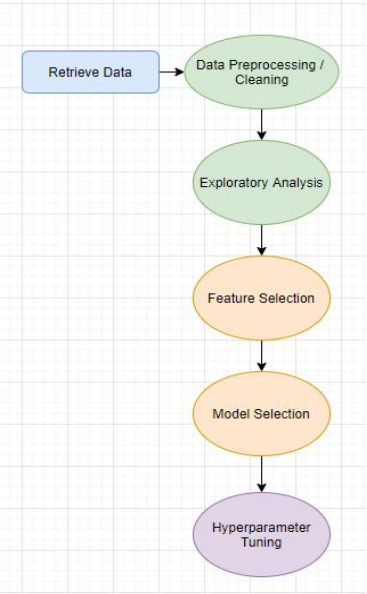
Mini-Project 4

By: Sunny Bhandal

Introduction

Predicting the loan eligibility of customers using supervised learning.

- **Potential Benefits**:
  - Reduce time by implementing an online approval system.
  - Opportunity to sell more products to those customers who are approved.



Hypotheses

- Does a person's education level affect their chances of getting a loan?
  - Potentially significant feature (9% difference).

Loan_Status	N	Y
Education		
Graduate	134	331
Not Graduate	47	80

- Does gender play a role in the approval of a loan?
  - Potentially insignificant.

Gender	Female	Male
Loan_Status		
N	36	140
Y	73	330

- Does credit history play a role in the approval of a loan?
  - Potentially significant (79% with credit history get approved).

Loan_Status	N	Y
Credit_History		
0.0	82	7
1.0	110	415

Bias

Upsampling because of the size of the dataset.

- The dataset mostly contains male customers (81%).
- The dataset mostly contains customers that are not self-employed (86%).
- The dataset mostly contains customer who have graduated (79%).
- The dataset mostly contains customers with a credit history (85%).

Data Handling

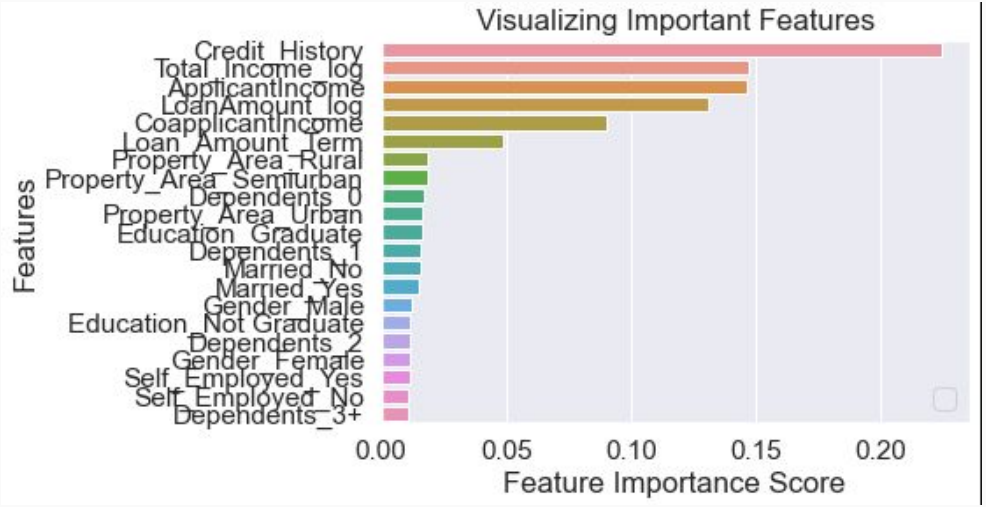
- Use mean/mode methods to fill missing data.
- Log transformations on numerical data that is highly skewed.
- Standard Scaler transformation on the remaining numerical data.
- Create subsamples based on different criteria to limit bias (gender, education, etc.).
- Encode categorical variables.

Initial Results

Model Accuracy: 78.4%

Classifier: Random Forest Classifier

Feature Importance:



Adjusted Model

Model Accuracy: 73.0%

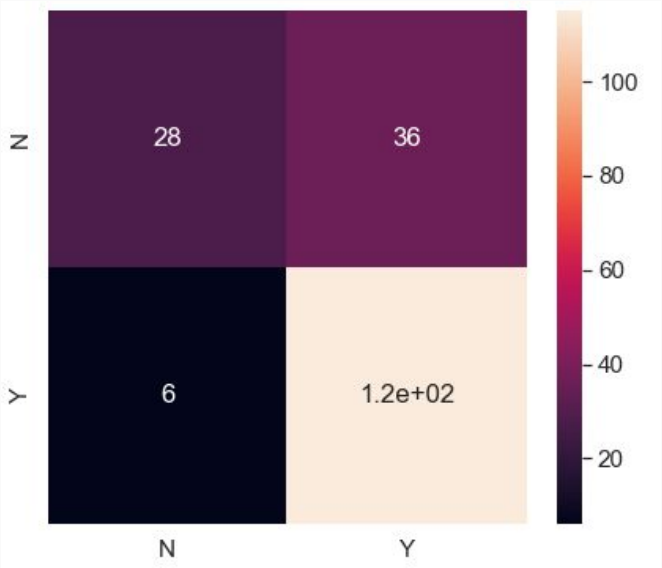
Classifier: Random Forest Classifier

Selection Method: SelectKBest

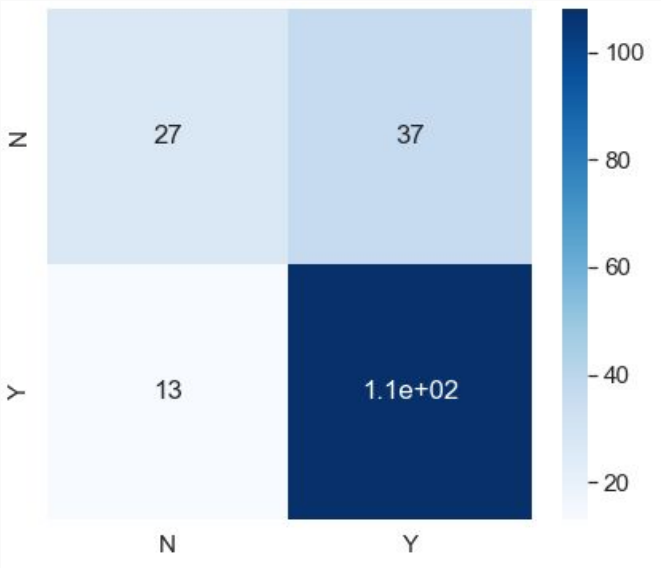
- Performance is worse without all variables.

Comparisons

Before



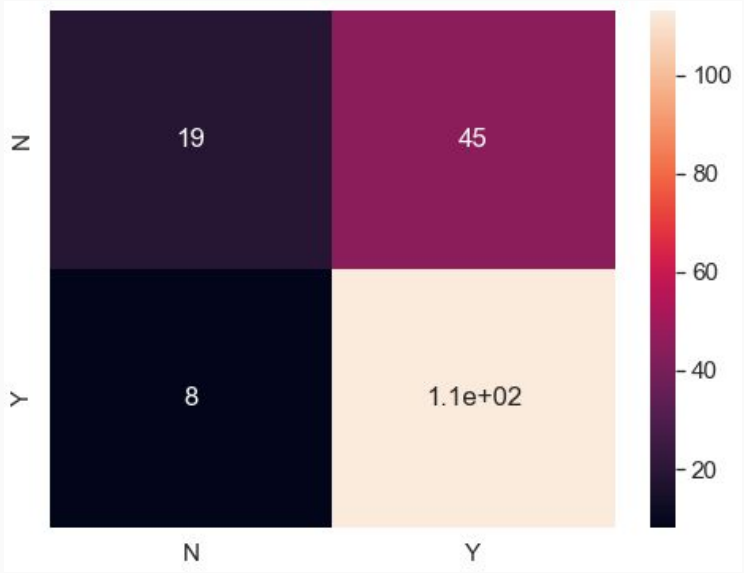
After



Deployment Model

Model Accuracy: 71.4%

Classifier: Random Forest Classifier



Deployment Model

Data:

```
json_data = {'Gender': 'Female',
 'Married': 'Yes',
 'Dependents': '2',
 'Education': 'Graduate',
 'Self_Employed': 'No',
 'ApplicantIncome': 3500.0,
 'CoapplicantIncome': 4000.0,
 'LoanAmount': 160.0,
 'Loan_Amount_Term': 360.0,
 'Credit_History': 1.0,
 'Property_Area': 'Urban'}
```

Output:

```
[[0.27, 0.73]]
```

(73% chance of approval)

Challenges

- PIPELINES
- Time (handling errors, bias, model performance)

Conclusion

Credit History and Total Income were the most significant features for predicting loan approval.