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 policy. 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 basic customer information (age, gender, job etc.), original policy information (transaction date, price, plan type etc.), and historical claim records.

2 Data

We provide anonymized historical data from 2015.02.01 to 2017.01.30. The data will be provided in four csv files. The files are coded in UTF-8, no header line, and the column separator is ‘,’. Detailed data information is listed in the following table.

File Name	Description
customer.csv	Customers' basic information, primary key is Customer ID
policy.csv	Policy information, primary key is Policy ID. Customer ID, Product ID are foreign keys
claim.csv	Historical claim information, User ID, Policy ID are two foreign keys and combined as the primary key
renewal.csv	Historical renewal records, primary key is Customer 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. Customer ID
2. Renewal label, 0 if the prediction is negative, 1 if the prediction is positive.

We use precision, recall, and F_1 as assessment indicators, they are defined as follows:

- PredictionSet is the set of users who will renew their insurance that your algorithm suggests.
- ReferenceSet is the set of users who actually renew their insurance.
- F_1 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 live defense performance and objective score.