Intel® oneAPI Hackathon for Open Innovation

Theme: **Open Innovation with Intel® oneAPI AI Analytics Toolkits**

Team Name: Blazer

Team Members: Ayesha Irfhana M

Theme Chosen and Motivation

Theme:-Open Innovation with Intel® oneAPI AI Analytics Toolkits-Insurance Fraud detection

Introduction:-

- Many use pricey GPUs to shorten training and inference times for machine learning on industrial datasets that are getting bigger and bigger.
- Our project's demonstration will demonstrate how you may speed up your machine learning workflow utilising Intel's optimised libraries, which are both affordable and streamlined.
- The claimant or the insurer may engage in a variety of illegal and dishonest practices in order to obtain favourable benefits. Insurance fraud is estimated to cost consumers several billion dollars annually.

Proposed Tech Architecture (Highlight usage of oneAPI data analytics toolkit)

- In this project,I have demonstrated the results of research by automating the evaluation of insurance claims using a variety of data methodologies.

IntelOneAPI toolkits:-

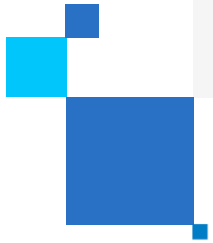
- This reference notebook will show you how to use multiple accelerated Python libraries included in the Intel® AI Analytics Toolkit to improve your ML workflow's training cycles, prediction throughput, and accuracy (AI Kit).
- The principal libraries we'll use in this notebook are:

1.Intel® Distribution of Modin*
2. Intel® Extension for Scikit-learn*
3. XGBoost Optimized for Intel® Architecture
4.Intel® Daal4py

Impact of inclusion of oneAPI

Hardware Requirements:-

- I have used 5th generation of Intel(R) Xeon(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz
- **The Intel® Distribution of Modin:-**It is used to process and study the data first. With support for datasets from 1 MB to 1 TB+, the Intel® Distribution of Modin is a distributed DataFrame library created to expand your pandas process naturally with the size of your dataset.
- **Intel® Extension for Scikit-learn:**-For Intel® CPUs and GPUs in single- and multi-node setups, the Intel® Extension for Scikit-learn delivers seamless integration with scikit-learn applications while shortening algorithm run times.
- **Intel® Daal4py library:-**With the help of the get gbt model from xgboost() method, it is quite simple to convert a tuned XGBoost model to Daal4py in just one line of code. after using Intel® Daal4py library,the execution time is lesser than before.It gives **predictions with better time performances.**



intel®

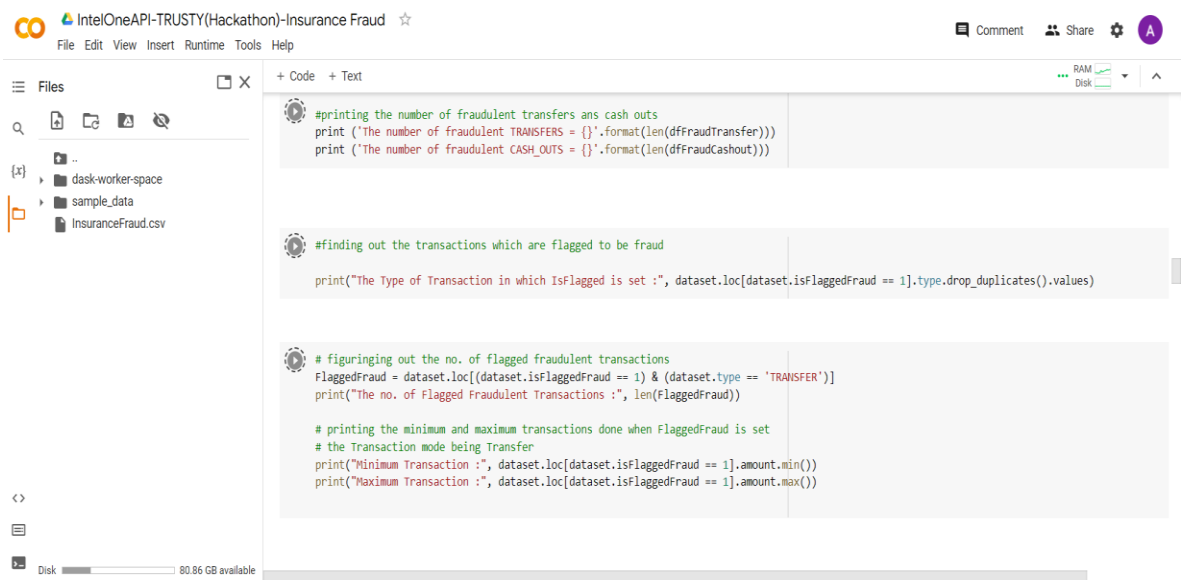
Innovation Quotient and Scalability

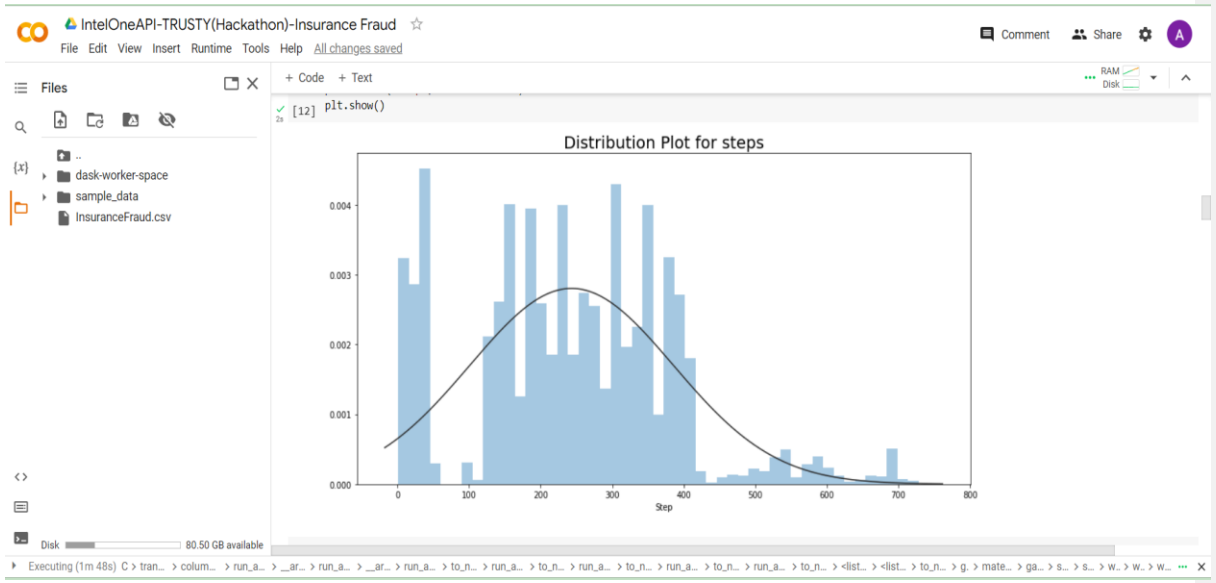
- Additionally, the system may be able to create heuristics for fraud signs. As a result, this strategy has a favourable effect on the entire insurance business by boosting both firm reputation and customer happiness.
- The dataset which I took was downloaded from kaggle (InsuranceFraud). For further reference of the dataset visit my github**.(Download the dataset fully and run the cells in colaboratory).**
- We can reduce losses for the insurance business by developing a model that can categorize insurance fraud.

Technologies used and screenshots:-

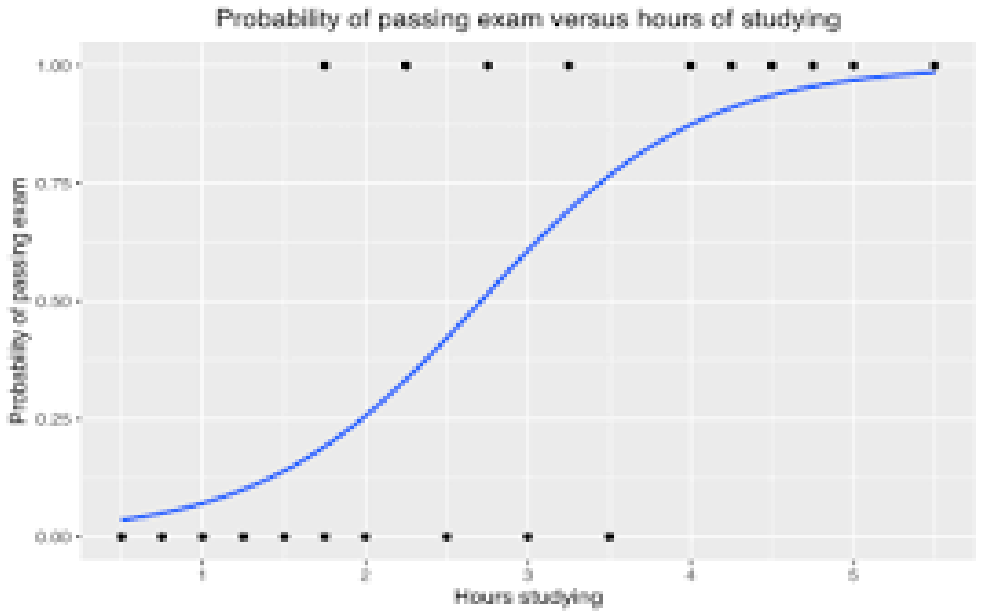
- It uses methods based on the XGBoost algorithm, an extreme gradient boosting algorithm, to spot fraudulent claims in the auto sector. To clean, explore, and extract pertinent features from the provided data, a variety of data analysis techniques including data cleansing, data exploration, and privacy-preserving are also utilised.

- Techniques such as **Logistic Regression,XGboosting and K-Means clustering** are used.

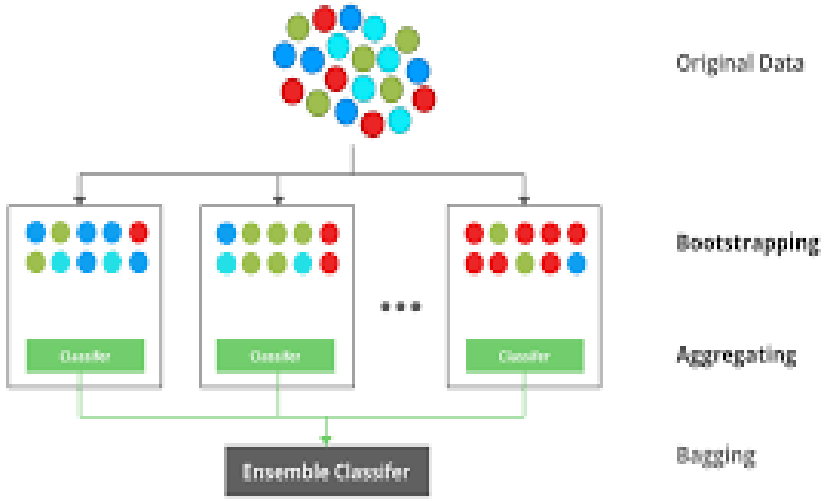




Machine Learning Algorithms

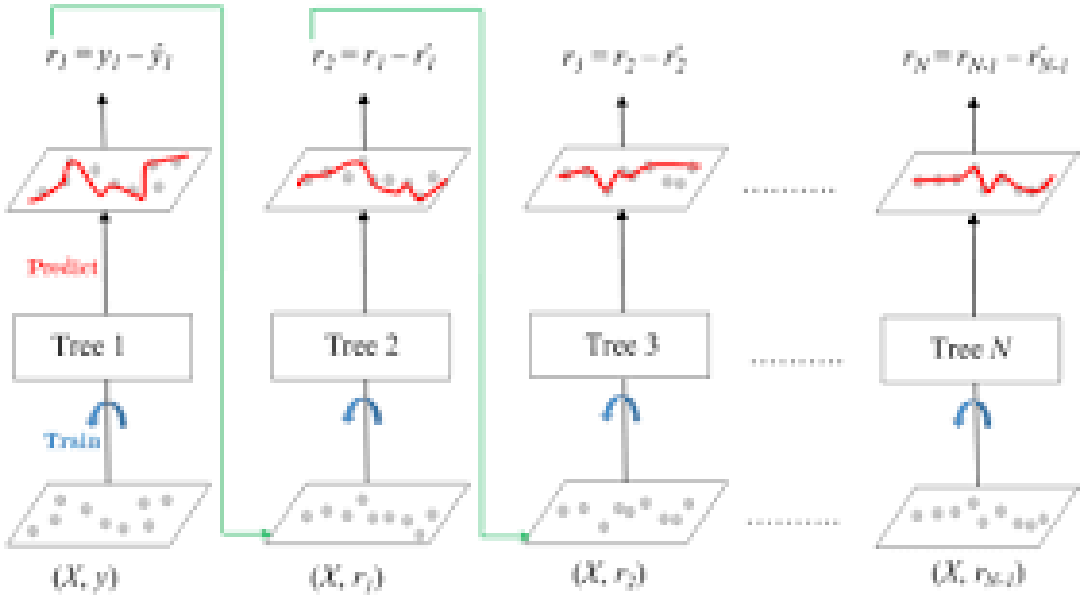


Logistic Regression

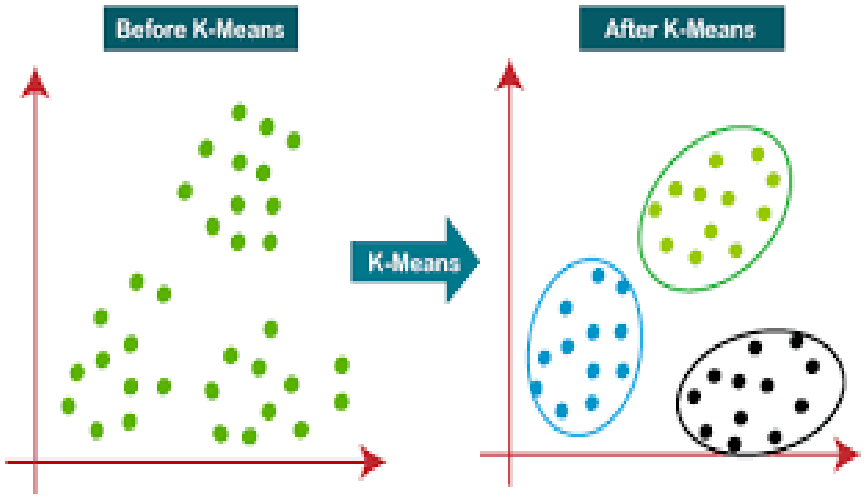


XGBoosting

Machine Learning Algorithms



Linear Regression



K-means clustering

Quick Summary

- *To quickly and accurately obtain fraud claims, we created an automated fraud detection application framework based on machine learning and XGBoost algorithms.*

- To clean, validate, and extract the pertinent features from given data, several data analysis techniques are utilised, such as data validation, data insertion, data preprocessing, and clustering.

- Insurance fraud reportedly costs consumers several billion dollars, according to recent reports. Therefore our project has an appropriate method that can identify potential frauds with high precision, accuracy, and is efficient.

Video and demo submission

Video Link:-

Demo link:- https://colab.research.google.com/drive/1DQ3W9pAem-shT8-fxhZTgfM8Tdl4L4Tk?usp=sharing


intel