**Problem Statement**

Loan Prediction is a common problem in the banking industry. Given a customer's information, the goal is to predict whether the customer will default on their loan. This is a binary classification problem, where the output is either "Yes" or "No".

The input data may include the customer's personal information (Gender, income, Employment status, Education), information about the loan (loan amount, Credit History, loan Amount term), and other relevant data.

The output of the model should be a prediction of whether the customer is likely to get the loan or not. This prediction will be used by the bank to make informed lending decisions and to manage their loan portfolio.

**Machine learning related works**

1. Loan approval prediction using machine learning:

* They used machine learning using python on a data set to provide bank system which will be approved for the loan and which will not with information like user education, depended, married, employed, and loan amount.
* They used pandas and numpy as a libraries to load and clean the data frame.
* They used matplotlib to visualize the data features.
* They used sklearn library to encode, train and split the data.
* They used more four modules to get the different accuracies:
  + Random Forest Classifier.
  + KNeighbors Classifier.
  + SVC.
  + Logistic Regression.
* They concluded that the Random Forest Classifier algorithm is the best module for this data with accuracy 82.5%, and the worst is KNeighbors Classifier algorithm with 63.75%.

2) Loan Approval Prediction:

* They used machine learning on a different data set to establish a decision for the loan approval of accuracy ranging from 75-85% but the best score is the Logistic Regression 88.70%.
* They used information from the data set in the module like user education, depended, married, employed, loan amount, and credit history.
* They applied some data processing (cleaning) on the data to give the best results.
* They used five different machine learning algorithms to define the best accuracy in this data set for the loan approval prediction:
  + Random Forest.
  + Naïve Bayes.
  + Decision Tree.
  + Logistic Regression.
  + K Nearest Neighbor.
* They concluded that the best module for this data is the Logistic Regression algorithm with accuracy 88.70%, and the worst module is the Random Forest algorithm with accuracy 79.03%.

Model Architecture

1. **Decision Tree**

**Figure 1**. Architecture of a Decision Tree Classifier Model [1].

1. **Logistic Regression**

Diagram

Description automatically generated

**Figure 2**. Architecture of a Logistic Regression Model [2].

1. Diagram

   Description automatically generated **Random Forest**

**Figure 3**. Architecture of a Random Forest Model [3].

1. **Support Vector Classifier**

Diagram

Description automatically generated

**Figure 4**. Architecture of a Support Vector Classifier Model [4].

[1]. “Decision tree algorithm in Machine Learning - Javatpoint,” *www.javatpoint.com*. [Online]. Available: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm. [Accessed: 01-Jan-2023].

[2]. S. Raschka and V. Mirjalili, *Python Machine Learning, Second edition*. s.l: Packt Publishing, 2017.

[3]. A. Gelzinis, A. Verikas, E. Vaiciukynas, M. Bacauskiene, J. Minelga, M. Hallander, V. Uloza, and E. Padervinskis, “Exploring sustained phonation recorded with acoustic and contact microphones to screen for laryngeal disorders,” *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)*, 2014.

[4]. R. Ruiz-Gonzalez, J. Gomez-Gil, F. Gomez-Gil, and V. Martínez-Martínez, “An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis,” *Sensors*, vol. 14, no. 11, pp. 20713–20735, 2014.