

You are working as a loan analyst at Goods-Fargo Bank. You decided to create a tool to determine whether a loan applicant would repay a loan he/she received. You have a **Data set with** several customer variables and whether a customer repaid (Repay variable) the loan. Create training and testing samples.

1. Use logistic regression to predict who is likely to repay the loan. Evaluate the following two models in terms of their prediction accuracy. In the first model, use all the predictors, i.e., full model. In the second model, use 4 predictors. For the second model, you are free to choose which predictors to leave out. In the third model use only significant predictors(using *statsmodels* library). Then evaluate these models based on their **overall classification accuracy**.

Recommend a final model and express the model as a mathematical equation relating the output variable to the input variables.