Overview

SafeTravel Inc. is one of the world's largest insurance companies specializing in travel insurance. During travel, there are a lot of risk factors - loss of baggage, airline cancellations, health issues etc. The potential customers are travellers who want to insure themselves against travel-related risks. They have different product offerings like 1-way travel insurance, 2-way insurance, insurance against cancellations and so on. They receive thousands of claims spread across different products.

Wrongly denying a genuine claim could lead to lawsuits against the company and approving the wrong claim would lead to a loss. Automatically predicting the claims could lead to a lot of benefits and solve some other supplementary problems too. As a team of data scientists consulting for SafeTravel Inc, you are now responsible for meeting their business outcome.

Problem Description

Insurance companies take risks over customers. Risk management is a very important aspect of the insurance industry. Insurers consider every quantifiable factor to develop profiles of high and low insurance risks. Insurers collect vast amounts of information about policyholders and analyze the data.

As a Data scientist in an insurance company, you need to analyze the available data and predict whether to sanction the insurance or not.

Dataset Description

A zipped file containing train, test and sample submission files are given. The training dataset consists of data corresponding to 52310 customers and the test dataset consists of 22421 customers. Following are the features of the dataset

• Target: Claim Status (Claim)
• Name of agency (Agency)
• Type of travel insurance agencies (Agency.Type)
• Distribution channel of travel insurance agencies (Distribution.Channel)
• Name of the travel insurance products (Product.Name)
• Duration of travel (Duration)
• Destination of travel (Destination)
• Amount of sales of travel insurance policies (Net.Sales)
• The commission received for travel insurance agency (Commission)
• Age of insured (Age)
• The identification record of every observation (ID)

Evaluation Metric

The evaluation metric for this task will be precision_score . Read up about it more here.

Submission Format

The user has to submit a csv file with the ID and Claim label. Sample submission file has been given to you. You can refer the sample submission file.