**Insurance Fraud Detection ProblemSet**

A large number of problems in data mining are related to fraud detection. Fraud is a common problem in auto insurance claims, health insurance claims, credit card transactions, financial transactions and so on.

The data in this particular case comes from an actual auto insurance company. Each record represents an insurance claim. The last column in the table tells you whether the claim was fraudulent or not.

Application Area- Insurance, Fraud Detection

Data Mining Task - Classification Model-Built Decision Tree and Random Forest Model

Number of Instances- 15,900

Number of Attributes- 31

Fraud Detection Model Steps;-

1. Initially, the dataset was divided into Train and Test data.
2. After excluding the target variable from both training and testing dataset, I performed

one-hot encoding to transform categorical data into numerical data for further analysis.

3) After building a Decision Tree Classification Model and performing hyperparameter

tuning and Model tuning, I was able to achieve accuracy of 81%.

4) After building a Random Forest Model and performed hyperparameter tuning, used

Randomized search and cross validation, I was able to achieve accuracy of 86%.

More details about which parameters are used and what are the results is added in the

presentation.