**Table of Contents.**

[1 Introduction 1](#_Toc414111172)

[2 Challenge. 1](#_Toc414111173)

[3 Your role. 1](#_Toc414111174)

[4 Details of the data 1](#_Toc414111175)

[5 Assignment 3](#_Toc414111176)

# Introduction

A bank is concerned about the potential for loans not to be repaid. If previous loan default data can be used to predict which potential customers are liable to have problems repaying loans, these "bad risk" customers can either be declined a loan or offered alternative products.

This example focuses on using existing loan default data to predict potential future defaulters, and looks at three different Bayesian network model types to establish which is better at predicting in this situation.

# Challenge.

For banks, it is important to predict the default probability of a loan with high accuracy. Whenever an applicant applies for a loan, predicting whether the loan will be repaid is an important activity for any bank. High accuracy is beneficial for both the banks and the loan applicants. In this exercise, we learn to apply Bayesian Networks to classify the sample data into default and non-default categories.

Bayesian networks enable you to build a probability model by combining observed and recorded evidence with "common-sense" real-world knowledge to establish the likelihood of occurrences by using seemingly unlinked attributes.

# Your role.

You are a credit modeler in a bank and you are entrusted with the responsibility of developing models that predict the default probability of a loan. In this phase, you have to confirm that you understand the business requirements (business understanding) as well as the data (data understanding) for the project.

# Details of the data

The bank has shared the data with your team.

The data is in the form of a text file “BankBayesLoan.txt”. The description of the various fields are given below.

The bank has stated that the target variable is “default”.



**Data Description**

The data is described below

| **SN** | **Name** | **Description** |
| --- | --- | --- |
| 1 | age | Age in years of the loan applicant |
| 2 | ed | Level of education of a loan application |
| 3 | employ | Years with current employer |
| 4 | address | Years at current address |
| 5 | income | Household income in thousands |
| 6 | debtinc | Debt to income ratio in thousands |
| 7 | creddebt | Credit card debt in thousands |
| 8 | othdebt | Other debt in thousands |
| 9 | default | Previously defaulted |

# Assignment

**Exercise 1.**

On the basis of the problem statement and the data provided, carry out the following steps in CRISP DM for the client –

1. Business Understanding
2. Data Understanding

The response for this exercise will be a report in the form of a word document, with separate sections for the deliverables below (refer to the lecture on CRISP-DM Methodology).

|  |  |
| --- | --- |
| **Stage** | **Deliverables** |
| **Business Understanding** | Data Mining Goals |
|  | Data Mining Success Criteria |
| **Data Understanding** | Data Exploration Report |
|  | Data Quality Report |

**Exercise 2.**

Carry out preliminary data analysis to identify variables that affect/ might affect the target variable (Default) and provide a report on your findings.