**Prediction of World Economic Development Indicators**

**Project Proposal By: Team 6**

**Priyal Chaudhari | Prashant Vikram Singh**

**INTRODUCTION**

"World development indicators is a primary world bank collection of development indicators, compiled from officially recognized international resources."

The World Bank contains Catalog of international economic, financial, and socio-economic data sets from the World Bank. Among the most useful of the many datasets offered is [World Development Indicators](http://data.worldbank.org/data-catalog/world-development-indicators) featuring time series data from 1960 for 207 countries in the areas of population, labor, education, economics, the environment and much more.

**Problem Addressed by the Project**

* To measure the progress of Global development, members of United Nations proposed a list of 8 Millennium Development goals
* For example:
  + To achieve universal primary education
  + To eradicate extreme poverty and hunger
* To achieve these goals and to improve country’s overall GDP and development we need to understand the complex relation between these Millennium goals and Economic indicators
* By seeing the predicted values of these economic indicators, we can craft a plan to overcome difficult economic problems
* Our model will help user to predict value of world economic indicators for a specific country for a specific year

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

**Link to Download data files:**

<http://data.worldbank.org/data-catalog/world-development-indicators>

**Next step:**

• Handle missing data

• Compute summary metrics

• Check for any observable anomalies and outliers.

## **Part 2: Machine Learning: Prediction**

A Prediction model script in a Jupiter notebook in R/Python that builds a prediction model for world development indicators which will use the data from 1970 to current year for predicting the values for future year.

Machine learning prediction approach:

* Variable selection using various methods like k-best, forward, backward
* Application of various ML prediction models
* Study of time series forecasting using AR, MA, ARMA, ARIMA
* Evaluating the model based on parameters like MAE, MAPE, RMSE

Deploy the best forecasting model on Azure ML

Created Restful APIs

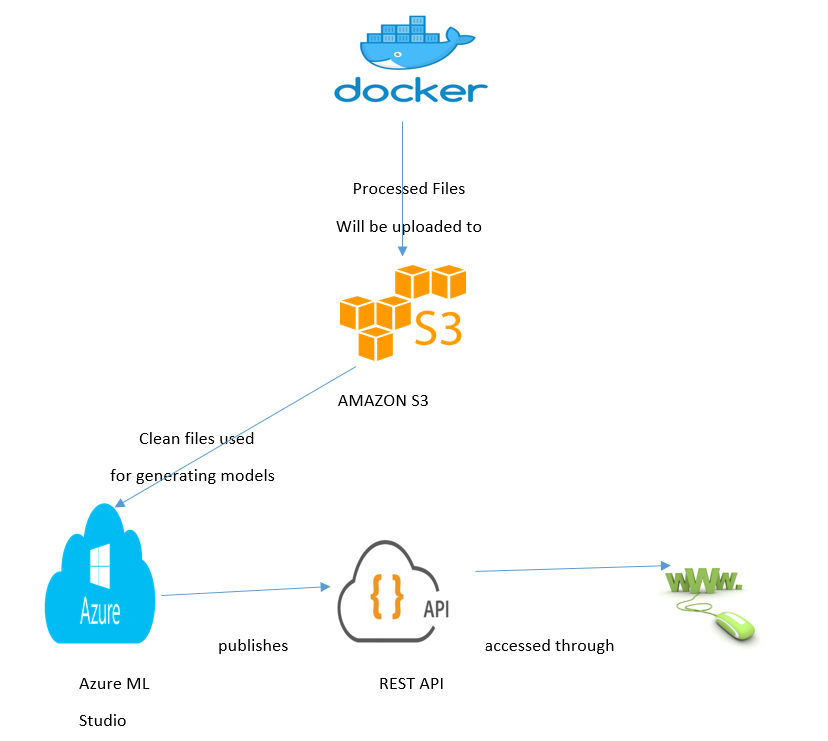
## **Part 3: Summarization**

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

**Tools**:

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

# System Architecture



# Proposal Link:

<https://youtu.be/jcBWAhXdfEA>