HW - Project Organization

Dinelka Nanayakkara

October 7, 2024

1 Git Respository

The git repository can be found here: https://github.com/dinelka97/BIOS-611-Project.

2 Data Description

The dataset used for this project relates to loan approval rates. The primary goal of using this data would be to build a model that could predict if an individual gets approved for a loan or not. Although supervised learning will be the primary task of this project, I aim to also perform unsupervised learning to explore patterns that might be in the data.

2.1 Variables

The following table provides a brief overview of the variables used in the dataset.

var_name	**label**	data_type
id	Unique identifier	num
person_age	Age	num
person_income	Income level	num
person_home_ownership	Home ownership status (own, rent, mortgage, other)	categorical
person_emp_length	Length of employment	num
loan_intent	Purpose of obtaining loan	categorical
loan_grade	Credibility of loan	categorical
loan_amnt	Loan amount	num
loan_int_rate	Loan interest rate	float
loan_percent_income	Loan amount as a percentage of income	float
cb_person_default_on_file	Default indicator	categorical
cb_person_cred_hist_length	Length of credit history	numeric
loan_status	Loan Approval Status	categorical

2.2 Objectives

I plan to work on the following as part of the project:

- If we consider person_home_ownershipm person_emp_length, and person_income as one variable - can we build clusters based on this?
- Does a past credit default decrease the possibility of loan approval by a great extent?
- Can we classify loan amounts and credibility of an individual?

- I hypothesize that those with a low income do not have a good credit history and they are more prone to apply for loans, thus having a longer credit history.
- Build a model using logistic regression, random forest, and neural networks to predict if a loan is approved or not.