**Forecasting Credit Worthiness of Clients in Consumer Finance**

The topic selected here pertains to consumer finance and credit card portfolios.

A screenshot of a computer

Description automatically generated

1. This project aims to use retail portfolios in consumer finance and forecast the risk associated with clients credits based on their loan and credit history, hence predicting the clients' creditworthiness.

2. The question to be addressed here is "Can we develop a model based on retail portfolios in consumer finance that automates the detection of high-risk behavior in clients' loans and, accordingly, whether or not they would default in the foreseeable future?"

A screen shot of a computer

Description automatically generated

1. The analytic approach here could be first an anomaly detection that, based on the client's history of credit and loan data, raises an alert to the lenders, and then a secondary analysis such as logistic regression with time-series forecasting could be used to foresee if in a foreseeable future, e.g., 3 months, the client would default on a specific loan book.

2. Examples of the data required to address this problem would be

- the history of the account balance,

- the history of loan or credit payments to different lenders,

- number of missed payments with specific days-past-due (DPD) bucket details as the higher DPD, the higher the risk of default,

- and the history of the client's credit score on the loan book over time.

3. The data collection for these types of problems could occur through collecting the client's information tied to their social security number. If one were to solve this problem in connection with a specific firm, then they would have to go through ethical policies and protocols of data collection to be in accordance with the privacy of the clients' data. Further, some other considerations, such as stationarity of the variables and etc, should be taken into account since the data will be used for time-series forecasting.

4. The data related to one's history of credit balance, payments, credit, and missed payments from different lenders could be in different formats (different file types or different data structures), frequencies (different on how granular the data frequency is), and types and names of variables (the variables could be named differently or could refer to different things and need transformation) on record.

5. To model the data, one could use the data per individual and split them into training and test sequences of data. First, a logistic time-series forecasting model will be trained on the training sequence using cross-validation to achieve high accuracy, and then the trained model will be used to see if it can perform accurately enough in forecasting the default flag for the test portion of the data.