Group Name: Tech Geeks
Name: Vaishnavi Udanshiv
Email: vaishnaviudanshiv22@gmail.com
College: Munster Technological University
Specialization: Data Science

Problem Description: The problem is to develop a predictive model that can assess the credit worthiness of potential future customers of a financial institution. The available data set consists of 807 past loan customer cases, each with 14 attributes including financial standing, reason for the loan, employment, demographic information, foreign national status, years of residence in the district, and the outcome/label variable Credit Standing, which classifies each case as either a good loan or bad loan. The objective is to build a model that accurately predicts the credit standing of new loan applications, using the available data as the training set. The model should be able to identify the key factors that determine creditworthiness and provide insights to help the financial institution make better lending decisions.

Business Understanding:

In the given problem statement, the business understanding stage involves understanding the financial institution's objective of assessing the creditworthiness of potential future customers. This stage requires a thorough understanding of the financial domain, including the lending process, loan evaluation criteria, and credit risk management. The data scientist needs to work closely with Maeve, the manager of the financial institution, to identify the specific problem to be solved, such as accurately predicting the credit standing of loan applications, which can reduce the risk of default and improve the profitability of the institution. It is also essential to define the key performance indicators, such as accuracy, precision, and recall, that the model should achieve to meet the business objectives. By gaining a deep understanding of the business context and goals, the data scientist can build a model that not only provides accurate predictions but also helps the financial institution make better lending decisions by identifying the key factors that determine creditworthiness.

Project Lifecycle:

Data Cleaning : 2-3 Days

Exploratory Data Analysis: 7 Days

Model Selection and Building: 7 Days

Model Evaluation : 2-3 Days

Final Presentation : 7 Days

Data Intake Report:

The dataset contains information on 807 past loan customer cases, with 14 attributes for each case. The following is a description of the dataset attributes:

1.   ID: A unique identifier for each customer case.

2.   Checking.Acct: The status of the checking account for the customer, with values "No Acct", "0Balance", and "Low".

3.   Credit.History: The credit history status of the customer, with values "All Paid", "Current", "Delay", and "Critical".

4.   Loan.Reason: The reason for the loan, with values "Car New", "Furniture", "Small Appliance", "Education", and "Car Used".

5.   Savings.Acct: The status of the savings account for the customer, with values "Low", "MedLow", "Medium", and "No Acct".

6.   Employment: The employment status of the customer, with values "Short", "Very Short", "Long", and "No Job".

7.   Personal.Status: The personal status of the customer, with values "Single", "Married", and "Divorced".

8.   Housing: The housing status of the customer, with values "Own", "Rent", and "Other".

9.   Job.Type: The job type of the customer, with values "Management", "Skilled", "Unskilled", and blank.

10. Foreign.National: A binary attribute indicating whether the customer is a foreign national or not.

11. Months.since.Checking.Acct.opened: The number of months since the checking account was opened.

12. Residence.Time.In.current.district: The number of years the customer has resided in the current district.

13. Age: The age of the customer in years.
14. Credit.Standing: The outcome variable indicating the credit standing of the customer, with values "Good" and "Bad".

The dataset contains information on customers' financial standing, employment, personal status, and demographic information. The primary objective is to build a predictive model to assess the creditworthiness of potential future customers, using the available data as the training set. The data scientist will use this dataset to build a model that accurately predicts the credit standing of new loan applications and identifies the key factors that determine creditworthiness to help the financial institution make better lending decisions.