

By : Niken Iarasati Winasih

LOAN DATA PREDICTION ANALYSIS

AGENDA

- Background
- Objective
- Data Collection
- Data Understanding
- Data preprocessing
- Modeling
- Insight

The background is a light gray with a subtle, crumpled paper texture. In the top right corner, there is a small orange outline circle and a larger solid orange circle. In the bottom left corner, there is a large solid orange circle partially visible.

BACKGROUND

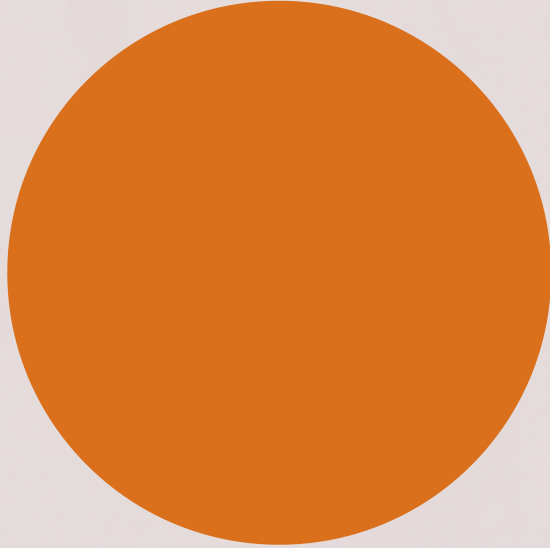
Checking the suitability of prospective borrowers against predetermined criteria

Analyze the credit risk at which the borrower will return the interest and principal rates

OBJECTIVE



conduct analysis on loan
customers

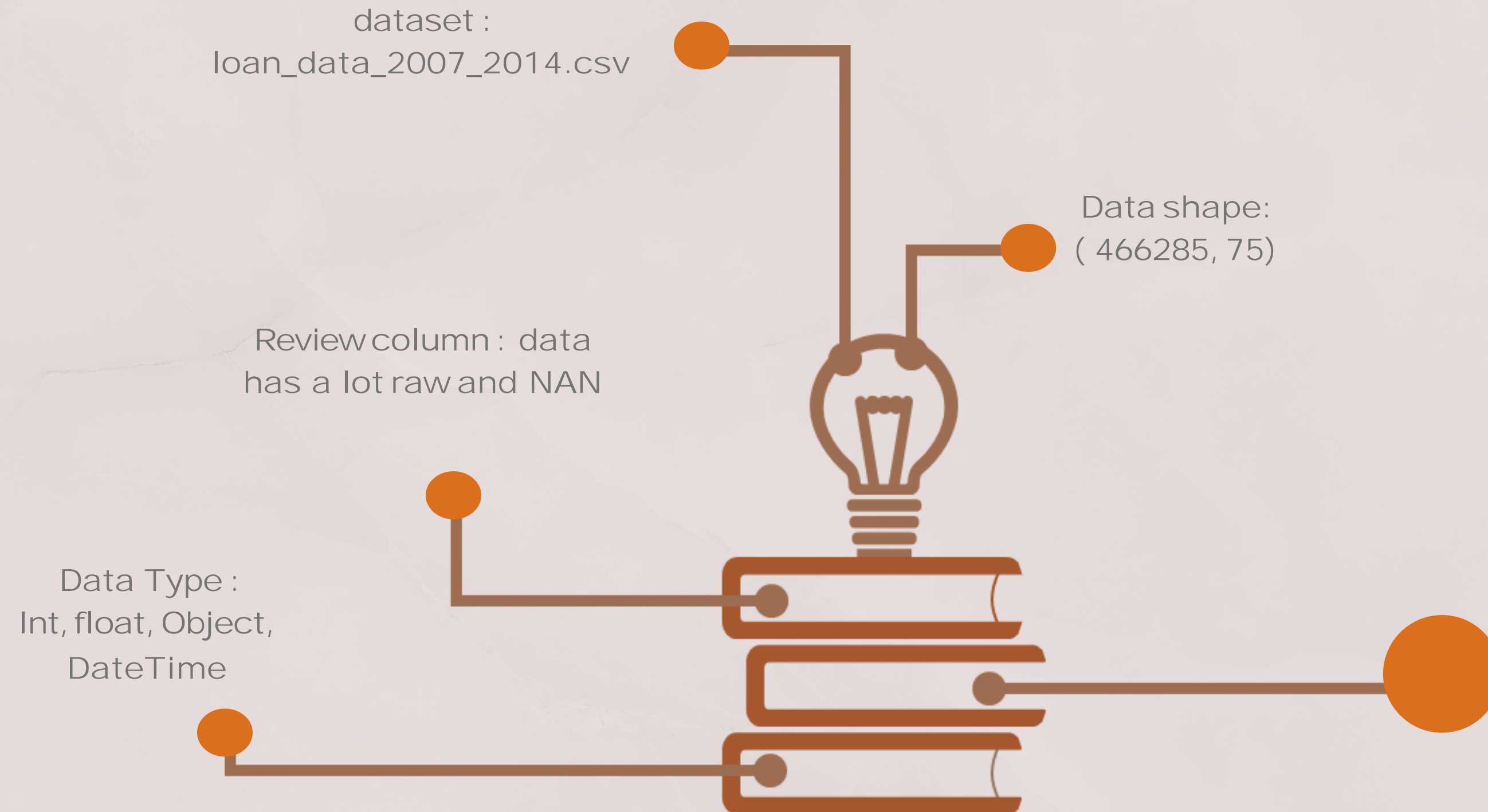


create a loan data
prediction analysis
model



provide insight and
formulate solutions

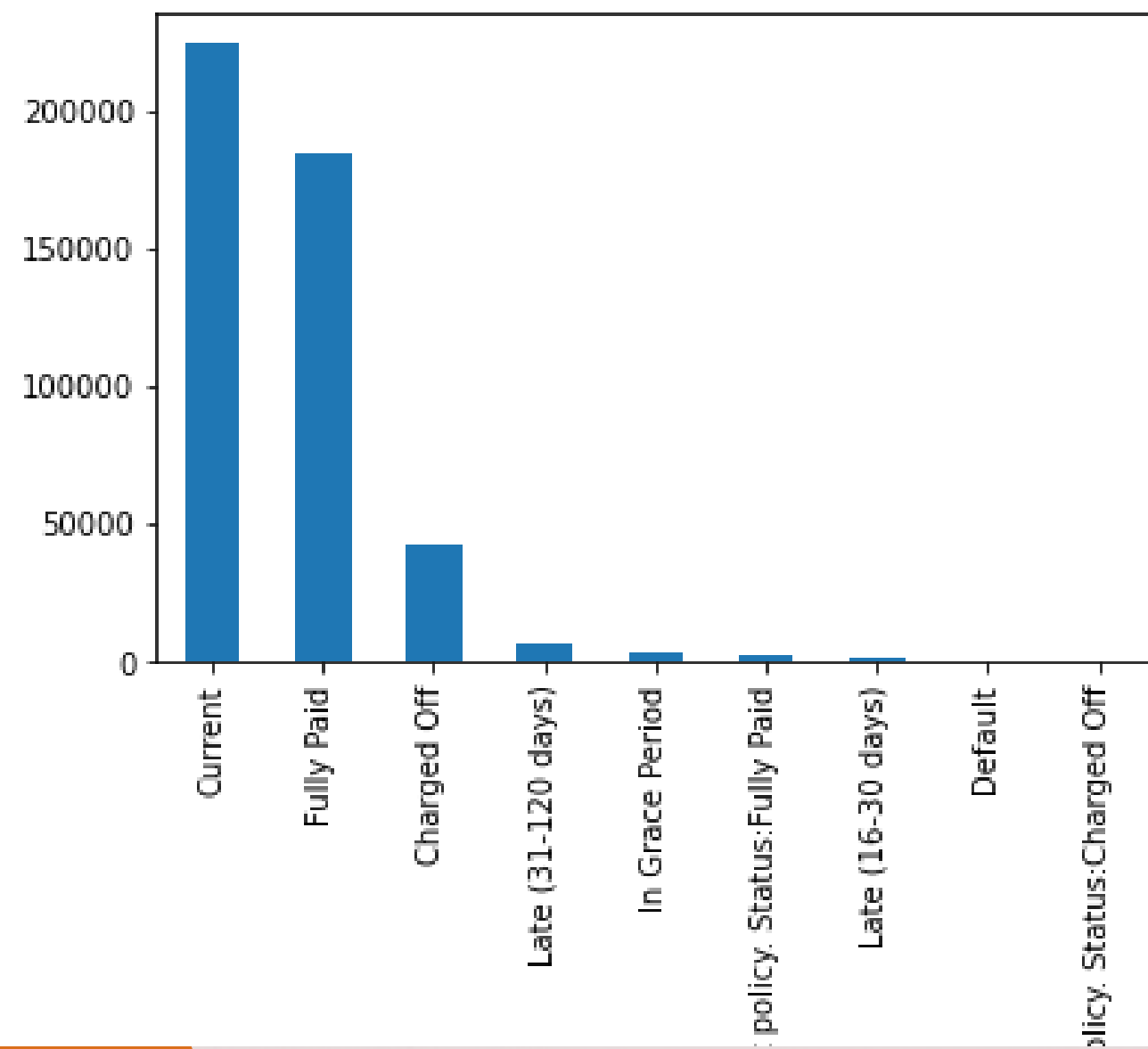
DATA COLLECTION



DATA UNDERSTANDING

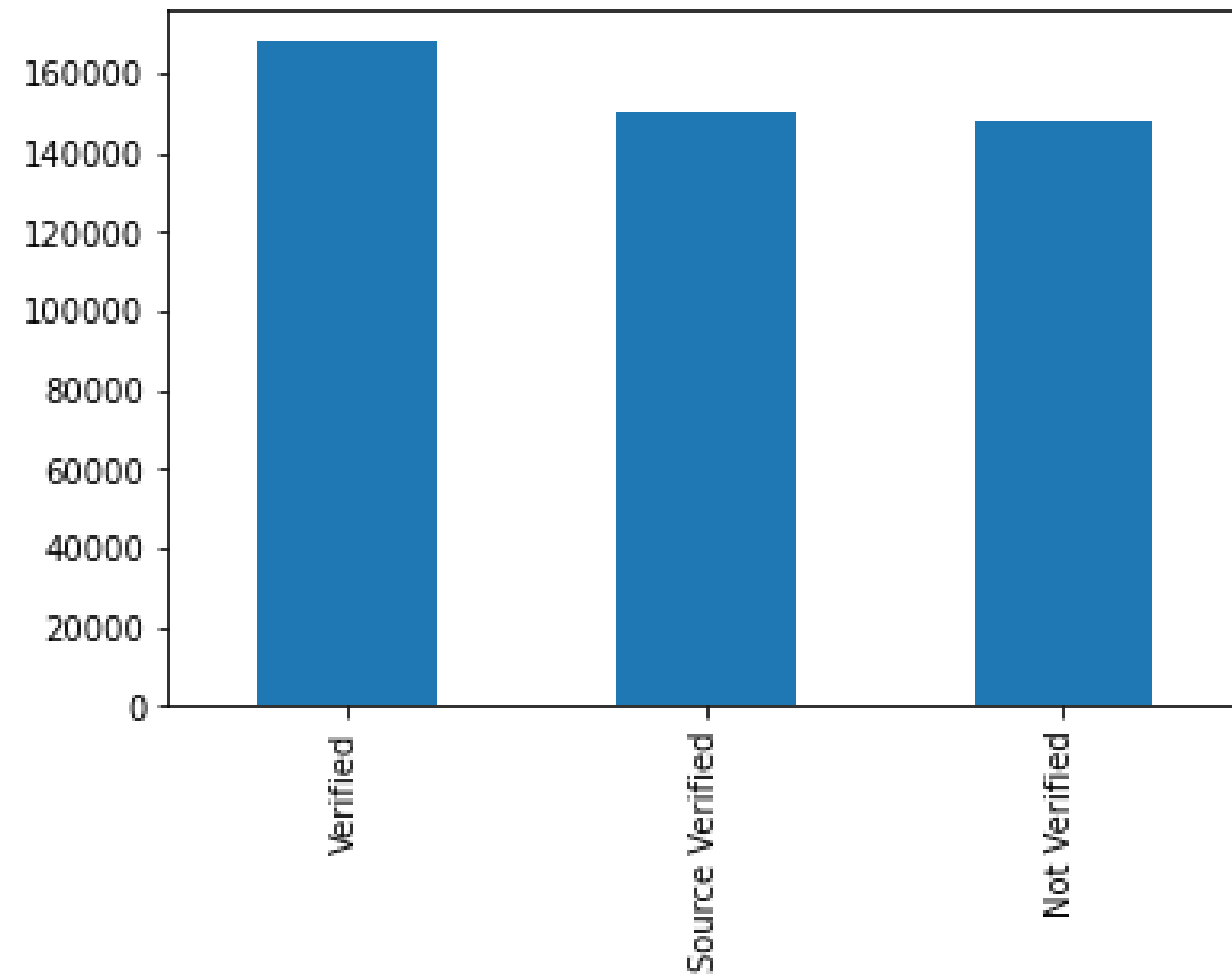
loan status

[13]: <AxesSubplot:>



```
Current                224226
Fully Paid             184739
Charged Off            42475
Late (31-120 days)     6900
In Grace Period        3146
Does not meet the credit policy. Status:Fully Paid  1988
Late (16-30 days)      1218
Default                832
Does not meet the credit policy. Status:Charged Off  761
Name: loan_status, dtype: int64
```

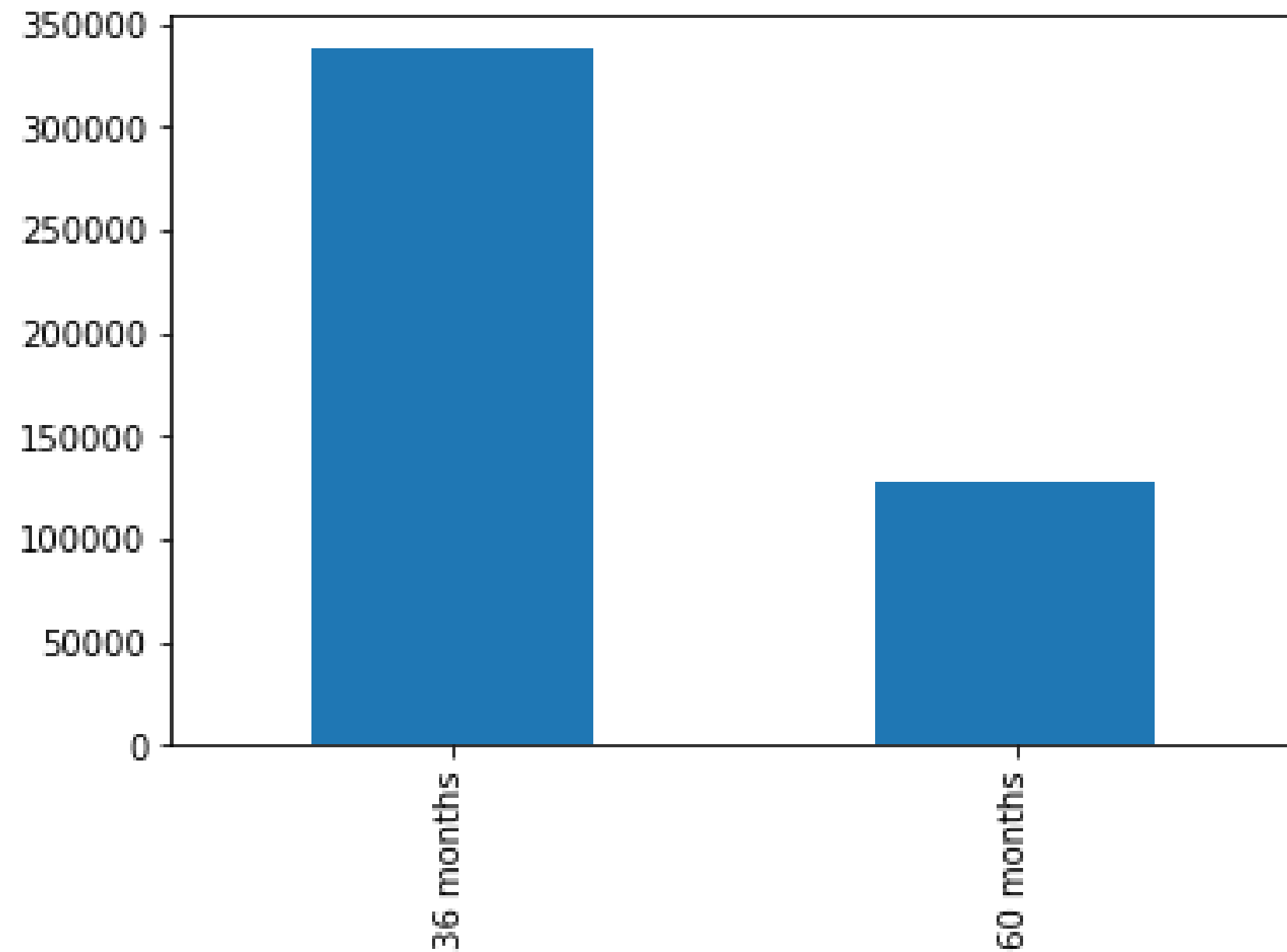

verivication status



```
Verified          168055
Source Verified   149993
Not Verified      148237
Name: verification_status, dtype: int64
```

term

<AxesSubplot:>

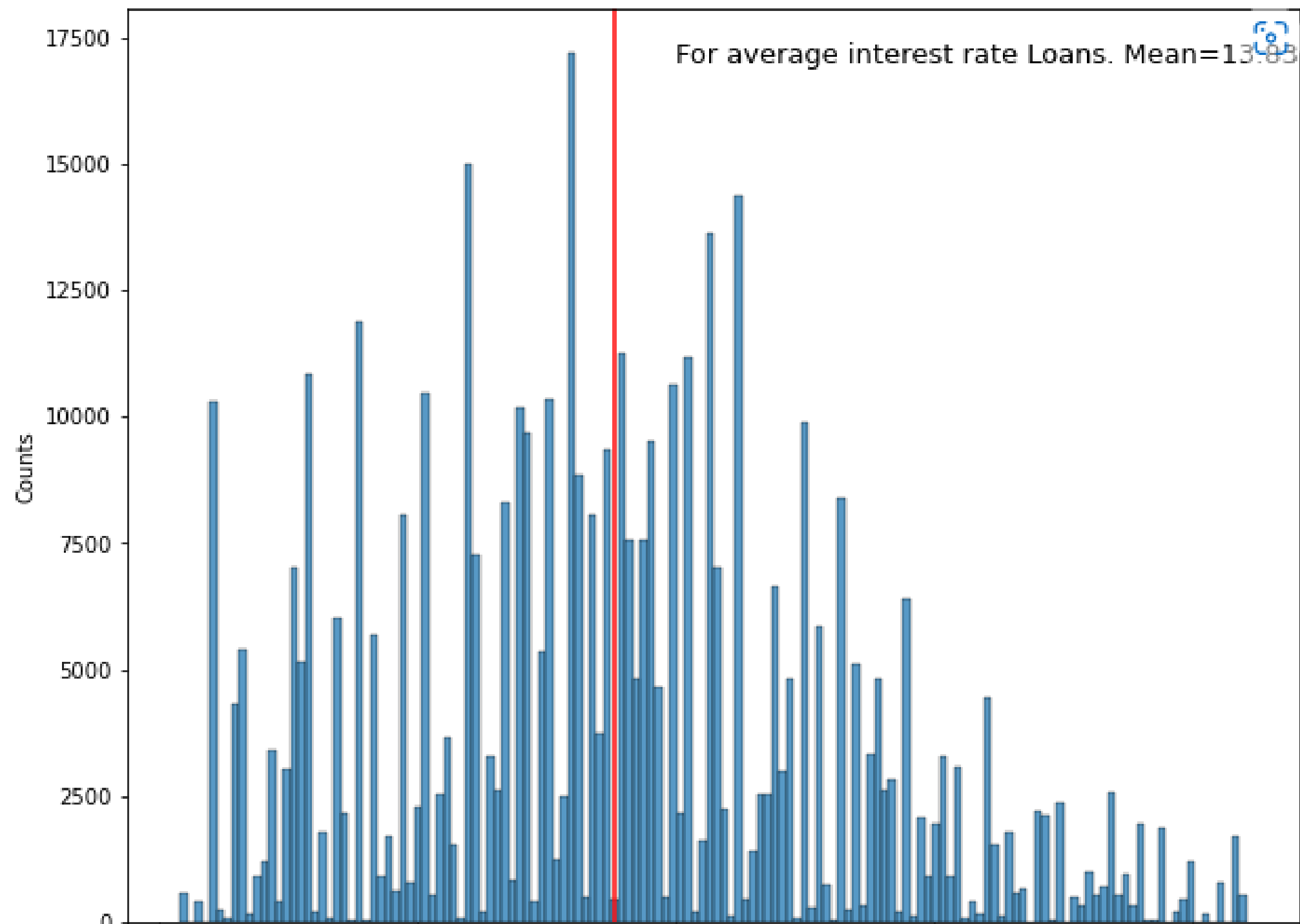


36 months 337953

60 months 128332

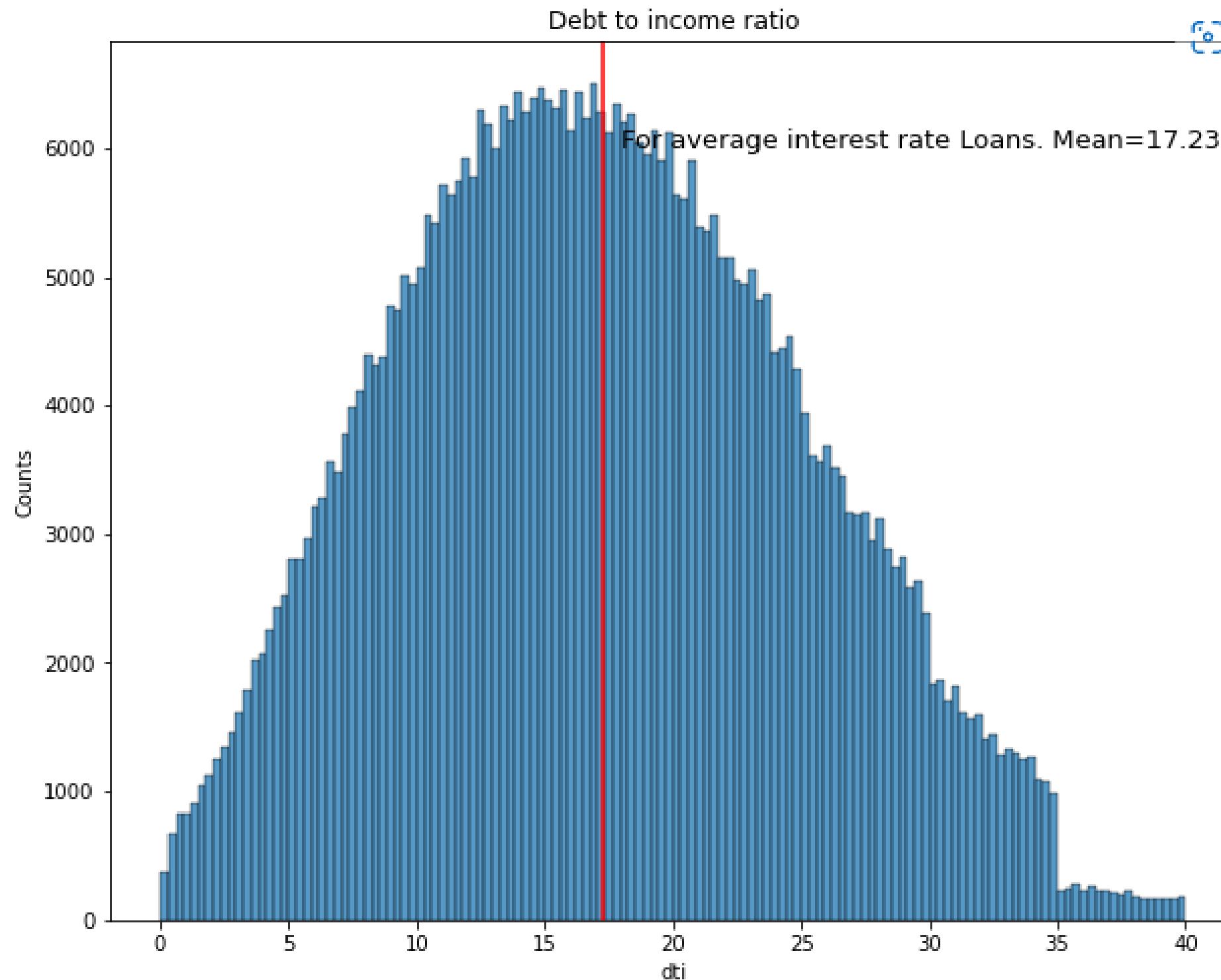
Name: term, dtype: int64

Average interest rate Loans



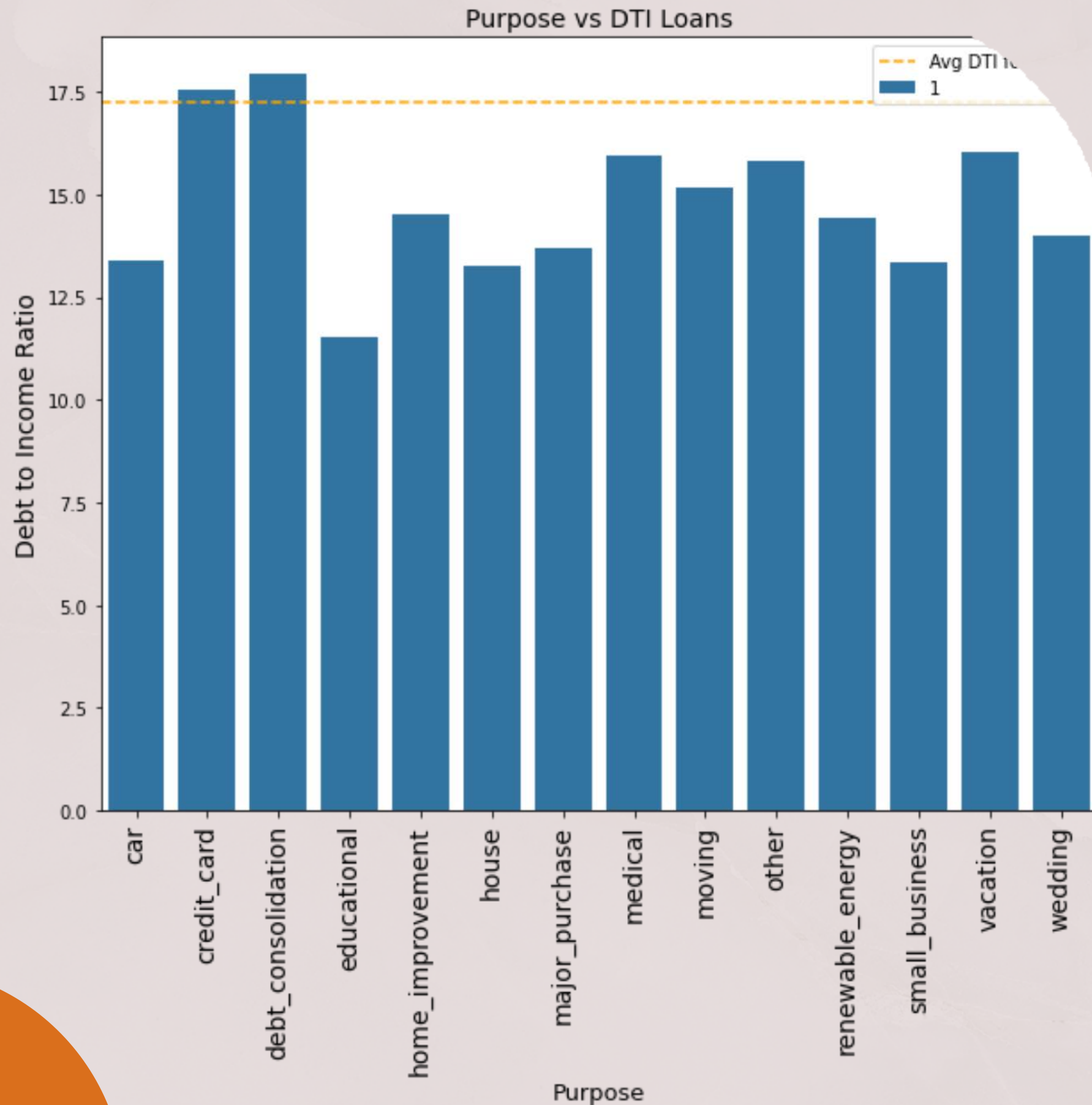
rata rata bunga adalah sebanyak 13%

Average interest rate Loans



Rasio Hutang terhadap Pendapatan sebesar 17%

Purpose vs DTI



The higher the DTI ratio, the higher the risk.

Debt consolidation and credit card debt have above-average debt-to-income ratios.

DATA PREPROCESSING



Handling missing
value di emp_title
dengan 'other'



fill the nan column
in dti with 0



fill the nan column
in funded_amnt
with 0

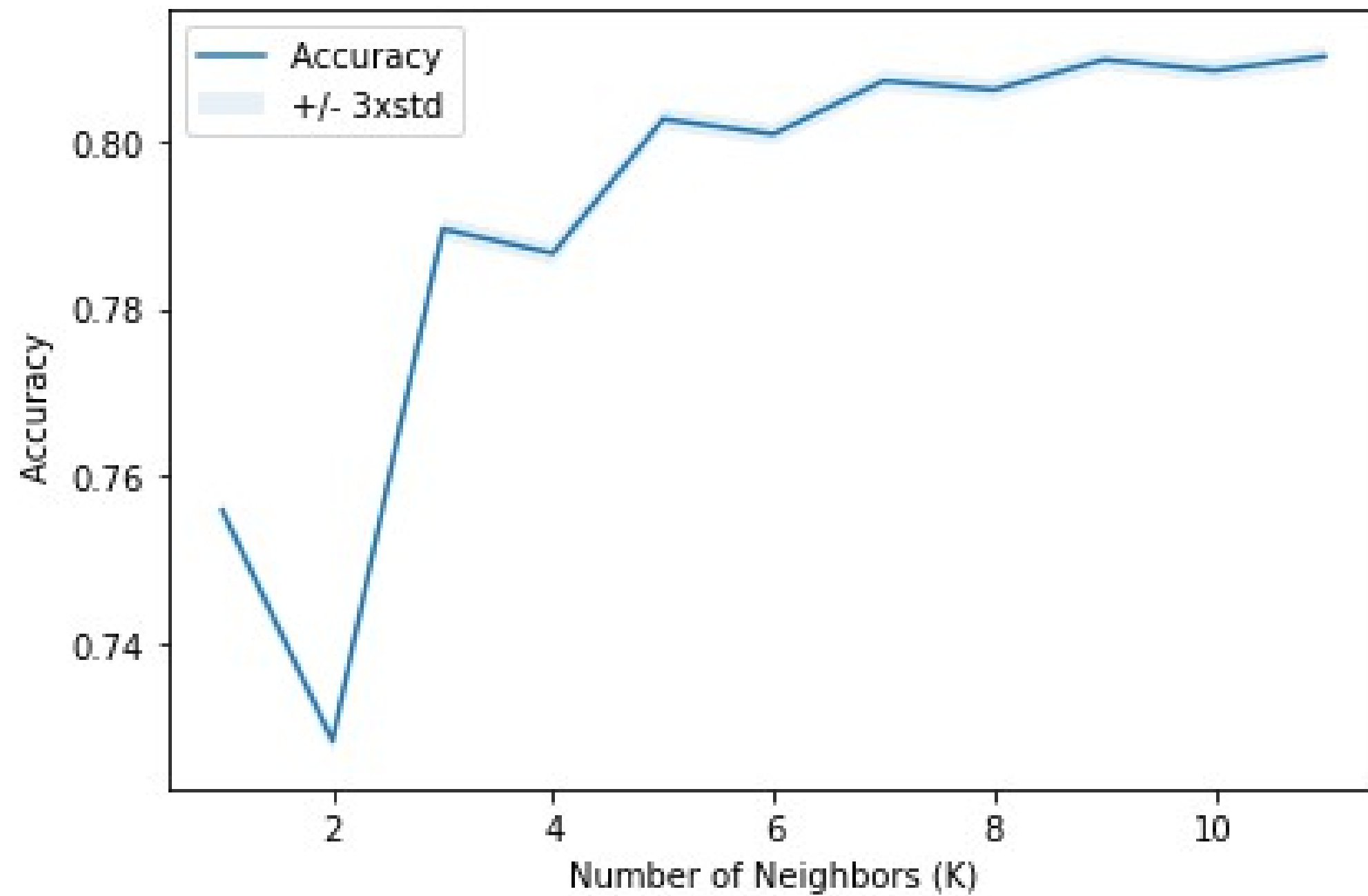


scaling



MODELING

With KNN

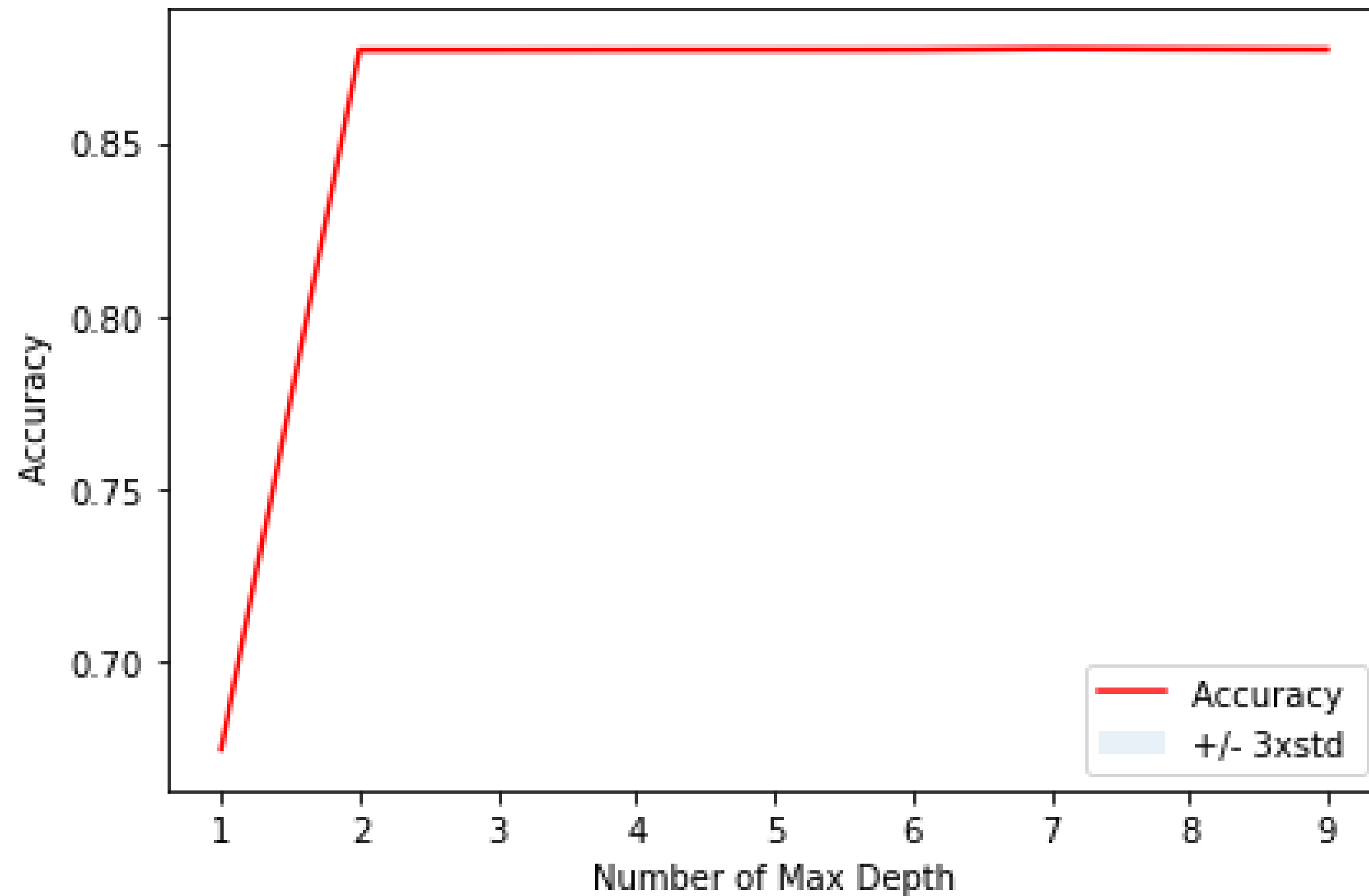


The best accuracy was with 0.8101268537482441 with k= 11

Train set Accuracy: 0.8394034764146391

Test set Accuracy: 0.8072316287249215

With Decision tree



Train set Accuracy: 0.8770360401900126

Test set Accuracy: 0.876877875119294

Debt Consolidation was identified as a primary risk factor, accounting for ~58% of identified high-risk cases.

Recommended a refined policy_code indexing system to distinguish between high-risk (0) and healthy (1) loan applications.

Feature independence was verified through correlation matrix analysis, ensuring model stability and avoiding multicollinearity.

Occupational analysis revealed a higher propensity for debt consolidation among specific professions, such as educators.

Recommended a refined policy_code indexing system to distinguish between high-risk (0) and healthy (1) loan applications.



INSIGHT



The background is a light gray with a subtle, crumpled paper texture. There are five orange circles of varying sizes and styles: a large solid circle in the top right, a medium solid circle in the bottom left, a large hollow circle in the top left, a medium hollow circle in the bottom right, and a small hollow circle in the middle right.

THANK YOU