**Predicting the Claim size of Liability Insurance claims**

**Background & ProblemStatement:**

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 lawsuits, 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 cannot 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 provider 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 the size of the 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 the 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-less than 200K    & 200K and above.**

**The motivation for predictive analytics:** While predicting the cases “200KandAbove” which are likely to result in litigation helps early intervention with suitable defense 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 the 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](https://sct.insofe.edu.in/media/question_img_upload/Train-1542969243754.csv) (14418 rows, 2columns) & [Test.csv](https://sct.insofe.edu.in/media/question_img_upload/Test-1542969243754.csv) (4806 rows) (ClaimID and Target attribute: ”ClaimSize”, but Test.csv doesn’t have Target attribute as it is to be predicted)
2. [Train\_ClaimDetails.csv](https://sct.insofe.edu.in/media/question_img_upload/Train_ClaimDetails-1542969243754.csv) (14418 rows, 55 columns )& [Test\_ClaimDetails.csv](https://sct.insofe.edu.in/media/question_img_upload/Test_ClaimDetails-1542969243754.csv) ( 4806 rows,55 columns)
3. [Train\_Policy\_Demographics.csv](https://sct.insofe.edu.in/media/question_img_upload/Train_Policy_Demographics-1542969243754.csv) ( 14418 rows, 10 columns ) & [Test\_Policy\_Demographics.csv](https://sct.insofe.edu.in/media/question_img_upload/Test_Policy_Demographics-1542969243754.csv) (4806 rows ,10 columns)