**Big Data Medicare Fraud Detection**

**Atharva Kousadikar, Debpriyo Roy, Harsh Sindhwa**

**Technology Used:**

* Azure HD Insight Cluster
* Apache Spark
* Pycharm
* Tableau
* Power BI

**Benefits:**

* Visualized Big Data on Tableau and Power BI to get powerful insights.
* Processed data cleaning and modelling on imbalance data through various processes.
* Used Azure HD Insight to collaborate and save work.
* Used Apache Spark on Virtual Machine (Cent OS 7) to process data faster and to create final table.

**Drawbacks:**

* Imbalance Data need more targeted values to make the model more accurate.

**Challenges:**

* Merging and applying joins on different Big data, as it takes time and sometimes might result in duplicate redundant values.

**Working Example:**

First, we took 3 datasets and analysed them sop that we get features from all the datasets. Then we uploaded that datasets on AZURE HD INSIGHT for to collaborate and save our work. We then did Data cleaning and processing on Apache Spark on Virtual Machine to create Final Table for Data modelling and then we did Visualization on Power BI to gain the Insights.