverse-event-reporting-system-faers-data.html>

Pipeline: **Luigi**

Data Download from website

Preprocess and Missing Value Analysis

Transformation

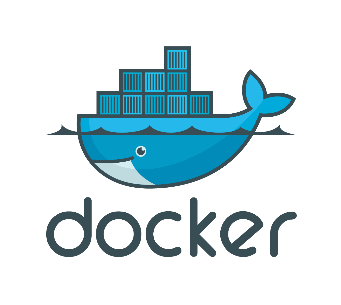
Upload to S3 or Azure Storage

Model Training

Classification

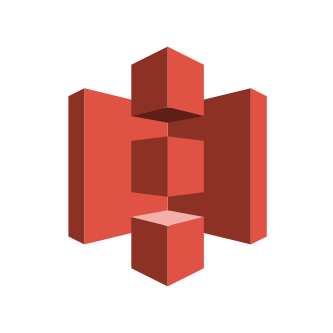
Test Results

# System Architecture



Processed Files

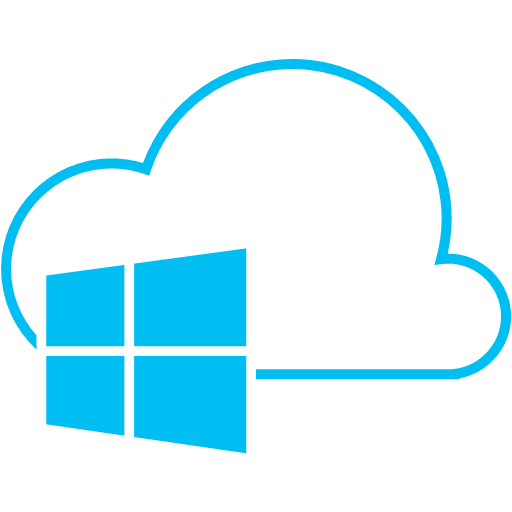
Will be uploaded to



AMAZON S3

Clean files used

for generating models

 publishesaccessed through 

Azure ML REST API

Studio

Youtube link : <https://www.youtube.com/watch?v=N8OrCHuN7BM>