Car Insurance Fraud Detection Based on Existing Data

Context:

Car insurance fraud can be difficult to detect due to the wide range of fraudulent activities that can occur and because they are far less common than legitimate claims. Fraud can be as simple as lying about a residential address or as extreme as faking an accident. Fraudulent claims are unethical and cause significant revenue loss for insurance companies each year. A predictive model for determining if claims are fraudulent will give insurance companies more confidence in pursuing a challenge to the claim and hopefully preventing more loss from these individuals in the future.

Criteria for success:

The objective is to create a predictive model to detect car insurance fraud to significantly reduce revenue loss. The model should be able to analyze if a claim is fraud or not accurately.

Scope of solution space:

In order to predict fraud, an existing dataset of car insurance claims will be used as fraudulent claims are identified as well as many details about the claim pertaining to the individual who filed.

Constraints:

The claims that are classified as fraudulent in the dataset do not provide the type of fraud that was committed. Additionally, looking at one existing dataset may be limiting.

Stakeholders:

Insurance company executives

Data sources:

Publicly available car insurance claim dataset with fraudulent claims identified. Downloaded from Kaggle