**Insurance Claim Detector**

**Overview**

Insurance Claim Detector is used by the insurance company to decide whether the claimed insurance is **Fraud** or **Genuine**. It is an insurance company which provides insurance to Hospitals, depending upon those hospital's details. This model can be used in such a way that by entering hospital details the model will predict whether that particular hospital which claimed for insurance is Fraud or Genuine.

**Technical Aspect**

This project is divided in to three parts:

**1. Pre-Processing:**

**a. EDA:**

- Import Libraries

- Import data

- Understanding data using Visualizations

- Relationship between data if any using correlation

**b. Feature Engineering:**

- Dealing with Missing data

- Removing Duplicates

- Categorical Features Encoding

- Normalization of Features

**c. Feature Selection:**

This can be done using different methods such as Decision Tree classifier, Random forest classifier, RFE with random forest classifier, Extra Trees classifier. In this project i have used Extra Trees Classifier to find the top features that are effecting the result much. Using domain knowledge and Feature selection we have dropped some of features which are not impacting much.

**d. Imbalanced data treatment:**

Now when we have observed our data is imbalanced, we can say this because it has 75% of Genuine data and 25% of Fraud data which may cause biasness in the model towards Genuine case as these are more. So it is compulsory to balance the data so i have balanced data using Random Over Sampling. Now my entire data is ready for model building. Now split the data in to train and test with trained data as 80% and test data as 20%.

**2. Model Building:**

- Model is built on this train data using Random Forest Classifier and Validation is done on test data and achieved an accuracy of 90.23% and F1-score as 90%.

- Model generated the output for the insurance data set in which 159054 claims are Genuine(0) among 156389 actual Genuine’s and 159054 claims are Fraudulent(1) among 156487 actual Fraudulent’s.

**3. Model Deployment:**

- Model Deployment is done using Flask.

**Installation**

The Code is written in Python 3.7.