Classification
for Housing Loans

Overview and the problem:

Housing is one of the necessities of life that must be owned. Our company provides a financing service for people to obtain housing. And the agreement with the customer to pay this amount in installments within a specified period of time
We face the problem of defaulting on some customers' payments, which makes us form a team to analyze customers and find out who will default, and this is very difficult, so we will do this process by learning the machine and trying to predict through data of customer.

DATA:

The data source is: https://www.kaggle.com/burak3ergun/loan-data-set

The Raw data contains 13 features: Loan_ID , Gender , Married , Dependents , EducationSelf_Employed , ApplicantIncome , CoapplicantIncome , LoanAmount , Loan_Amount_Term , Credit_History , Property_Area , Loan_Status .

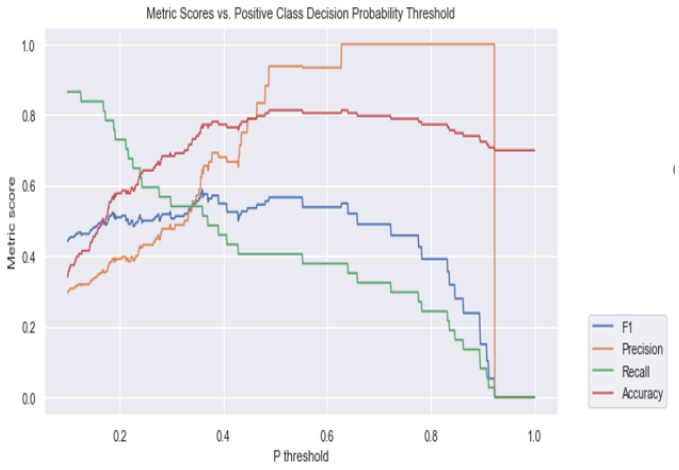
The data is customer data such as gender, marital status and education..

Experiments:

Baseline	train	0.76
	val	0.76
Decision Tree	Train	0.814
	Val	0.756
Random Forest	Train	0.808
	Val	0.813
Knn	Train	0.812
	Val	0.813
Logistic Regression	train	0.816
	val	0.814
Naïve Bayes	train	0.794
	val	0.812

SVM	train	0.688
	val	0.699

Logistic Regression:



Conclusion:

- We have used many algorithms to predict the loan status
- The highest rated algorithm is **Logistic Regression**