loan approval prediction

Aziz Wafi

Outline

- Introduction
- Proposed Vision
- Potential Impact
- Introduction to Dataset
- Next Steps

LOAN
APPROVAL

Introduction

- **In the domain of loan approval prediction, the challenge is to determine whether a loan application should be approved or rejected.**

- **This problem holds significant importance in minimizing risks and enhancing decision-making processes.**

Proposed Vision

- **Our approach involves leveraging machine learning to provide solution by building predictive models that analyze historical loan data and various applicant attributes.**

- **By employing these methods, we aim to improve the overall decision-making process in loan approval scenarios.**

Potential Impact

- **Anticipated impact includes reducing default rates, optimizing loan approval processes, and enhancing overall business value.**

- **Quantifiable measures, such as potential cost savings or increased efficiency, will highlight the tangible benefits of our solution.**

Introduction to Dataset

- **Datasets for this project may include historical loan data with details on approved and denied applications, along with attributes such as credit scores, income, and employment status.**

- **I will use "Loan-Approval-Prediction-Dataset" which is published public on Kaggle.**

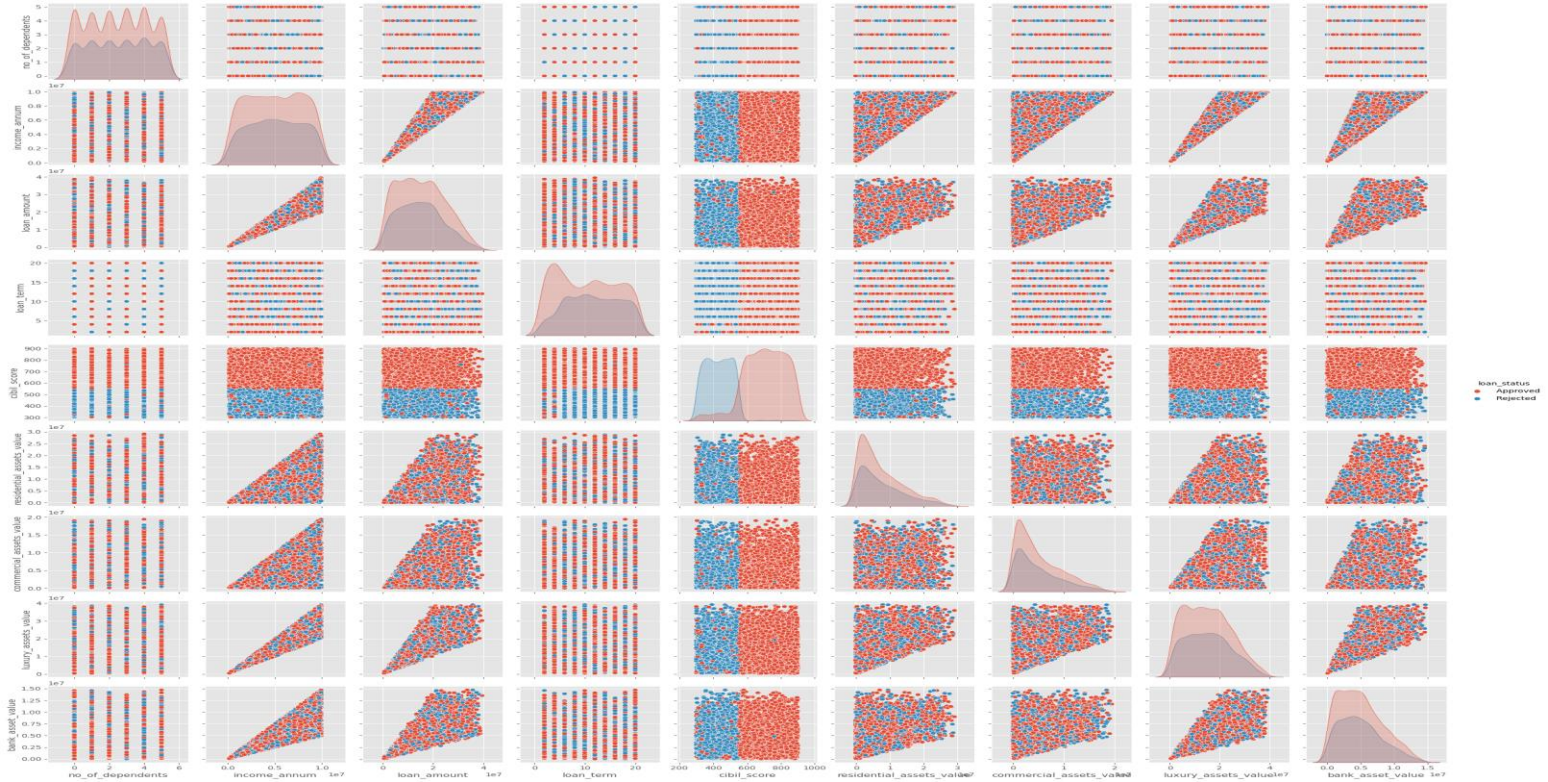
loan_id	no_of_dependents Number of Dependents of the Applicant	education Education of the Applicant	self_employed Employment Status of the Applicant	income_annum Annual Income of the Applicant	loan_amount Loan Amount	loan_term Loan Term in Years	cibil_score Credit Score
1 4269	0 5	Graduate 50% Not Graduate 50%	true 0 0% false 0 0%	200k 9.90m	300k 39.5m	2 20	300
1	2	Graduate	No	9600000	29900000	12	778
2	0	Not Graduate	Yes	4100000	12200000	8	417
3	3	Graduate	No	9100000	29700000	20	506
4	3	Graduate	No	8200000	30700000	8	467
5	5	Not Graduate	Yes	9800000	24200000	20	382
6	0	Graduate	Yes	4800000	13500000	10	319
7	5	Graduate	No	8700000	33000000	4	678
8	2	Graduate	Yes	5700000	15000000	20	382
9	0	Graduate	Yes	800000	2200000	20	782
10	5	Not Graduate	No	1100000	4300000	10	388
11	4	Graduate	Yes	2900000	11200000	2	547
12	2	Not Graduate	Yes	6700000	22700000	18	538
13	3	Not Graduate	Yes	5000000	11600000	16	311

Introduction to Dataset

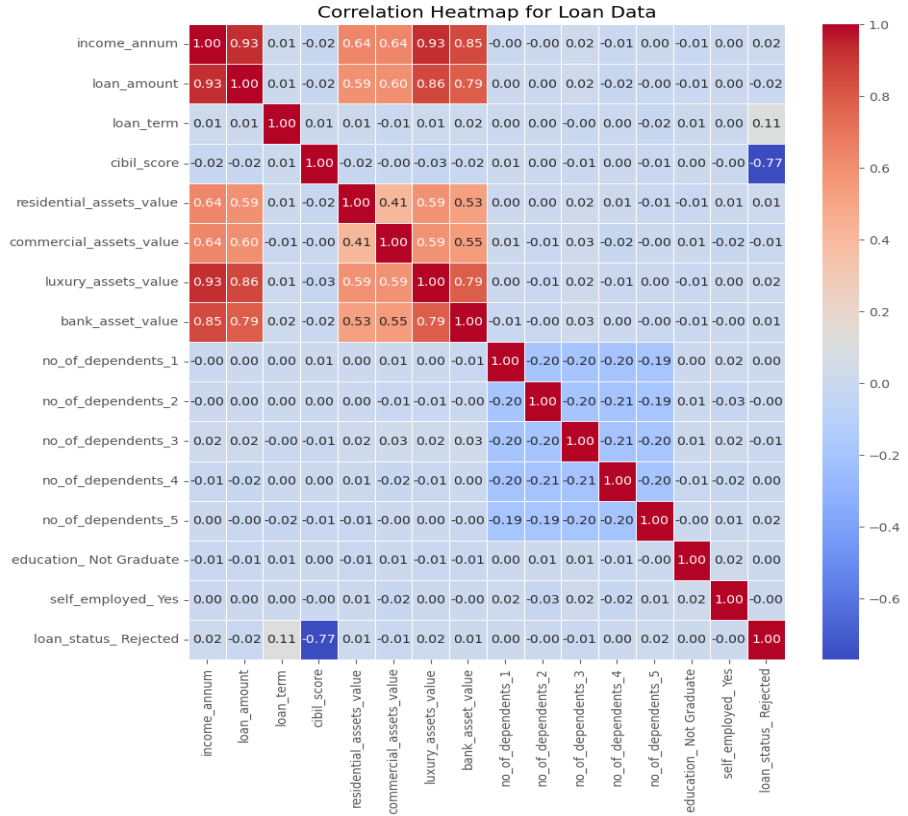
Preprocessing and EDA steps:

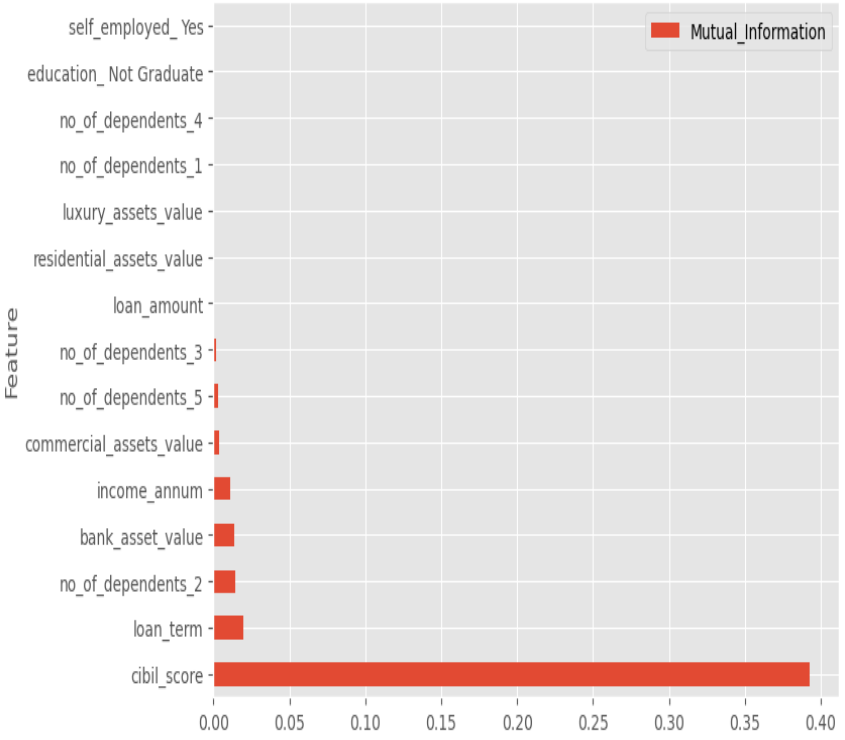
- Trailing whitespaces from all column names and remove non important columns.
- Type casting, missing values and duplicated checking.
- Summary statistics, distributions analysis, and correlation analysis.
- One hot encoding and Normalization.

Introduction to Dataset



Introduction to Dataset





Potential Impact

So, based on our analysis, we can conclude that the most important features influencing the loan approval decision are:

- **Cibil score, Long term, Bank asset value, Annual income, Commercial assets value, Number of dependents.**
- **It appears that these features significantly contribute to distinguishing between approved and rejected loan applications. On the other hand, other features do not provide substantial new information for our problem.**

Next Steps

- Split data
- Baseline model (Navie method)
- Logistic regression
- Ensemble (random forest)