Allstate Purchase Prediction Challenge

Predict a purchased policy based on transaction history

$50,000 · 1,568 teams · 3 years ago

Training Data

3 files

- sampleSubmission.csv
- test_v2.csv.zip
- train.csv.zip

sampleSubmission.csv

Download File

File size 924.99 KB

Data Introduction

Files

The training and test sets contain transaction history for customers that ended up purchasing a policy. For each customer_ID, you are given their quote history. In the training set you have the entire quote history, the last row of which contains the coverage options they purchased. In the test set, you have only a partial history of the quotes and do not have the purchased coverage options. These are truncated to certain lengths to simulate making predictions with less history (higher uncertainty) or more history (lower uncertainty).

For each customer_ID in the test set, you must predict the seven coverage options they end up purchasing.

What is a customer?

Each customer has many shopping points, where a shopping point is defined by a customer with certain characteristics viewing a product and its associated cost at a particular time.

- Some customer characteristics may change over time (e.g. as the customer changes or provides new information), and the cost depends on both the product and the customer characteristics.
- A customer may represent a collection of people, as policies can cover more than one person.
- A customer may purchase a product that was not viewed!

Product Options

Each product has 7 customizable options selected by customers, each with 2, 3, or 4 ordinal values possible:

Option name	Possible values
A	0, 1, 2
B	0, 1
C	1, 2, 3, 4
D	1, 2, 3
E	0, 1
F	0, 1, 2, 3
G	1, 2, 3, 4

A product is simply a vector with length 7 whose values are chosen from each of the options listed above. The cost of a product is a function of both the product options and customer characteristics.

Variable Descriptions

customer_ID - A unique identifier for the customer
shopping_pt - Unique identifier for the shopping point of a given customer
record_type - 0=shopping point, 1=purchase point
day - Day of the week (0-6, 0=Monday)
time - Time of day (HH:MM)
state - State where shopping point occurred
location - Location ID where shopping point occurred
group_size - How many people will be covered under the policy (1, 2, 3 or 4)
homeowner - Whether the customer owns a home or not (0=no, 1=yes)
car_age - Age of the customer's car
car_value - How valuable was the customer's car when new
risk_factor - An ordinal assessment of how risky the customer is (1, 2, 3, 4)
age_oldest - Age of the oldest person in customer's group
age_youngest - Age of the youngest person in customer's group
married_couple - Does the customer group contain a married couple (0=no, 1=yes)
C_previous - What the customer formerly had or currently has for product option C (0=nothing, 1, 2, 3,4)
duration_previous - how long (in years) the customer was covered by their previous issuer
A,B,C,D,E,F,G - the coverage options
cost - cost of the quoted coverage options