LOAN
ELIGIBILITY PREDICTION

By: Rika Sahriana

CONTENT:

BACKGROUND

BACKGROUND

Loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don't have assurance if the applicant is able to repay the loan with no difficulties.

PROBLEMS

A Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. To automate this process, we need to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers.

About Loan Eligibility Prediction

Loan eligibility is defined as a set of criteria basis which a financial institution evaluates to decide eligibility of a customer for a particular loan.

Criteria

Loan amount | Dependents | Marital status
Applicant income | Loan amount term | Co-applicant income
Gender | Credit history | Property area

Loan Application
Document Submission
Document Verification
Loan Approval

OBJECTIVE

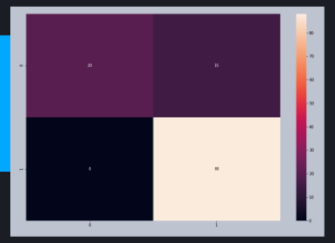
Data Introduction

For this problem, I use Loan Eligible Dataset from kaggle.

- Loan_ID
- Married
- Education
- ApplicantIncome
- Loan_Amount
- Property_Area
- Loan_Status
- Gender
- Dependents
- Self_Employed
- CoapplicantIncome
- Credit_History
- Loan_Amount_Term

Confusion Matrix

Confusion Matrix Logistic Regression



RESULT SUMMARY

Result Summary

Conclusion

The best model based on the analysis that has been done is logistic regression with a model accuracy level of 87.80% and a cross validation score of 80.94%. So that in implementing the eligibility loan prediction, we can use this model to get best prediction

BACKGROUND

Loans are the core business of banks. The main profit comes directly from the loan’s interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don’t have assurance if the applicant is able to repay the loan with no difficulties.

PROBLEMS

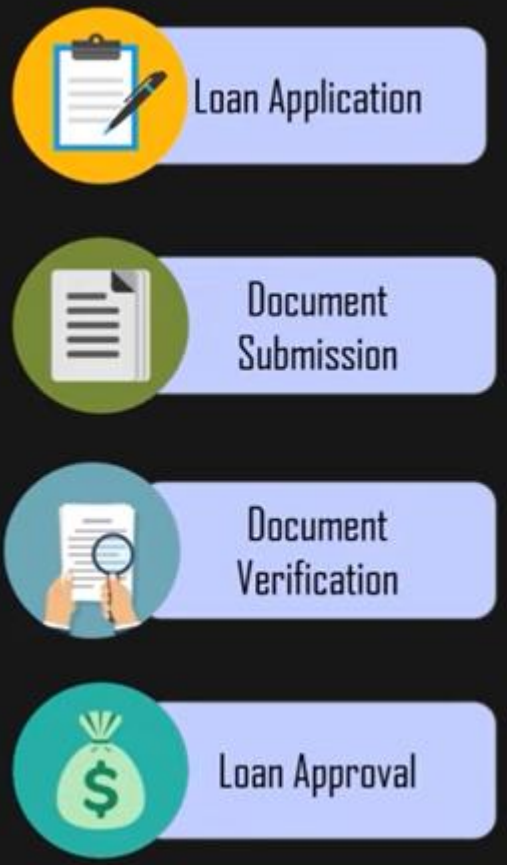
A Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. To automate this process, we need to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers.

About Loan Eligibility Prediction

Loan eligibility is defined as a set of criteria basis which a financial institution evaluates to decide eligibility of a customer for a particular loan.

Criteria

Loan amount	Dependents	Marital status
Applicant income	Loan amount term	Co-applicant income
Gender	Credit history	Property area



Data Introduction

For this problem, I use Loan Eligible Dataset from kaggle.

- Loan_ID
- Married
- Education
- ApplicantIncome
- Loan_Amount
- Property_Area
- Loan_Status
- Gender
- Dependents
- Self_Employed
- CoapplicantIncome
- Credit_History
- Loan_Amount_Term

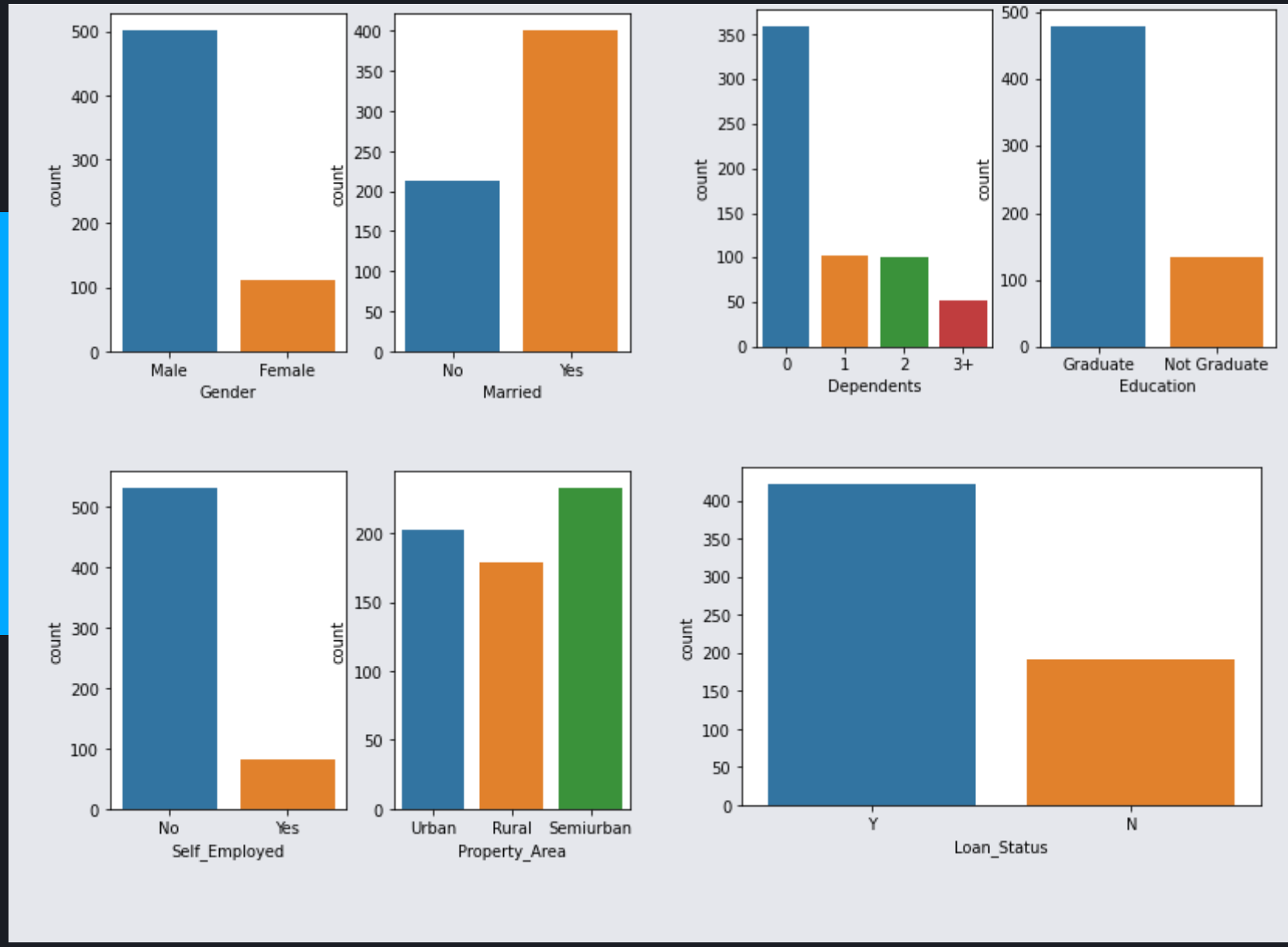
Data Introduction

Null value in each column

Key Name	Null.sum
Loan_ID	0
Married	3
Education	0
ApplicantIncome	0
Loan_Amount	22
Property_Area	0
Gender	13
Dependents	15
Self_Employed	32
CoapplicantIncome	0
Credit_History	50
Loan_Amount_Term	14

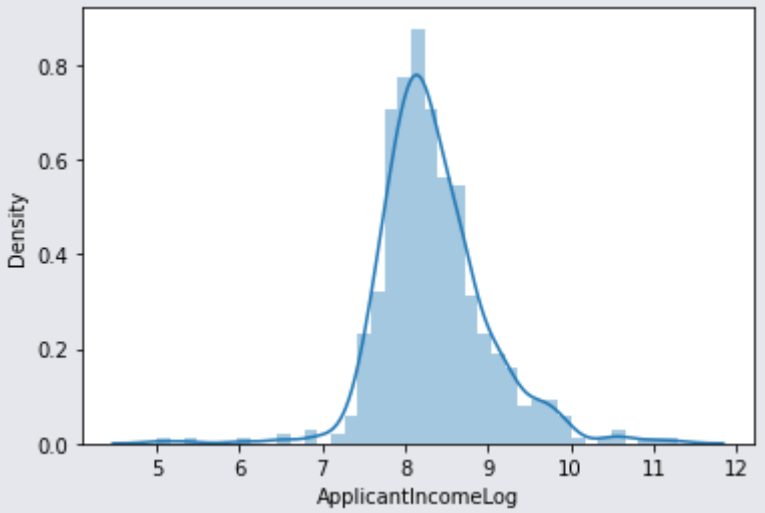
Exploratory Data Analisys

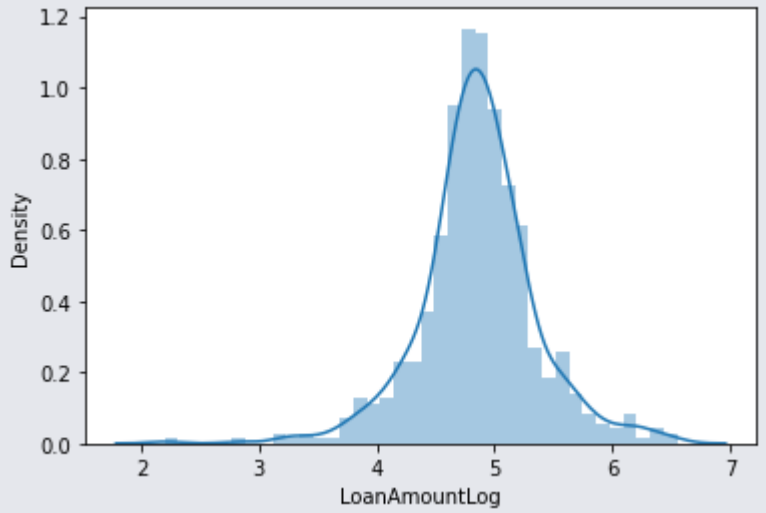
Categorical Data Visualization

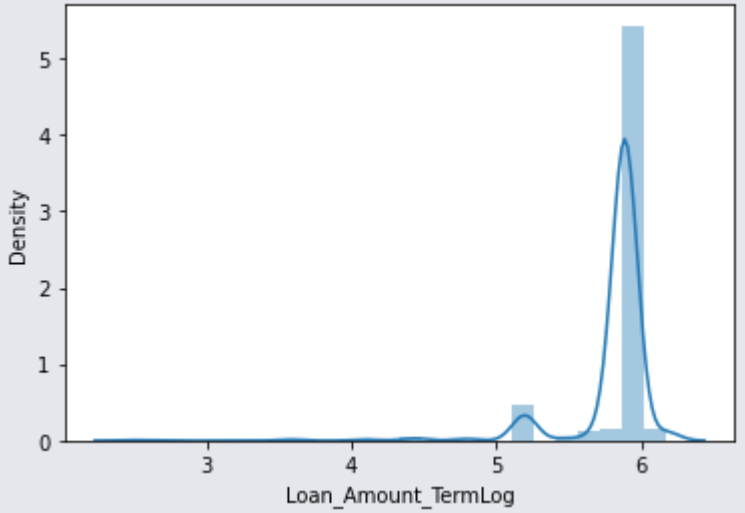


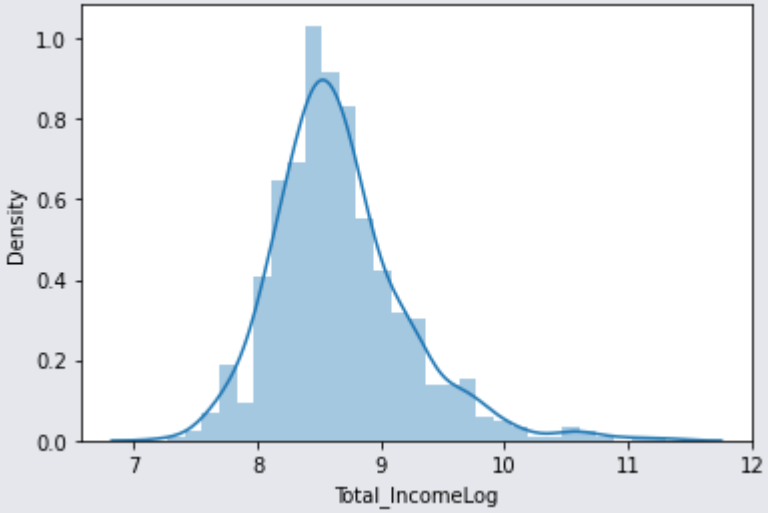
Exploratory Data Analisys

Numeric Data Visualization









Exploratory Data Analisys

Coorelation Matrix

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Total_Income	ApplicantIncomeLog	CoapplicantIncomeLog	LoanAmountLog	Loan_Amount_TermLog	Total_IncomeLog
ApplicantIncome	1	-0.12	0.57	-0.045	-0.014	0.89	0.79	-0.25	0.43	-0.024	0.72
CoapplicantIncome	-0.12	1	0.19	-0.06	-0.0017	0.34	-0.24	0.57	0.2	-0.043	0.38
LoanAmount	0.57	0.19	1	0.039	-0.0077	0.62	0.55	0.048	0.9	0.045	0.69
Loan_Amount_Term	-0.045	-0.06	0.039	1	0.0014	-0.07	-0.031	0.014	0.087	0.94	-0.055
Credit_History	-0.014	-0.0017	-0.0077	0.0014	1	-0.014	0.019	0.0024	-0.025	-0.0079	0.019
Total_Income	0.89	0.34	0.62	-0.07	-0.014	1	0.64	0.023	0.5	-0.042	0.85
ApplicantIncomeLog	0.79	-0.24	0.55	-0.031	0.019	0.64	1	-0.37	0.49	-0.015	0.74
CoapplicantIncomeLog	-0.25	0.57	0.048	0.014	0.0024	0.023	-0.37	1	0.13	0.013	0.15
LoanAmountLog	0.43	0.2	0.9	0.087	-0.025	0.5	0.49	0.13	1	0.085	0.66
Loan_Amount_TermLog	-0.024	-0.043	0.045	0.94	-0.0079	-0.042	-0.015	0.013	0.085	1	-0.03
Total_IncomeLog	0.72	0.38	0.69	-0.055	0.019	0.85	0.74	0.15	0.66	-0.03	1

1.0 0.8 0.6 0.4 0.2 0.0 −0.2

Train Test Split

I use 20% data as Test data and 80% as Train data

	shape
X	(614, 11)
x_train	(491, 11)
x_test	(123, 11)

Training Model

To get best prediction, I use 5 Method. There are

	Training Accuracy	Testing Accuracy	Cross Validation Score
Logistic Regression	79.42	87.80	80.94
DecisionTree Classifier	100.0	78.04	71.17
RandomForestClassifier	100.0	83.73	78.17
ExtraTrees Classifier	100.0	80.48	76.22
Support Vector Machine	68.02	71.54	69.70

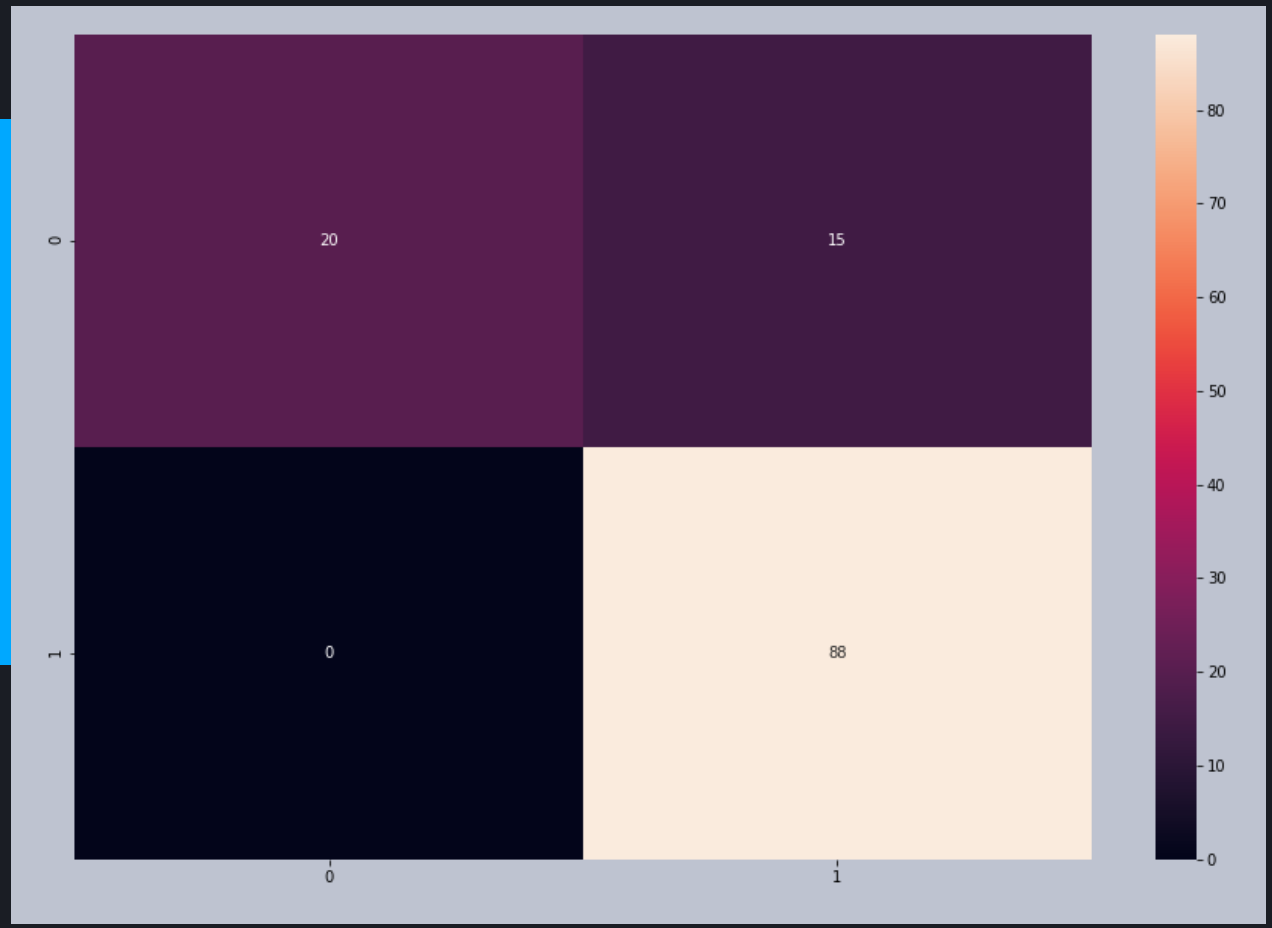
Hyperparameter Tuning

To improve accuracy model, i try to use 2 method hyperparameter tuning

	Training Accuracy	Testing Accuracy
GridSearchCV	79.42	87.80
RandomizedSearchCV	79.42	87.80

Confusion Matrix

Confusion Matrix Logistic Regression



Result Summary

Conclusion

The best model based on the analysis that has been done is logistic regression with a model accuracy level of 87.80% and a cross validation score of 80.94%. So that in implementing the eligibility loan prediction, we can use this model to get best prediction