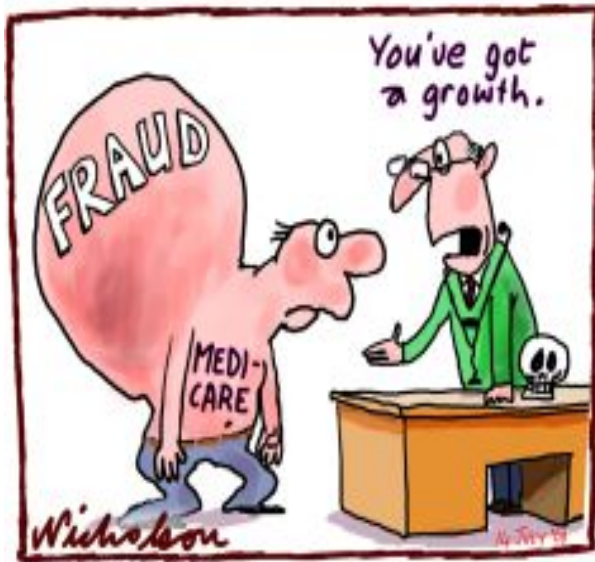# Fraud Detection in Medicare

*Presented by Shwetha Reddymalli*

INTRODUCTION

- Medicare fraud is an act of claiming reimbursement of healthcare services through Medicare to which one is not entitled. In USA we have Medicare Fraud Strike Force that deals with data resources from all the sources available like Federal, State and Local law and perform data analytics in order to eradicate or prevent fraud, waste and abuse.

- US health expenditure is about $3.6 trillion. Medicare spending increased to $750.2 billion which is almost 21% of total national health expenditure where in Fraud impact is estimated to be $80 billion which sums up to 10% of total Medicare funds.

PROBLEM STATEMENT

- Misrepresentation of Medicinal services is the primary issue that causes misfortunate incidents in the industry. We will be focusing on various fraud patterns involved in the industry such as fraud by service providers, insurance subscribers, insurance carriers and other conspiracy frauds.

- The main objective of this project is to build a Data model that helps in showing the connections between various datasets that are required for data analysis using anomaly analysis and geodemographic metrics.

- We will be considering various tools like **Tableau, Power BI, Spark and Jupyter**.

METHODOLOGY
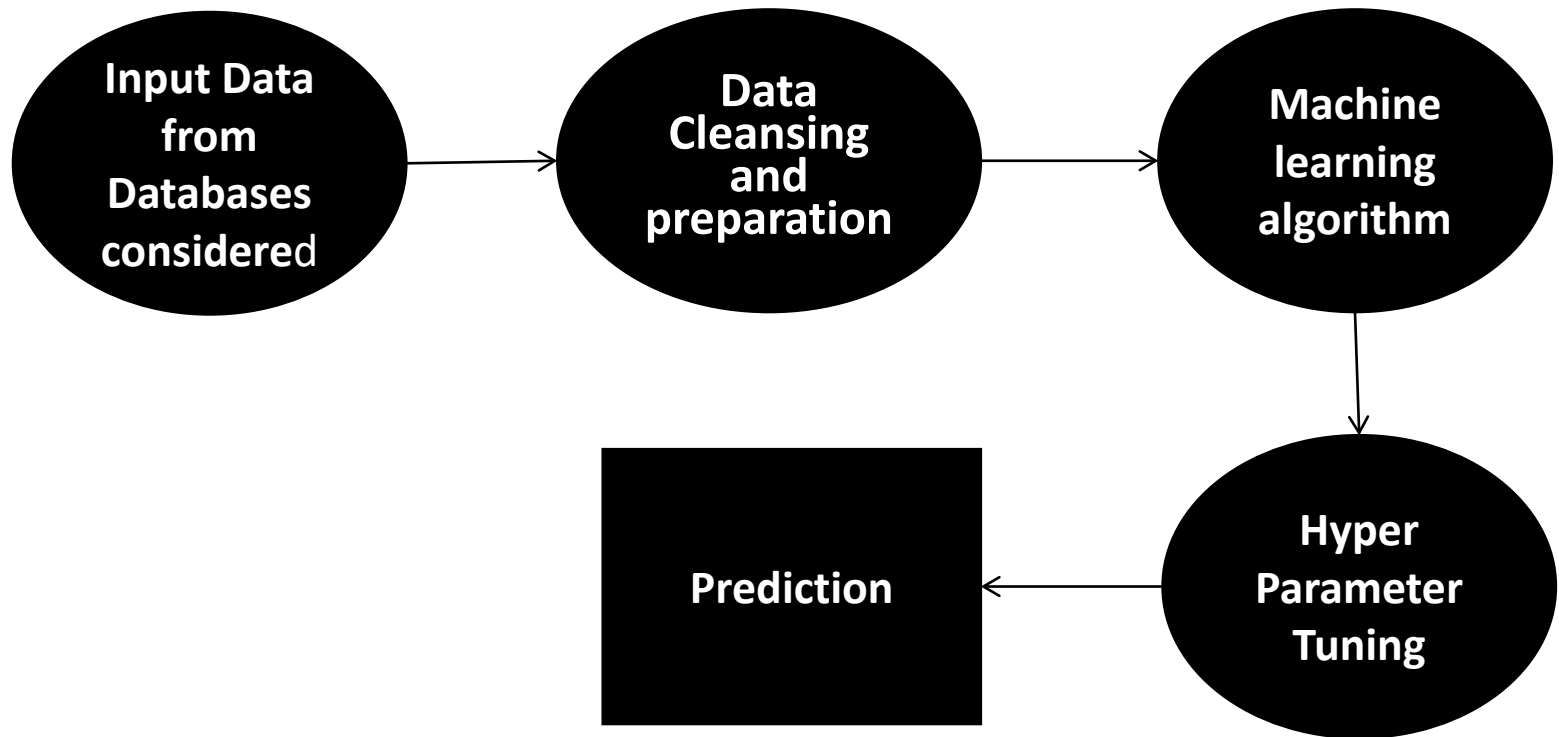
DATA SOURCES

We are considering multiple datasets, each having numerous rows and columns those contains details about physicians, drugs, payments and fraud details and the project idea is adopted from GitHub as a reference. These have huge data and are spirally interesting to work on.

- CMS Part D datasets: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber.html

- LEIE Datasets: https://oig.hhs.gov/exclusions/exclusions_list.asp

- FDA Datasets:
  https://www.fda.gov/Drugs/InformationOnDrugs/ucm079750.htm#collapseOne

SOLUTION OR FUTURE SCOPE

- Cross validation usage for sampling the data into train-test split.

- The main scope is to build a real time fraud detection pipeline using Machine learning algorithms.

- Hyper tuning parameter considered helps in increasing the overall performance of the algorithm adopted for analysis.

- We have considered the real-time data; hence the model needs to be readapted without a pause to the prediction service as we have continous interacting users.

THANK YOU