

PREDICTING THE TOTAL AMOUNT OF AUTO INSURANCE POLICY CLAIM

Problem Statement

A car insurance policy can include several types of coverage, which depend on the type of plan you purchase, such as bodily injury and property damage. An insurance claim is a request made to an insurance company for compensation. An insurance claim commences an inquiry to determine the amount of payment, which the insurance company will pay on behalf of the customer; the coverage may involve attorney fees, hit-and-run costs and other legal fees etc. . In this project, I will investigate the significant predictive variables for the response variable “claim amount”. The model will predict the total amount of insurance policy claims.

Dataset

The data set downloaded from Emcien data repository and is in a comma-separated value file format.

(http://dyzz9obi78pm5.cloudfront.net/app/image/id/560ec66d32131c9409f2ba54/n/Auto_Insurance_Claims_Sample.csv)

This data set includes the records of the Auto Insurance claims for years 2011 to 2019; there are 9134 records in the data set; each row of the file is a unique policyholder and customer comprising of 25 features excluding the customer ID.

Scope of the Project

The goal of the project is to develop a model to predict the total claim amount for an auto insurance claim. Predicting the claim cost is a fundamental issue because it policy premiums are contingent on this prediction. Since the policy premiums are among the most important decision variables for customer, an optimized policy premium becomes important to keep the existing customer and to appeal the ones. .

; hence, auto insurance companies are the potential stakeholders of this project.

Technical Approach

➤ **Method:** This study will analyze the potential predictive variables for the response variable of insurance total claim.

➤ **Working with Data:**

After cleaning and wrangling the data, the project will focus on the following:

1. Related with the car such as vehicle type, and vehicle class
2. Associated with the customer such as gender, and education
3. Connected with policy plans such as policy type and sales channel

The project will be capitalized on the EDA techniques to analyze the data. To be able to build a predictive model, a regression model will be implemented along with machine learning techniques.

Project Outcomes

Project outcomes will include a jupyter notebook containing code, a report, and the slide deck.