**Problem Statement:**

Default of Credit Card Clients Data Set

Estimate the probability of Default

**Dataset:**

This advanced level data set has 30000 rows and 24 columns.

The data set could be used to estimate the probability of default payment by credit card client using the data provided.

**Data Preprocessing:**

1. Handling NULL values

There are no null values so there is no need to replace them with the mean or to delete any rows or columns.

1. The first row was filled with X1, X2, and so on with the next row being the names of the names of the variables so it had to be fixed
2. Feature scaling was done where ever it was needed

75% of the data was used for training the model and the remaining 25% was used for testing

**Models and Techniques Used:**

1. Decision Trees

Only manager an accuracy of 73% even after using Bagging, model was overfitted

1. kNN

Accuracy: 79%

1. Kernel SVM

Accuracy: 82%

1. Logistic Regression

Accuracy: 81%

1. Random Forest Classification

Accuracy: 82.48%

1. SVM

Accuracy: 81%

1. LightBGM

Accuracy: 82.5%

1. XGBoost

Accuracy: 82.1%

Most of the models managed to get around 80% but the decision trees and the random forest were overfitted meaning that more variation is needed

**Final Results:**

Ensemble methods like Random Forest, LightBGM, and XGBoost demonstrated the best performance, with accuracies ranging from 82.1% to 82.5%.