SkillForge

Minor Project

Project Name:

Data Science July Minor Project

Project Description:

Problem statement: Create a classification model to predict whether credit risk is good or bad.

Context: As a banking. Financial institution, is interest is to know the potential financial whereabouts of the customers in order to determine whether the credit risk associated with them is good or bad.

This dataset consists of 21 features of the customers. It could be used to predict if the customer could be given credit. Many features require data cleaning

Dataset:

https://drive.google.com/file/d/1QIuRa7EObS8qMT5ewAddFP4h1XbjhhSU/view?usp=sharing

Target Variable: Class

Steps to Consider:

1) Remove handle null values (if any).
2) Apply the necessary data preprocessing like
    i)   Handling null values
    ii)   Handling duplicate records
    iii) Label Encoding or Dummy variable creation for encoding categorical columns
3) Select x (independent variable) and y (dependent variable) – class (good or bad)
4) Split data into training and test data.
5) Apply the following models on the training dataset and generate the predicted value for the test dataset
    a) Logistic Regression

    b) KNN Classification
    c) SVM Classifier with linear and rbf kernel
6) Predict the class for test data
7) Generate Confusion matrix and classification report for each of these models.
8) Report the model with the best accuracy.