CoverWallet data-science challenge

The CoverWallet data-science challenge consists on predicting the account value of a given account with the data that the users provide during the online application and the initial quotes given. This prediction is one of the core parts of the lead scoring model. The challenge provides training and test data and submit their predictions of account value with the test data.

The process is also very important and it's very interesting for us to see your reasoning during the model development process. Plotting skills, notebook's flow, environment management, coding practices, etc will be also considered.

Data

We provide 2 data sets: accounts and quotes.

Accounts: contains a sample of our customers with the information they provided in the initial application form.

- *account_uuid*: uuid of the account
- *state*: state of the business
- *industry*: industry of the business. When blank, not indicated.
- *subindustry*: subindustry of the business. When blank, not indicated
- *year_established:* year the business was created
- *annual_revenue*: annual revenue of the business
- *total_payroll*: total payroll to the workers of the business
- *business_structure*: type of business
- *num_employees*: total number of employees of the business

Quotes: contains the quotes that were given to the user after submitting the online application form. Some of the quotes is what the user decided to finally buy.

- *account_uuid*: uuid of the account
- *product*: product type the user has requested
- *premium*: price given to the product
- *carrier_id*: insurance carrier that provides the product
- *convert*: 1 if the user has bought the product, 0 otherwise. This variable is notprovided in the test data.

The account value of a given account is defined as the sum of the premium of those products that the user has bought (convert==1). See example_account_value.csv

Submission

You have to submit for each account uuid in the accounts_test.csv, the expected account value that user will have. The file sample_submission.csv contains an example of the expected submission.

Additionally, you also have to submit a very brief document about the 3 major insights you have found. Along with the insights or within the notebook (in the case you have

use a notebook) you should also upload the EDA you have done for the analysis with some insightful charts. Plotting skills are very useful

Optional tasks

Optional 1: In general, when creating models you will face a common problem: serving it to enable stakeholders to consume it under demand. This is the case of this optional task: choose a framework and create an API service to consume the model via API Rest.

Remember to submit a quick guide of how to launch the API and consume it.

Evaluation criteria

We will use the RMSE to evaluate the predictions of the candidates. You will have to submit your **results**, the **training code** and the **insights document.**