# Predicting the “Fraud in auto insurance claims” & Pattern extraction

## Problem Description

A major general insurance company has a business problem with the significant number of claims being reported are fraudulent in nature and it is leading to leakages. So, the Insurer decided to predict the fraudulent ones before even processing the claims to allocate costs appropriately, keep the thorough investigation process in place and design proper action plans for the claims etc.

Insurance fraud refers to any claim with the intent to obtain an improper payment from an insurer. Motor and health insurance are the two prominent segments that have seen a spurt in fraud. Frauds can be classified from a source and/or nature point of view.

Sources can be policyholder, intermediary and/or internal with the latter two being more critical from internal control framework point of view. Frauds can be classified into nature wise, for example, application, inflation, identity, fabrication, staged/contrived/induced accidents etc.

Fraud affects the lives of innocent people as well as the insurance industry and thus it may be of interest to the health of the Insurance Industry and Society. In fact, Insurers report certain classified cases to Regulators and Law enforcement agencies like Police, Crime Bureaus and others as mandated by the Regulators/Government and required by Law. With the advent of organized gangs and/or collusion, the problem has become more complex and sophisticated, and the frauds have been difficult to detect and to prove, if detected.

The framework of prediction of fraud and pattern extraction will be useful for insurance companies, regulatory bodies, intelligence departments etc.

Prediction at the time of processing claims will reduce costs and minimize losses.
The intelligence arising out of ever-improving prediction algorithms will help to retrofit in terms of improvement of the underwriting process, the exercise of a good selection of policyholders based on identified profile attributes, strengthening of internal risk management mechanisms and finally, clear guidance and communication to employees and other stakeholders involved.
At the Industry level, the shared aggregate information helps build appropriate intelligence and resilience while paving the way for collective effort for prevention as well as minimizing losses and matching the efforts of perpetrators.

At the Regulator and Law enforcement level, the intelligence arising out of prediction will help revamp the Regulations/Laws and plan not only enforcement but Industry based initiatives/systems for resilience and to share information for the consumption of the Industry and the Society.
Prediction at the time of processing claims will reduce costs and minimize losses for the insurance company. Hence, the prediction of fraud plays a very important role in auto insurance claims. The company wants to understand the hidden patterns in the data which lead to the construction of the investigation process and the claim settlement decision.

Beyond building a model to predict fraud you will have to identify the patterns for fraud which will help in-turn to the company to take action accordingly to initiate the investigations on claim classes to identify if fraud exists and also to handle the fraudulent cases while settling the claims.

You are expected to create an analytical and modelling framework to predict the fraud in auto insurance claims based on the demographic, policy, claim, and vehicle-related features provided in the datasets and generate the top 20 patterns for fraud on target attributes using the decision tree algorithms only while answering other questions too cited below.

Objectives
You are expected.
1. To do exploratory Data Analysis using visualizations
2. To report the results/observations from learning curves
3. To build the analytical framework to predict if a claim is fraudulent or not
4. To extract the top 20 patterns for fraudulent claims