**Insurance Claims- Fraud Detection**

**1.0.Introduction:**

Insurance fraud refers to a variety of unethical behaviours that a person may engage in ,

in order to obtain a favourable outcome from an insurance company. Given the variety of fraud types and the low ratio of known frauds in typical samples, detecting insurance fraud is a difficult task. When developing detection models, the cost of false warnings must be balanced against the cost of loss avoidance. Not all claims are analysed thoroughly.

* 1. **Machine learning in Fraud detection**

Artificial Intelligence includes Machine Learning (AI). Artificial Intelligence's goal is to construct a computerised system that can perform complicated analysis and not only replace but also improve on human input. Machine Learning uses artificial intelligence to “give” systems the ability to learn and grow from their experiences without the need for additional programming. Fraud detection is a knowledge-intensive process that determines if a transaction or claim is authentic or fraudulent. Deep anomaly detection is a common type of machine learning used in the insurance sector. Anomaly detection works by examining the customer's normal, authentic claims and creating a model of how a typical claim appears. The model is then used to analyse big datasets. This method of anomaly detection can also be used to build other components of artificial intelligence. Predictive analytics is one such , that can be used to design the program.This program not only design and analyse the features it also identify the fraudulent claims.

**2.0.Objective**

The techniques in the machine learning is to improve the accuracy of detection on various imbalanced datasets. With the goal of generating higher predictive performance, the impact of feature engineering, feature selection, and parameter tweaking is also analyzed.

Predictive analysis involves certain steps to obtain accuracy detection

* Training
* Testing
* Validation

All this steps are involved along with the algorithm on the partial training datasets and then it is tested on the test datasets, finally it is then examined by some random splits .The datas in the datasets are handled by certain rules.

**3. Summary**

The difficulty with machine learning fraud detection is that frauds are significantly less frequently than legitimate insurance claims. This type of datas are called imbalanced data. Model is proposed to find accuracy of fraud detection on different algorithm to handle those imbalanced data. The best out of that are fitted in to the model to predict the accuracy.

The final model fitted DecisionTree which obtained 81.33% score and its f1score is 86% which better than the KNN classifier .

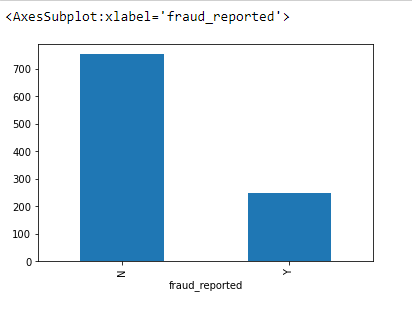
**3.1.About Dataset**

The dataset is about the auto claim insurance dataset along with the customer details for which they have claimed their insurance. Model is predicted whether the claim is authentic or fraud

**4. Data Handling**

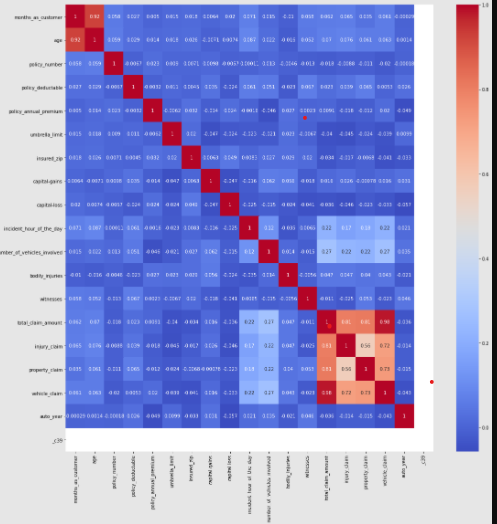
**4.1.Exploratory Data Analysis**

From figure it is clear that there is imbalance problem number of count is less than the no of cases reported fraud



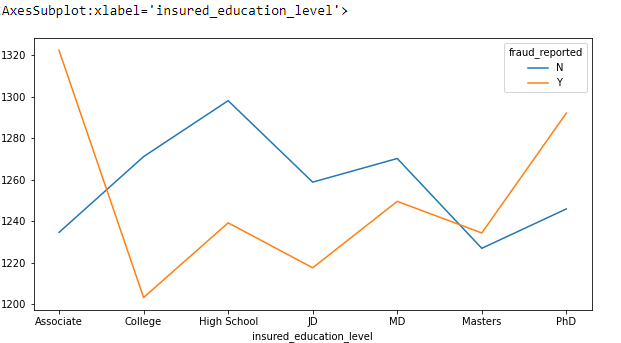
**4.2.Correlation between variables**

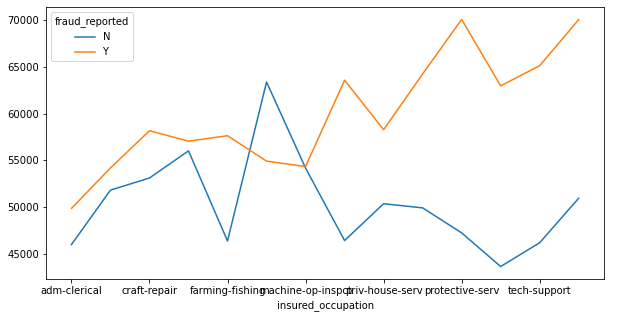
**Correlation between some variables are carried out and heatmap was plotted according to the map some variables which are not much correlated are dropped from the columns**

****

**5.Data Visualization**

**Counts of the variables are grouped and then they are plotted .Some of the plots**

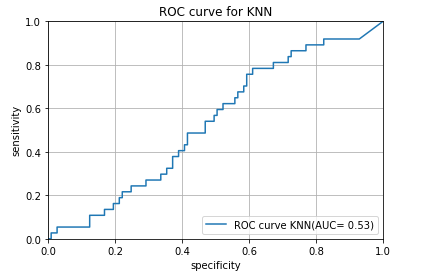
****

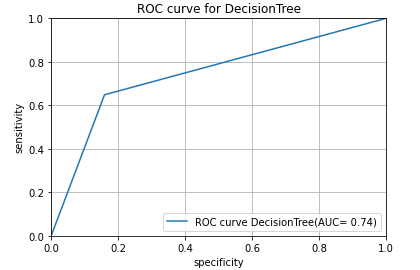
**6.Modelling**

Two different types model are used in this dataset

* KNN(K-NearestNeighbors)
* DecisionTree

Hyperparameter tuning and selection was done for the models using GridSearchCV.after 10 fold the model is fitted in to the training model and tested along with accuracy score,f1score, precision and recall value .

****

****

**Conclusion**

Fraud accounted for between 15 percent and 17 percent of total claims payments for auto insurance bodily injury in 2012, according to an Insurance Research Council (IRC) study. The study estimated that between $5.6 billion and $7.7 billion was fraudulently added to paid claims for auto insurance bodily injury payments in 2012, compared with a range of $4.3 billion to $5.8 billion in 2002.

This model is built to detect the auto fraud claimin doing so number of frauds are less and it has reduced the loss for business and insurance companies since it showed the accuracy of authentic insurance claim.