Mini project : Loan prediction

Asmaa Chraibi

Project scope

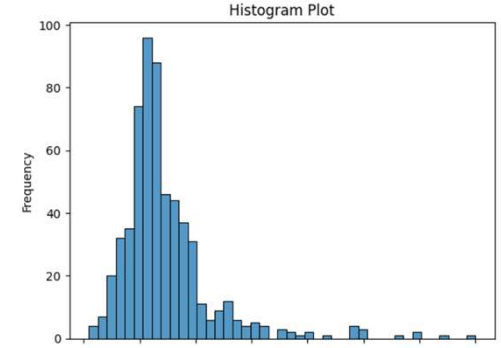
Context: the client is a financial institution

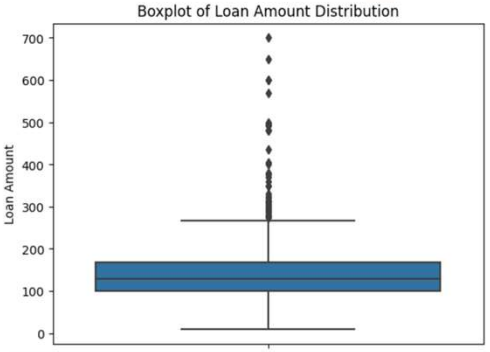
- Develop a prediction model
- Use Pipelines
- Deploy the model

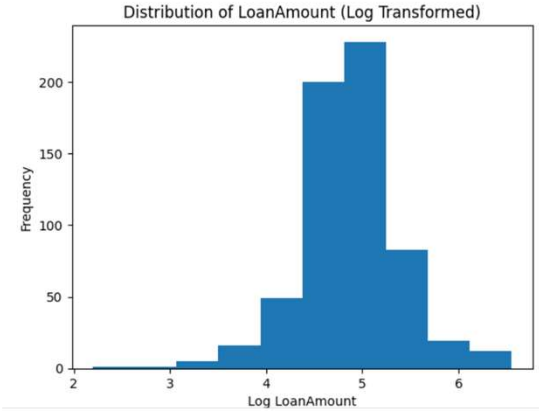
Process steps

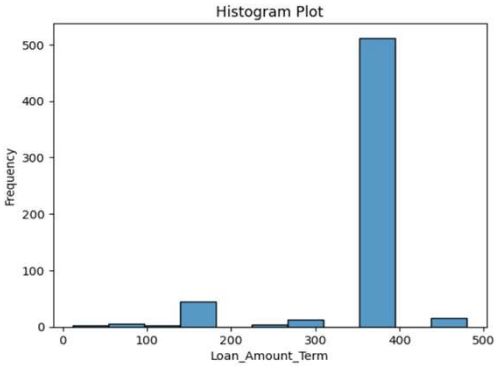
- EDA
- Cleaning/ selecting features
- Testing models
- Selecting model and deployment

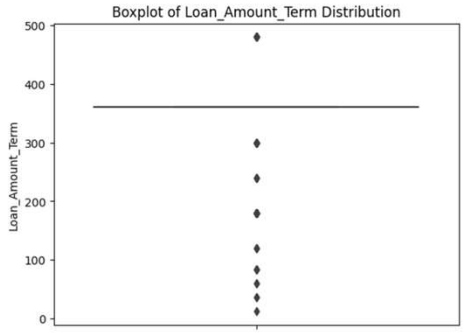
EDA insights snapshot











Features selection

	Education	Self_Employed	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status	family_size	total_income
0	Graduate	No	5.298317	360	1.0	Urban	Y	1	8.674026
1	Graduate	No	4.852030	360	1.0	Rural	N	3	8.714568
2	Graduate	Yes	4.189655	360	1.0	Urban	Y	2	8.006368
3	Not Graduate	No	4.787492	360	1.0	Urban	Y	2	8.505323
4	Graduate	No	4.948760	360	1.0	Urban	Y	1	8.699515

Models tested with Grid search

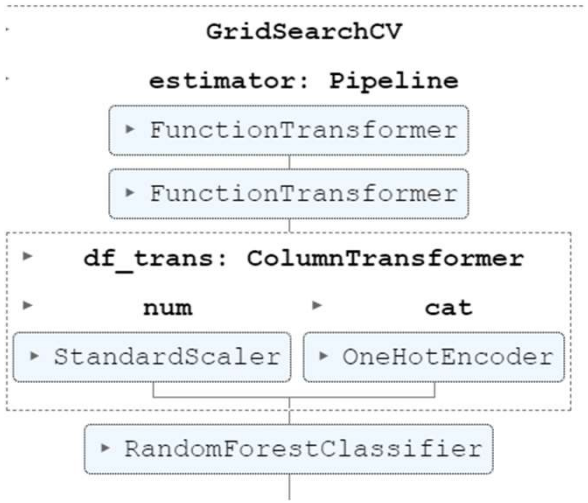
Model(with GS)	Train score	Test score
Random Forest C {'max_depth': 3, 'n_estimators': 100}	0.7983	0.8536
SVC {'C': 1, 'gamma': 0.1}	0.8004	0.8536
XGB classifier {'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 200}	0.8065	0.8699

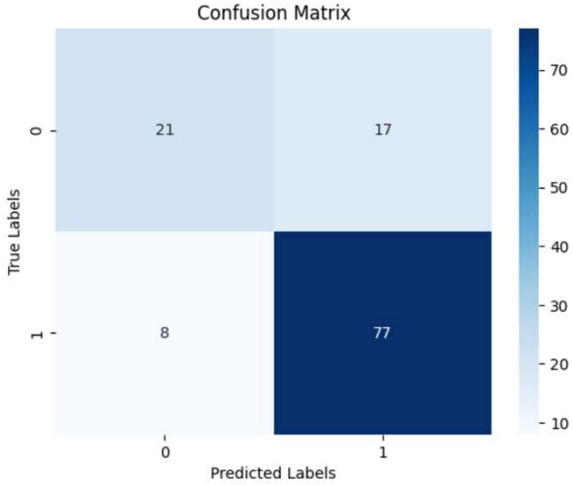
Pipeline results (RF)

```
Training Accuracy: 0.769857433808554
Testing Accuracy: 0.7967479674796748
Classification Report (Default Threshold):
              precision    recall  f1-score   support

           0       0.72      0.55      0.63        38
           1       0.82      0.91      0.86        85

    accuracy                           0.80       123
   macro avg       0.77      0.73      0.74       123
weighted avg       0.79      0.80      0.79       123
```





Deployment

```
json_data = [
    {
    "Loan_ID" : "LP001002",
    "Gender" : "Male",
    "Married" : "No",
    "Dependents" : "0",
    "Education" : "Graduate",
    "Self_Employed" : "No",
    "ApplicantIncome" : 5849,
    "CoapplicantIncome" : 200,
    "LoanAmount" : 1.0,
    "Loan_Amount_Term" : 360.0,
    "Credit_History" : 1.0,
    "Property_Area" : "Urban",
      }
]
```

```
(test_env) C:\Users\asmaa\OneDrive\Desktop\LHL\Projects\Project6\mini-project-IV\notebooks>python app1.py
 * Serving Flask app 'app1'
 * Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
[{'Loan_ID': 'LP001002', 'Gender': 'Male', 'Married': 'No', 'Dependents': '0', 'Education': 'Graduate', 'Self_Employed':
 'No', 'ApplicantIncome': 5849, 'CoapplicantIncome': 200, 'LoanAmount': 1.0, 'Loan_Amount_Term': 360.0, 'Credit_History'
: 1.0, 'Property_Area': 'Urban'}]
127.0.0.1 - - [26/May/2023 16:10:27] "POST /predict HTTP/1.1" 200 -
127.0.0.1 - - [26/May/2023 16:11:15] "GET /predict HTTP/1.1" 200 -
127.0.0.1 - - [26/May/2023 16:11:22] "POST /predict HTTP/1.1" 200 -
{'Loan_ID': 'LP001002', 'Gender': 'Male', 'Married': 'No', 'Dependents': '0', 'Education': 'Graduate', 'Self_Employed':
'No', 'ApplicantIncome': 5849, 'CoapplicantIncome': 200, 'LoanAmount': 1.0, 'Loan_Amount_Term': 360.0, 'Credit_History':
 1.0, 'Property_Area': 'Urban'}
127.0.0.1 - - [26/May/2023 16:21:46] "POST /predict HTTP/1.1" 200 -
```

```
PS C:\Users\asmaa\OneDrive\Desktop\LHL\Projects\Project6\mini-project-IV\notebooks> & C:/Users/asmaa/anaconda3/python.exe c:/Users/asmaa/OneDrive/Desktop/LHL/Projects/P
oject6/mini-project-IV/notebooks/test.py
...
request successful
...
{'prediction': '[1]'}
```

Future steps

- Pipeline with XGB classifier
- Review model best F1 score instead of accuracy
- Try a sequential model
- More EDA
- Collect more data from Clients
- Time