# COMP 6237 Data Mining

The brief of the group project

Team member:

YUEFU LIU [yl2e18@soton.ac.uk](mailto:yl2e18@soton.ac.uk) (Team leader)

YULIN LI [yl8n18@soton.ac.uk](mailto:yl8n18@soton.ac.uk)

PENGXIANG WEI [pw1a18@soton.ac.uk](mailto:pw1a18@soton.ac.uk)

WEIQIN DING [wd2m18@soton.ac.uk](mailto:wd2m18@soton.ac.uk)

DANFENG GUO [dg1a18@soton.ac.uk](mailto:dg1a18@soton.ac.uk)

We participate in a kaggle competition:

Liberty Mutual Group - Fire Peril Loss Cost

([https://www.kaggle.com/c/liberty-mutual-fire-peril/data）](https://www.kaggle.com/c/liberty-mutual-fire-peril/data%EF%BC%89)

**Abstract**

Liberty Mutual Insurance has provided a wide range of insurance products and services designed to meet customers' ever-changing needs for over 100 years.

Within the business insurance industry, fire losses account for a significant portion of total property losses. High severity and low frequency, fire losses are inherently volatile, which makes modeling them difficult. In this challenge, your task is to predict the target, a transformed ratio of loss to total insured value, using the provided information. This will enable more accurate identification of each policyholder’s risk exposure and the ability to tailor the insurance coverage for their specific operation.

Because we seek to tap innovation both inside and outside the company, certain eligible Liberty Mutual employees are encouraged to participate in this challenge for development purposes. Refer to the competition rules for the full details.

**the detail of dataset**

This data represents almost a million insurance records and the task is to predict a transformed ratio of loss to total insured value (called "target" within the data set).

There are 300 features for each record. Some of the values are missing. The total size of the training dataset is 1.4GB and the testing dataset is 1.4GB

GitHub address: <https://github.com/data-mining-not-found>

Expected results