**Capstone Two: Project Proposal**

# Vehicle Insurance Claim Fraud Detection

Apparently, it's a real database of an American insurance company Vehicle insurance fraud involves conspiring to make false or exaggerated claims involving property damage or personal injuries following an accident. Our objective is to classify claims as fraudulent or legitimate.

Some common examples include staged accidents where fraudsters deliberately “arrange” for accidents to occur; the use of phantom passengers where people who were not even at the scene of the accident claim to have suffered grievous injury, and make false personal injury claims where personal injuries are grossly exaggerated.

Problem statement formation

* Auto Insurance fraud is a major issue in the auto insurance industry. It is essential to detect and prevent fraud to ensure the industry's integrity and fair treatment of policyholders. In the United States alone, it is estimated that around 8,898 cars were intentionally set on fire in 2020. Car insurance scams cost companies around $29 billion every year, causing significant damage to the industry.

Context

* According to the Insurance Information Institute (III), insurance fraud is a deception committed against an insurance company for financial gain. Insurance fraud detection is a challenging problem, given the variety of fraud patterns and relatively small ratio of known frauds in typical samples. In 2020, an estimated 8,898 cars were intentionally set on fire in the US. Insurance companies rake in the dough, but they spend a lot of money fighting fraud, too. Car insurance scams cost companies hourly labor, investigations, bureaucratic costs, and other complicated and expensive work. This is then passed on to customers in the form of increased premiums. Car insurance scams damage companies to $29 billion per year. While building detection models, the savings from loss prevention need to be balanced with the cost of false alerts. Machine learning techniques allow for improving predictive accuracy, enabling loss control units to achieve higher coverage with low false positive rates.

Criteria for success

* The target is to detect if a claim application is fraudulent or not.
* We want to shift through our recent automobile insurance claims looking for anomalies that are suspicious and potentially fraudulent insurance claims.
* We will use machine learning algorithms and apply models to this data to identify the most suspicious claims

Scope of solution space

* Machine learning algorithms can be trained to recognize unusual behavior or trends in claims data, policyholder information, and other relevant data.
* Predictive modeling involves using statistical techniques to predict the likelihood of an event occurring. We can use this approach to assess the probability of a claim being fraudulent based on historical data.

Constraints

* Despite these measures, it's important to note that fraud detection is an ongoing challenge, and new tactics may emerge over time.
* Insurance companies continually adapt their strategies to stay ahead of evolving fraudulent activities and protect the interests of both insurers and policyholders.
* Fraud detection is a complex and evolving field. Regularly reassessing and updating detection strategies based on emerging patterns and industry trends is essential for continued success in mitigating fraudulent activities.

Data sources

* Dataset contains vehicle dataset contains 33 columns and 15420 rows of vehicle dataset - attribute, model, accident details, etc along with policy details - policy type, tenure, etc. The target is to detect if a claim application is fraudulent or not - FraudFound\_P
* [Dataset.csv](https://www.kaggle.com/datasets/shivamb/vehicle-claim-fraud-detection)