IRG61104 Advanced Machine Learning
Universiti Kuala Lumpur
January 2024
F.M.N

Mini Project

Weight: 50%
Due: 9 February 2024

1 Introduction

The objective of this assignment is to build a deep learning model to assess credit risks based on the provided credit risk dataset.

1.1 Goal

Your final goal is to predict predict whether a loan will **default** or **not**.

2 Dataset

Feature	Description
`person_age`	Age
`person_income`	Annual income
`person_home_ownership`	Home ownership
`person_emp_length`	Employment length (in years)
`loan_intent`	Loan intent
`loan_grade`	Loan grade
`loan_amnt`	Loan amount
`loan_int_rate`	Loan interest rate
`loan_status`	Loan status (0 is non default 1 is default)
`loan_percent_income`	Percent income
`cb_person_default_on_file`	Historical default
`cb_person_cred_hist_length`	Credit history length

Table 1: Credit risk dataset.

2.1 Credit Risk Dataset

The credit risk dataset encompasses crucial attributes such as individual age, income, home ownership status, employment length, loan details (including intent, grade, amount, and interest rate), and credit history metrics. As shown in Table 1, the dataset provides a foundation for you to explore and analyse factors influencing credit risk. The dataset can be downloaded from here

3 Objective

Build a **deep learning model** to assess credit risks based on the provided credit risk dataset.

4 Instructions

4.1 Data Exploration [10 points]

1. Load the dataset and perform exploratory data analysis.
2. Check for missing values, outliers, and data distributions.
3. Visualise the distribution of the target variable (`loan_status`).

4.2 Data Preprocessing [10 points]

1. Handle missing values (if any) appropriately.
2. Encode categorical variables using *one-hot* encoding or *label* encoding.
3. Standardise or normalise numerical features.
4. Split the dataset into training and testing sets.

4.3 Model Building [10 points]

1. Design a deep learning model for binary classification (predicting `loan_status`: 0 or 1).
2. Choose an appropriate architecture with input layers, hidden layers, and output layers.
3. Compile the model with an appropriate loss function and optimiser.
4. Train the model on the training dataset.

4.4 Model Evaluation [10 points]

1. Evaluate the model on the testing dataset.
2. Compute and interpret the following evaluation metrics:
   (a) Accuracy
   (b) AUC
   (c) Precision
   (d) Recall
   (e) F1-score
3. Generate a confusion matrix to visualise the model's performance.

4.5 Model Improvement [10 points]

1. Experiment with different hyperparameters (e.g., learning rate, number of layers, number of neurons) to optimise the model's performance.

4.6 Discussion [10 points]

1. Discuss the challenges of credit risk assessment and the importance of interpretability in such models.
2. Explore ethical considerations and potential biases in credit risk models.

5 Assignment Collaboration Policy

Encourages collaboration while emphasising originality. Students can discuss, share resources, but must submit unique work. Plagiarism is strictly prohibited. Acknowledge significant contributions. Violations, including plagiarism, will be penalised.

6 Submission

Please submit the following materials:

- Submit a Jupyter notebook containing your code, explanations, and visualisations.
- Include a summary of your findings and insights from the analysis.
- Reflect on the model's performance and discuss potential improvements.

Made using LaTeX