Prediction of Home Insurance premium price

Group No: 11

Student Names: Rajendra Kumar Rajkumar
KV Vishagan Renganathan

Background Information and Introduction

Homeowners insurance is a form of property insurance that covers losses and damages to an individual's house and to assets in the home. Homeowners insurance also provides liability coverage against accidents in the home or on the property. A customer can get information about the policy premium price, which he needs to pay for getting his/her property covered either directly from the insurance company or from insurance aggregator websites. Because an insurance company is a business, it is obvious that the rate charged must cover losses and expenses, and earn some profit. But to be competitive, insurance companies must also offer the lowest premium for a given coverage. Moreover, all states in the United States have laws that regulate what insurance companies can charge, and thus, both business and regulatory objectives must be met, which the rating algorithms of respective companies should comply with.

Data Collection

The data will be collected from reliable sources like Kaggle, Datahub and Data.gov. If required, web scraping will be employed to extract data.

Problem Statement

With many homeowners insurance companies in the market, customers tend to get confused about choosing the best among the various premium prices offered by companies. On the other hand, insurance companies rely on their rating algorithm to rate policies, but are in constant need to enhance the same to survive in market. So, it becomes important to have a way to cross verify whether the charged premium amount is economical and profitable for both customer and company.

Possible Solution

The possible solution to address this problem is to build a model using Data Mining techniques, which will predict homeowners insurance premium prices from past records which will help customers to compare the prices obtained from a company/aggregator site with predicted price and chose the right company, whose premium comes close to the predicted value. Also, the model will

help insurance companies enhance their rating algorithm by comparing the price generated through algorithm and the predicted price by fine tuning it to arrive close to the predicted price. Historical data of home insurance premium prices will be used to build the model, where various input predictors will used to determine the output variable which is nothing but the premium price. The data will be categorized as training, validation and test. The training data will be used to build the model. The validation data will be used to assess the model performance and the test data will be used on the deployed model to assess the overall efficiency.