

# Enrich Machine Learning Challenge

## Context

This dataset contains 100k transactions. A transaction is consisted of:

- *transaction\_description* - string describing a transaction
- *transaction\_amount* - the amount spent/received
- *transaction\_account\_type* - the type of account used in the transaction

The goal is to train the model that will be used to predict *transaction\_class* for a given transaction.

## Dataset

Column Information:

> *transaction\_description* - a string (can contain special characters)

> *transaction\_amount* - float (positive or negative or zero)

> *transaction\_account\_type* - can take one of these values:

*credit\_card*  
*transaction*  
*insurance*  
*foreign*  
*investment*  
*loan*  
*savings*  
*term\_deposit*  
*mortgage*  
*Superannuation*

> *transaction\_class* - label assigned to transaction (value that we want to predict), can take one of these values:

- payment
- transfer
- bank-fee
- cash-withdrawal
- interest
- refund

There are some limitations on the *transaction\_amount* these labels imply. We differentiate between **credit** (*transaction\_amount* < 0) and **debit** (*transaction\_amount* > 0) transactions.

Payment, bank-fee, cash-withdrawal transactions are credit transactions.

Interest, refund transactions are debit transactions.

Transfer can be both

## Assignment

Train a model using the dataset provided, that will be able to predict *transaction\_class* based on the input. You should use [f1 metric](#) to evaluate your model. After you train the model, run all the transactions in the scorecard file that was provided through your model and append *transaction\_class* prediction (and confidence scores in these predictions if possible) to the scorecard.

You are free to choose tools, language, frameworks, model architecture as you wish and find convenient.

As the result of your work you should provide us with:

- Scorecard result - .csv file with columns *transaction\_description*, *transaction\_amount*, *transaction\_account\_type*, *transaction\_class* (predictions from your model)
- Code - code used to train the model (github preferably)