

## Programming Assignment 1: Are we losing customers?

Customer churn occurs when customers stop doing business with a company, also known as customer attrition. It is also referred to as loss of clients or customers.

You are given **sensitive** information of 9,000 of an European Bank, EBQ. Your task is to build an Artificial Neural Network (ANN) based on the dataset such that later the ANN model can predict correctly which of the 1,000 customers (in a separate **unlabeled judge set**) the bank is going to lose. This predictive analysis is vital for the EBQ bank to revise their business strategy towards customer retention. What do you think?

Anyway, you are recruited by the bank to do the analysis. And, the head of the bank only trusts heads, i.e., brains.... I mean neural networks for making any decisions. And luckily you were in Dr. B's class and know something(?) about the ANN that you could successfully convince the head of the bank during the interview that the answer to life, the universe, and everything is the artificial neural network, "and not the number 42". He loved your answer. He put a lot of faith in you. Now, can you solve his problem?

### What you are going to submit

A zip file containing the following items:

1. A CSV file (**judge-pred.csv**) containing the predicted labels (by your best performing Artificial Neural Network in terms of accuracy) of the 1000 customers in exact same order of the judge set. Please follow the format below:

CustomerID,Exited

12345,1

34567,0

...

2. Jupyter notebook (\*.ipynb), or the Python (\*.py) file(s), with source name "**test-code**" that reads judge.csv file, and generates judge-pred.csv.
3. Jupyter notebook (\*.ipynb), or the Python (\*.py) file(s), with source name "**training-code**" that builds the artificial neural network of your choice, and reports training and test performance.
4. A Word/PDF file, with name "**Report**" describing the specifications of the artificial neural networks you built with its test performance (on a random 20% test data you set aside from the 9000 customer dataset) in the following format. It means you are going to have to run the experiment with a different artificial neural network architecture (varying either number of hidden layers, or number of perceptrons in hidden layers, or activation functions in each layer,

or number of epochs, etc.). Please try several changes in a way that you achieve a reasonably good performance (e.g., 80% test accuracy just to put a number here).

### Report

	Configurations	Accuracy		Precision		Recall		F1-score	
		Train set	Test set	Train set	Test set	Train set	Test set	Train set	Test set
ANN-1	11x10, ReLU 10x5, ReLU 5x4, ReLU 5x1, Sigmoid Epoch=100 Scaling=yes								
ANN-2	11x11, ReLU 11x9, ReLU 9x5, ReLU 5x3, Sigmoid 3x1, Sigmoid Epoch=100,000 Scaling=yes								
...									

Here, ANN-i is the best model showing accuracy 0.79 (just an example), and we are going to use it to predict the **judge.csv** dataset.

### What you are program to accomplish the assignment

- Download the datasets:
  - [dataset.csv](#) – Containing 9000 customer information (labeled)
  - [judge.csv](#) – Containing 1000 customer information (unlabeled)
- Preprocess the dataset, and have a 80-20 split (training and test). *Before doing that, please investigate which feature(s) might or might not have relations to the target variable “Exited”.*

*You may want to drop those (very) distantly related features from your dataset. For instance, CustomerID, Surname may not have any correlation with the target output variable.*

3. Design an artificial neural network, ANN using Keras library. For reference, please take a look here <https://keras.io/getting-started/sequential-model-guide/>. For simplicity, please use the 'adam' optimizer and 'binary\_crossentropy' as the loss function the optimizer will use.
4. Apply the ANN you designed above on the training set, and then apply the trained model on the test dataset.
5. Report the architecture, algorithm parameters, and performance into the Report document.
6. Write a program that has access to your best performing trained ANN (in terms of accuracy metric), and reads the **judge.csv** file. It will output **judge-pred.csv** containing only two fields: CustomerID, Exited for each of the judge instances present in the **judge.csv** file in the exact same order.

#### CAUTION

- Plagiarism is strictly checked for each submitted code. Please do not share source files with your fellow classmates. Both parties will be heavily penalized.

#### Graduate level tasks [**Please attempt only if you are a graduate student, otherwise it's optional**]

7. Repeat all the above tasks without using the Keras library to design and implement each of the ANN you planned, meaning please write the training algorithm from scratch using the binary cross entropy loss function which will be optimized by the stochastic gradient descent algorithm.