MiTH

Predicting the Claim size of Liability Insurance claims

Problem Statement:

Tort reform, to limit number and size of claims, is a contentious subject between consumer activists, claimants, Insurers and Lawyers. While there is a financial risk to the Society by way of wasting tax dollars or exodus of Insurers due to frivolous law suits, excessive court awards and exorbitant fees charged by Lawyers, leading to unsustainable Insurance rates, the ability and affordability of an individual or group of similar individuals affected for justice can not be overlooked. The “failure to warn” related laws concerning dangers of using their products/services were introduced in each line of business as part of the reform and the liability was limited as a result of reform where the manufacturer/service provided warned in an approved way.

In this context, the data of liability claims involving bodily injury settled is studied for various factors that contribute to size of claim. The data includes claims for General liability, medical professional liability, other professional liability, commercial automobile liability lines of business and the liability portion of commercial multi-peril insurance from insurance companies and self-insurers. There may be occasions where there is availability of some collateral sources of reimbursements to the injured person.

In this data, the claim size has been categorized into three levels viz., lessthan100K, 100K-lessthan200K & 200KandAbove .

Motivation for predictive analytics: While predicting the cases “200KandAbove” which are likely to result in litigation helps early intervention with suitable defence strategies and/or a settlement with claimants to avoid litigation, and reduce litigation expenses, it also helps allocation of appropriately experienced internal resources and attorneys. It also helps prepare expert witnesses to testify and reduce the claim costs. The claims predicted as “100K-lessthan200K”, potentially could develop into unexpected high severity claims of size more than $200K. So it is even more important, in such cases, to evaluate the mutual strengths to convince panel to force arbitration and avoid litigation and/or to explore the possibility of settlement. Identification of cases “lessthan100K” will help save the resources and allocate them to standard claim settlement operations team for expeditious settlement.

Data:

1. Train.csv & Test.csv (ClaimID and Target attribute:”ClaimSize”, but Test.csv doesn’t have Target attribute as it is to be predicted)

2. Train_ClaimDetails.csv & Test_ClaimDetails.csv

3. Train_Policy_Demographics.csv & Test_Policy_Demographics.csv

Supporting file : MITH_AttributeDetails.csv,

Objective:

You are expected to create an analytical and modeling framework to predict the claim size of each ClaimID as either of three classes ("**lessthan100K", "100K-lessthan200K" & "200KandAbove")** based on the quantitative and qualitative features provided in the data.

Evaluation Metric: Accuracy

Other Instructions:

1. Spend enough time on pre-processing and data understanding. Think of the problem from domain's perspective to build an efficient model.
2. Your final grader score carries much lower weightage than your overall approach which includes data exploration and model validation. Use your time wisely.

Attribute Information:

- Refer to the file "MITH_AttributeDetails.csv"