

The First ZhongAn Hackathon Competition Topic—Artificial Intelligence Track

1 Background

With the increasing demand for commercial health insurance, ZhongAn launched e-Health Insurance in 2015. It soon became the top selling product in the health care insurance market because of its price advantage and simple procedure for claim, and has largely influenced the entire health insurance industry. E-Health insurance is valid for one year, when the policy expires (30 days before the end of the policy and 15 days after the termination date), the user can choose whether to renew the policies. Precisely predicting customers' willingness to renew their insurance can provide support for marketing strategy and product design, and improve user experience. Since the launch of the product, ZhongAn has accumulated a large amount of insurance data and claims data, and ZhongAn Technology uses various statistical analysis methods and machine learning algorithms to forecast the renewal rate.

In this competition, the participants will predict whether a customer will renew his or her policy based on original policy information (transaction date, price, plan type etc.), basic customer information (age, gender, job etc.), and historical claim records.

2 Data

We provide anonymized historical data from ZhongAn e-Health Insurance. The data will be provided in four csv files. The files are coded in UTF-8, and the column separator is ','. Detailed data information is listed in the following table.

File Name	Description
policy.csv	Policy information, primary key is Policy ID
customer.csv	Customers' basic information, foreign key is Policy ID
claim.csv	Historical claim information, foreign key is Policy ID
renewal.csv	Historical renewal records, primary key is Policy ID

Sample data of the competition will be provided one month in advance. Please keep an eye on the official notice.

3 Evaluation

Each team is required to submit a result file in csv format. The result file should contain two columns:

1. Policy ID
2. Renewal label, 0 if the prediction is negative, 1 if the prediction is positive.

We use precision, recall, and F1 as assessment indicators, they are defined as follows:

- PredictionSet is the set of policies which will be renewed that your algorithm suggests.
- ReferenceSet is the set of policies which actually be renewed.
- F1 value would be the final objective assessment indicator to this problem.

$$\text{Precision} = \frac{|\cap (\text{PredictionSet}, \text{ReferenceSet})|}{|\text{PredictionSet}|}$$

$$\text{Recall} = \frac{|\cap (\text{PredictionSet}, \text{ReferenceSet})|}{|\text{ReferenceSet}|}$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Top five teams will be selected for the final oral defense with judges.

The final evaluation will consist of objective score (60 %) and live defense performance (40 %).