

Bank Loan Prediction

Brandon Lotus



Background

- Meningkatnya pengajuan pinjaman di bank dapat mengakibatkan masalah karena memiliki resiko yaitu Non-performing loan (NPL) atau kredit macet. NPL masih menjadi masalah utama di sektor perbankan dalam kurun waktu beberapa dekade terakhir.
- Penyebab paling sering tingginya NPL adalah kegagalan bank untuk mengidentifikasi serta memberikan keputusan debitur mana yang pantas dan mampu diberikan pinjaman.
- Di sisi lain, bank yang terlalu mengejar target penyaluran kredit menyebabkan pihak bank kurang memperhatikan aspek analisis kredit, seperti pengelolaan informasi yang kurang baik, kurang efektifnya kebijakan dalam pemberian kredit, dan SOP analisis kredit yang buruk.

— Apa itu Non-Performing Loan (NPL)?

- Non Performing Loan (NPL) adalah pinjaman dimana peminjam belum membayar cicilan bulanan jangka waktu tertentu. NPL terjadi ketika peminjam sulit untuk terus melakukan pembayaran kembali.
- Umumnya bank mengklasifikasikan pinjaman sebagai NPL ketika pembayaran jatuh tempo lebih dari 90 hari atau tergantung pada ketentuan perjanjian pinjaman.
- Ketika suatu pinjaman diklasifikasikan sebagai NPL, kemungkinan menerima pembayaran kembali secara signifikan lebih rendah.

— Apa dampak NPL pada bank?

- Bank menghasilkan uang dari bunga yang mereka kenakan atas pinjaman. Ketika bank tidak mendapatkan pembayaran bunga yang terhutang dari NPL, berarti bank memiliki lebih sedikit uang untuk membuat pinjaman baru.
- Ketika persentase NPL meningkat, investasi menjadi kurang menarik bagi calon investor karena profit di masa depan akan terganggu jika bank tidak memperoleh pendapatan dari bisnis kreditnya.

— Berapa potential loss yang bisa diterima bank?

- Di dalam dataset terdapat 67617 peminjam dengan tingkat charged off 26.7% (18.046 peminjam)
- Jumlah pinjaman terkecil dalam dataset \$15.000
- Interest rate di USA yaitu sekitar 3% (\$450)
- Potential loss yang dapat diterima oleh bank adalah :
- $\$450 \times 18.046 = \$8.120.700$ (~ Rp. 14 M)

— Bagaimana bank mencegah NPL?

- Sebelum memberikan pinjaman, umumnya bank melakukan assesment untuk memeriksa apakah peminjam memiliki potensi gagal bayar. Proses assesment secara manual akan sulit dilakukan ketika menangani jumlah data yang sangat banyak.
- Dalam project ini diharapkan dapat membantu proses assesment dengan menggunakan model machine learning. Model akan melakukan klasifikasi antara peminjam yang akan melunasi pinjaman atau yang akan menjadi pinjaman gagal.

Problems

- Apakah faktor-faktor yang dapat mengidentifikasi peminjam yang bisa / tidak melunasi pinjamannya?
- Bagaimana mengurangi resiko kerugian bank yang disebabkan oleh peminjam yang tidak melunasi pinjamannya?



Goals

- Mengidentifikasi faktor-faktor seorang peminjam bisa / tidak melunasi pinjamannya.
- Membuat model untuk memprediksi apakah peminjam akan membayar kembali pinjamannya berdasarkan fitur target 'Loan Status' dalam dataset.



Project Stages

01

Data Cleaning Preprocessing

- Check Data Types
- Handle Missing Values
- Describe Data
- Handle Outliers

02

Data Analysis Visualization

- Univariate Data Analysis
- Univariate Data Visualization
- Multivariate Data Analysis
- Multivariate Data Visualization
- Insights

Project Stages

03

Feature Engineering

- Encoding
- Recategorise
- Check Correlation
- Feature Selection
- Handle Imbalance
- Scaling

04

Machine Learning

- Split Data
- Model Training
- Model Tuning
- Evaluation Matrix
- Model Comparison

Project Stages

05

Dashboard

- Prediction
- Dataset
- Visualization

06

Conclusion Recommendation

- Conclusion
- Recommendation

Dataset

Loan Status	Current Loan Amount	Term	Credit Score	Annual Income	Years in current job	Home Ownership	Purpose	Monthly Debt	Years of Credit History	Number of Open Accounts	Credit Problems	Current Credit Balance	Maximum Open Credit	Bankruptcies	Tax Liens
Fully Paid	445412	Short Term	709	1167493	8	Home Mortgage	Home Improvements	5215	17	6	1	228190	416746	1	0
Fully Paid	262328	Short Term	726	1172718	10	Home Mortgage	Debt Consolidation	33296	21	35	0	229976	850784	0	0
Fully Paid	347666	Long Term	721	806949	3	Own Home	Debt Consolidation	8742	12	9	0	256329	386958	0	0
Fully Paid	176220	Short Term	726	1172718	5	Rent	Debt Consolidation	20640	6	15	0	253460	427174	0	0
Fully Paid	217646	Short Term	730	1184194	0	Home Mortgage	Debt Consolidation	10855	20	13	1	122170	272052	1	0
...
Fully Paid	781022	Short Term	741	2582404	6	Home Mortgage	Debt Consolidation	51648	26	25	0	669180	2299198	0	0
Charged Off	309474	Short Term	726	1172718	10	Home Mortgage	Debt Consolidation	13817	27	15	0	225872	892606	0	0
Charged Off	429132	Short Term	726	1172718	10	Home Mortgage	Debt Consolidation	28948	16	16	0	485279	656414	0	0
Fully Paid	103136	Short Term	742	1150545	6	Rent	Debt Consolidation	7316	19	12	1	109554	537548	1	0
Fully Paid	530332	Short Term	746	1717524	9	Rent	Debt Consolidation	9890	15	8	0	404225	738254	0	0

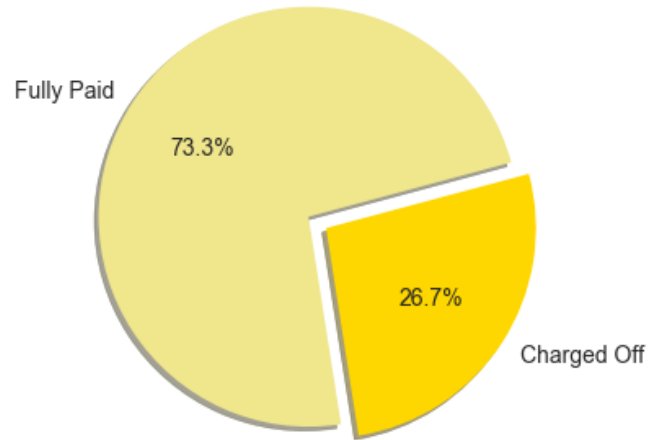
Dataset yang digunakan adalah 'Bank Loan Status Dataset' dari Kaggle yang berisi data historis peminjaman dari suatu bank. Data peminjaman merupakan pinjaman yang sudah ditutup baik karena jatuh tempo untuk pinjaman telah dilunasi atau charged off (ketika bank telah menentukan bahwa peminjam tidak akan pernah melunasi pinjaman mereka).

Dataset

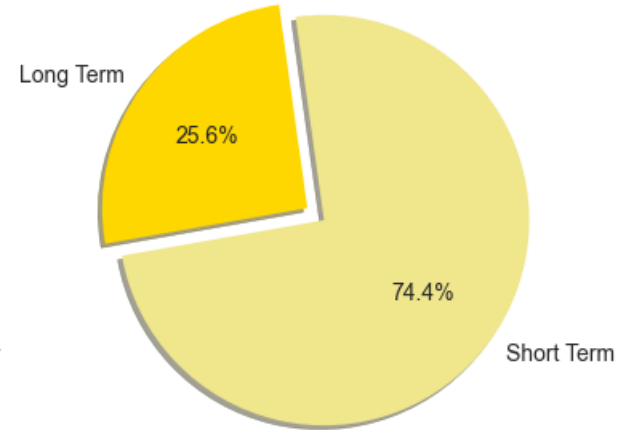
	Data Feature	Data Types	Null	Null Pct	Unique	Unique Sample
0	Loan ID	object	514	1	81999	[8236d82a-5679-40e7-b22a-07ec137ce399, 0565e42...
1	Customer ID	object	514	1	81999	[0fbe9c43-c7c2-40b3-aa9c-435ab6b0bde8, 916131c...
2	Loan Status	object	514	1	2	[nan, Fully Paid]
3	Current Loan Amount	float64	514	1	22004	[605902.0, 227524.0]
4	Term	object	514	1	2	[Short Term, Long Term]
5	Credit Score	float64	19668	20	324	[610.0, 724.0]
6	Annual Income	float64	19668	20	36174	[414295.0, 1914554.0]
7	Years in current job	object	4736	5	11	[7 years, 2 years]
8	Home Ownership	object	514	1	4	[Rent, nan]
9	Purpose	object	514	1	16	[renewable_energy, moving]
10	Monthly Debt	float64	514	1	65765	[5461.55, 7491.32]
11	Years of Credit History	float64	514	1	506	[50.4, 16.1]
12	Months since last delinquent	float64	53655	53	116	[69.0, 39.0]
13	Number of Open Accounts	float64	514	1	51	[25.0, 12.0]
14	Number of Credit Problems	float64	514	1	14	[4.0, 0.0]
15	Current Credit Balance	float64	514	1	32730	[624853.0, 159144.0]
16	Maximum Open Credit	float64	516	1	44596	[599852.0, 687368.0]
17	Bankruptcies	float64	718	1	8	[2.0, 6.0]
18	Tax Liens	float64	524	1	12	[7.0, nan]

Exploratory Data Analysis

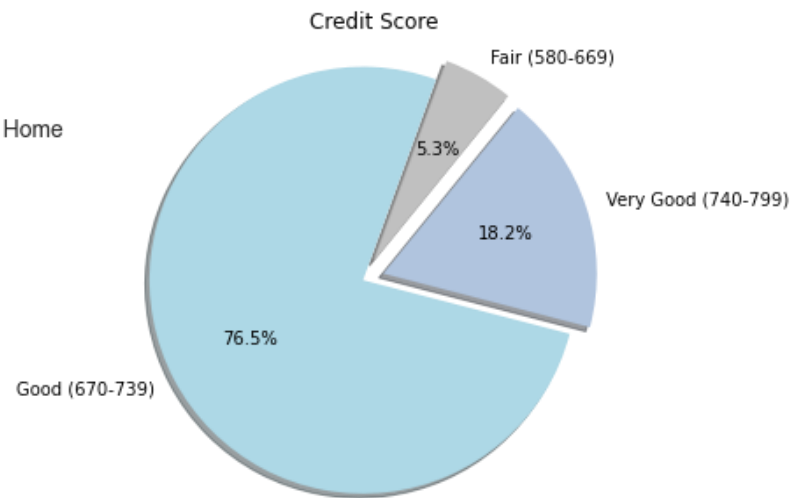
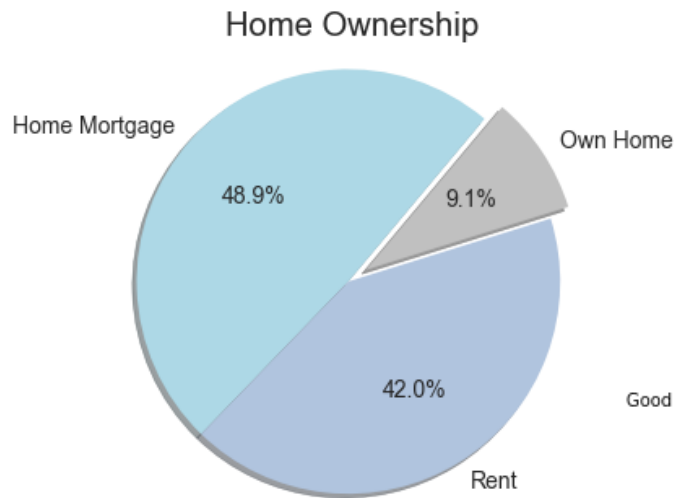
Loan Status



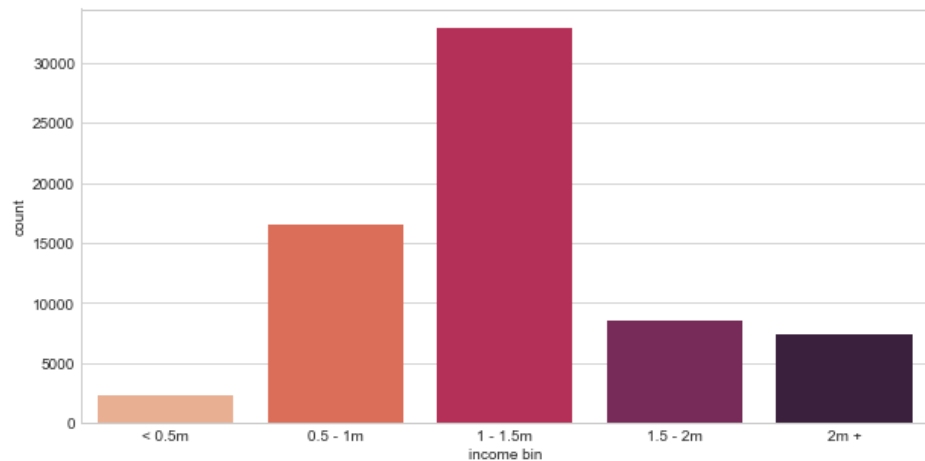
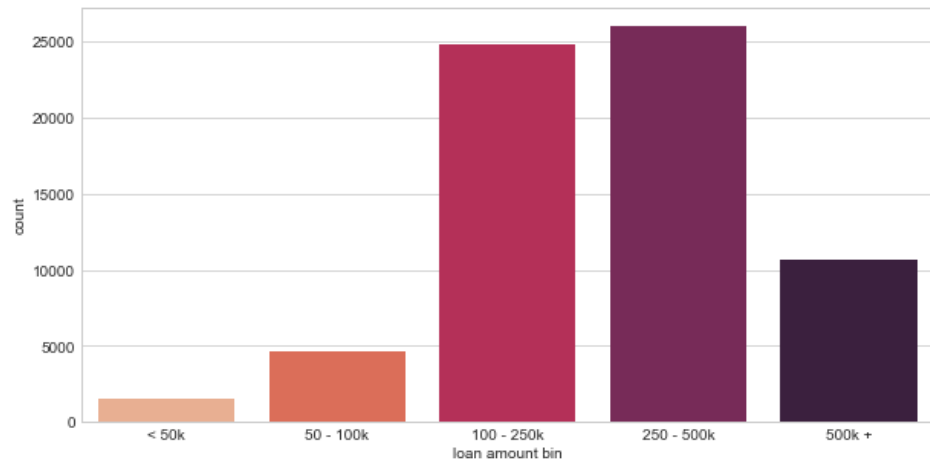
Term



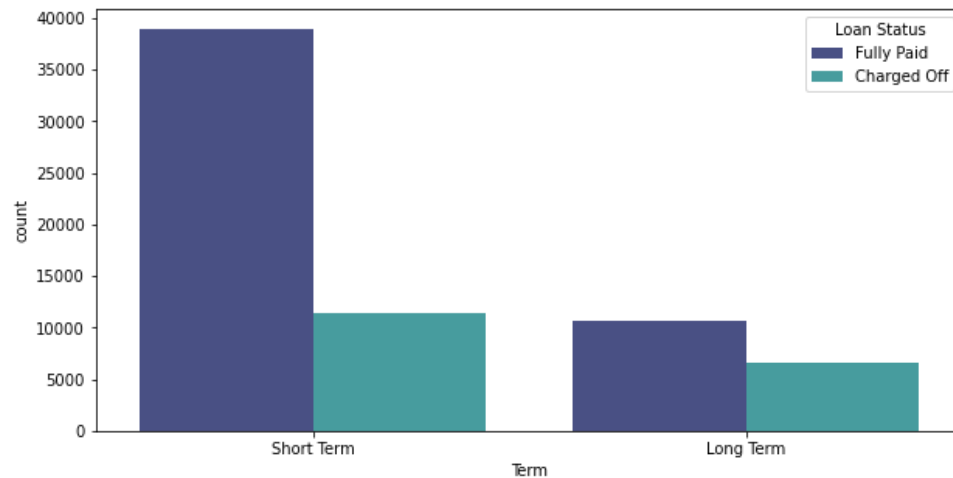
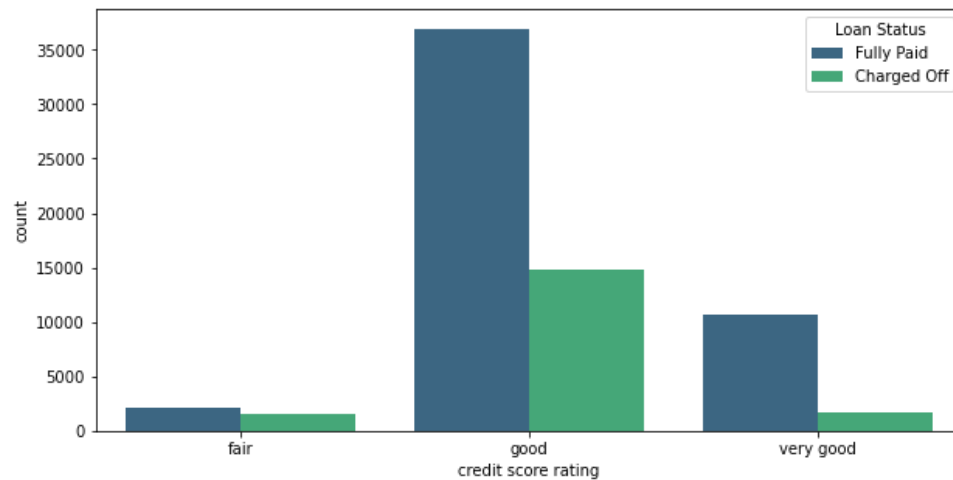
Exploratory Data Analysis



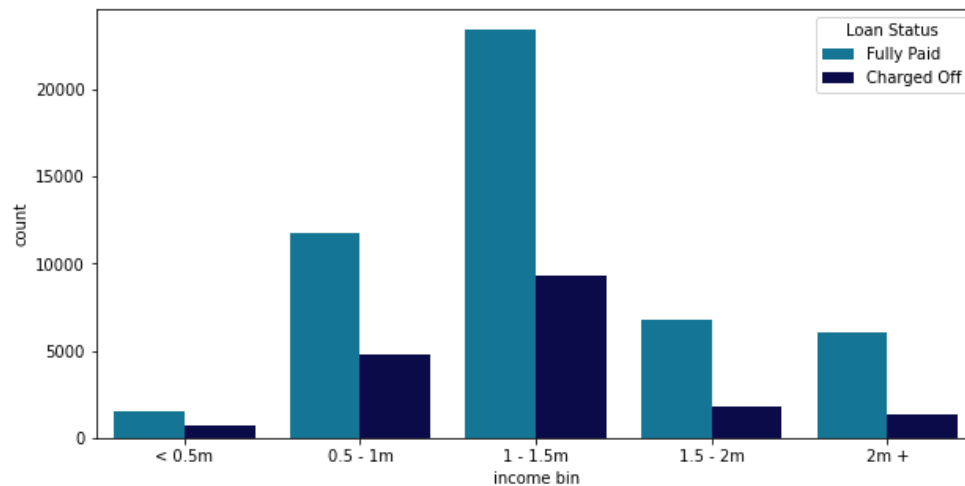
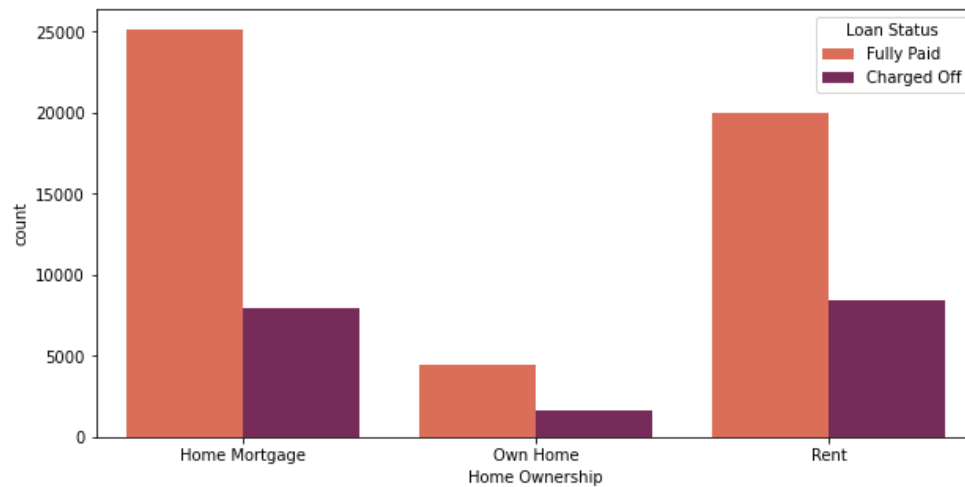
Exploratory Data Analysis



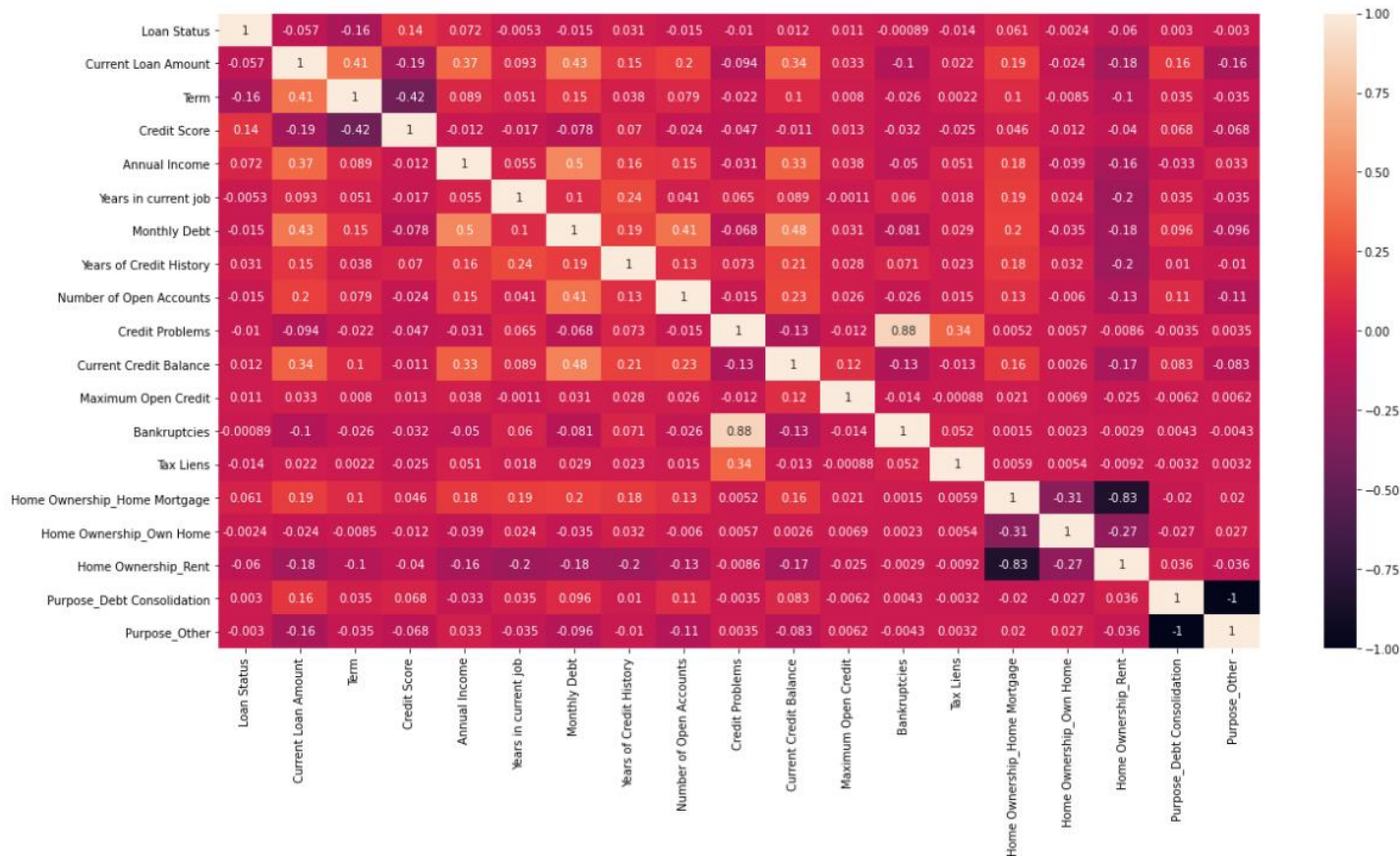
Exploratory Data Analysis



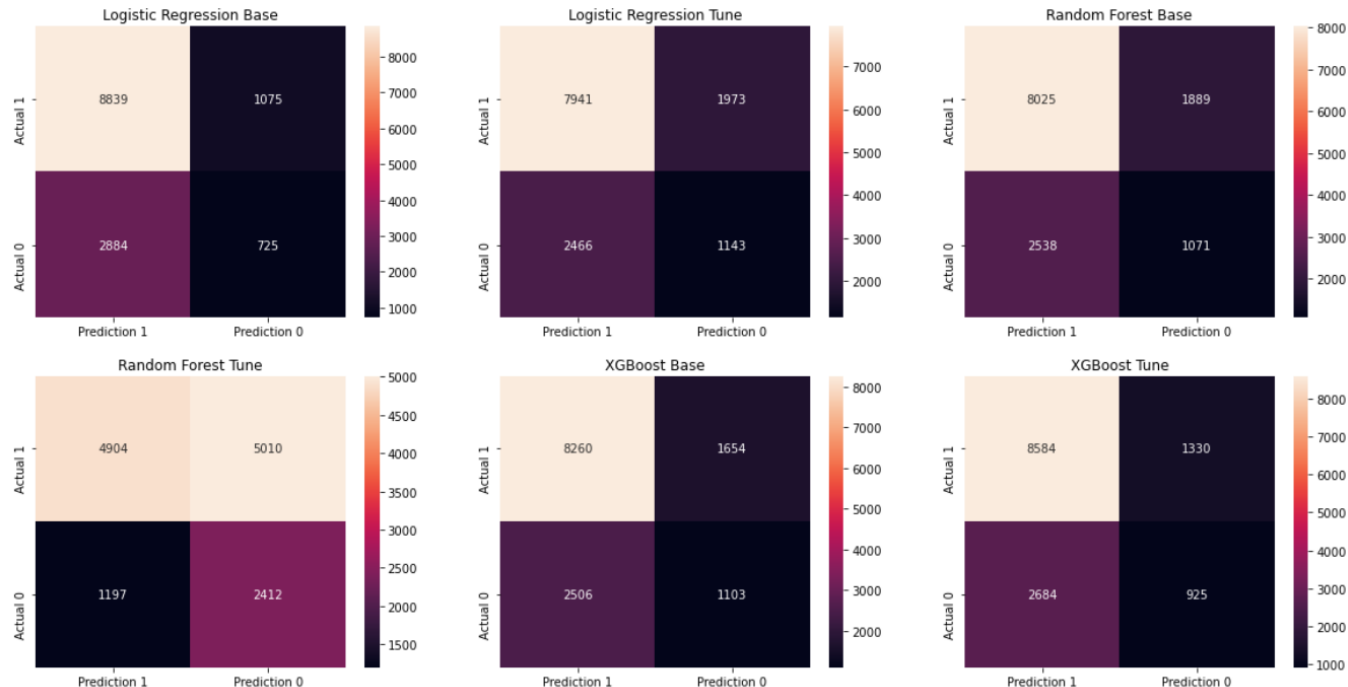
Exploratory Data Analysis



Feature Correlation



Model Comparison



	LogReg Base	LogReg Tune	Random Forest Base	Random Forest Tune	XGBoost Base	XGBoost Tune
Accuracy	0.707240	0.671744	0.672632	0.541004	0.692376	0.703172
Precision	0.753988	0.763044	0.759727	0.803803	0.767230	0.761803
Recall	0.891567	0.800989	0.809461	0.494654	0.833165	0.865846
F1-Score	0.817026	0.781556	0.783806	0.612426	0.798839	0.810499

- Model machine learning Random Forest setelah dilakukan hyperparameter tuning memiliki Precision Score 0,8 (80%) dan nilai False Positive paling rendah.
- Precision Score diutamakan untuk meminimalisir False Positive (aktual charged off, prediksi fully paid). Hal ini untuk menghindari resiko kerugian bank apabila peminjam tidak dapat melunasi pinjamannya.

Conclusion



Recommendation

- Model machine learning dapat diterapkan pada dashboard divisi peminjaman bank untuk membantu proses assessment pinjaman sebelum disetujui. Penggunaan machine learning dapat mempercepat proses assessment dan mengurangi resiko pinjaman yang tidak lunas.



Dashboard Predictor

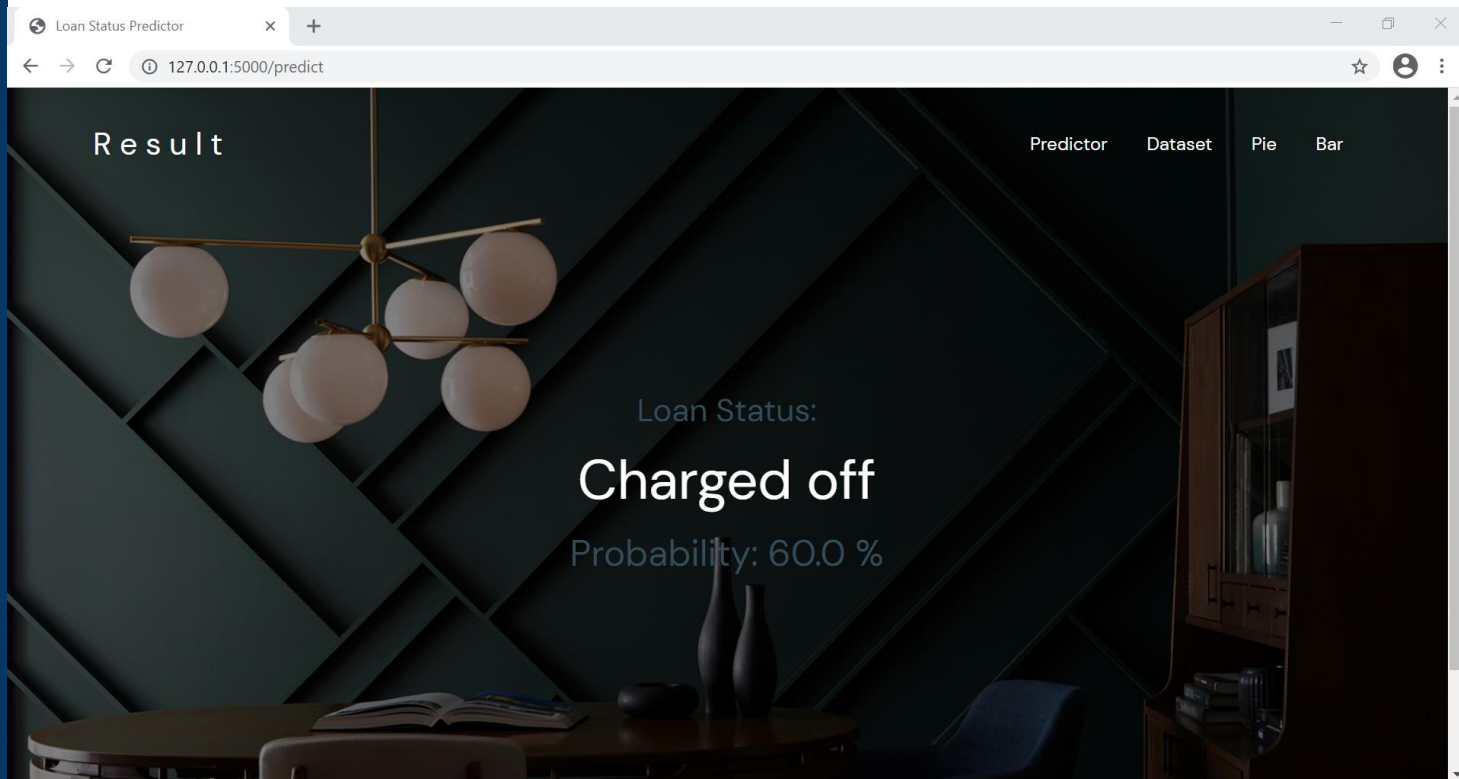
Loan Status Predictor

Predictor Dataset Pie Bar

Loan Status Predictor

Term	▼
Credit Score	▲▼
Annual Income	
Years of Credit History	
Home Ownership	▼
Current Loan Amount	

Dashboard Predictor



Dashboard Dataset

Loan Status Predictor

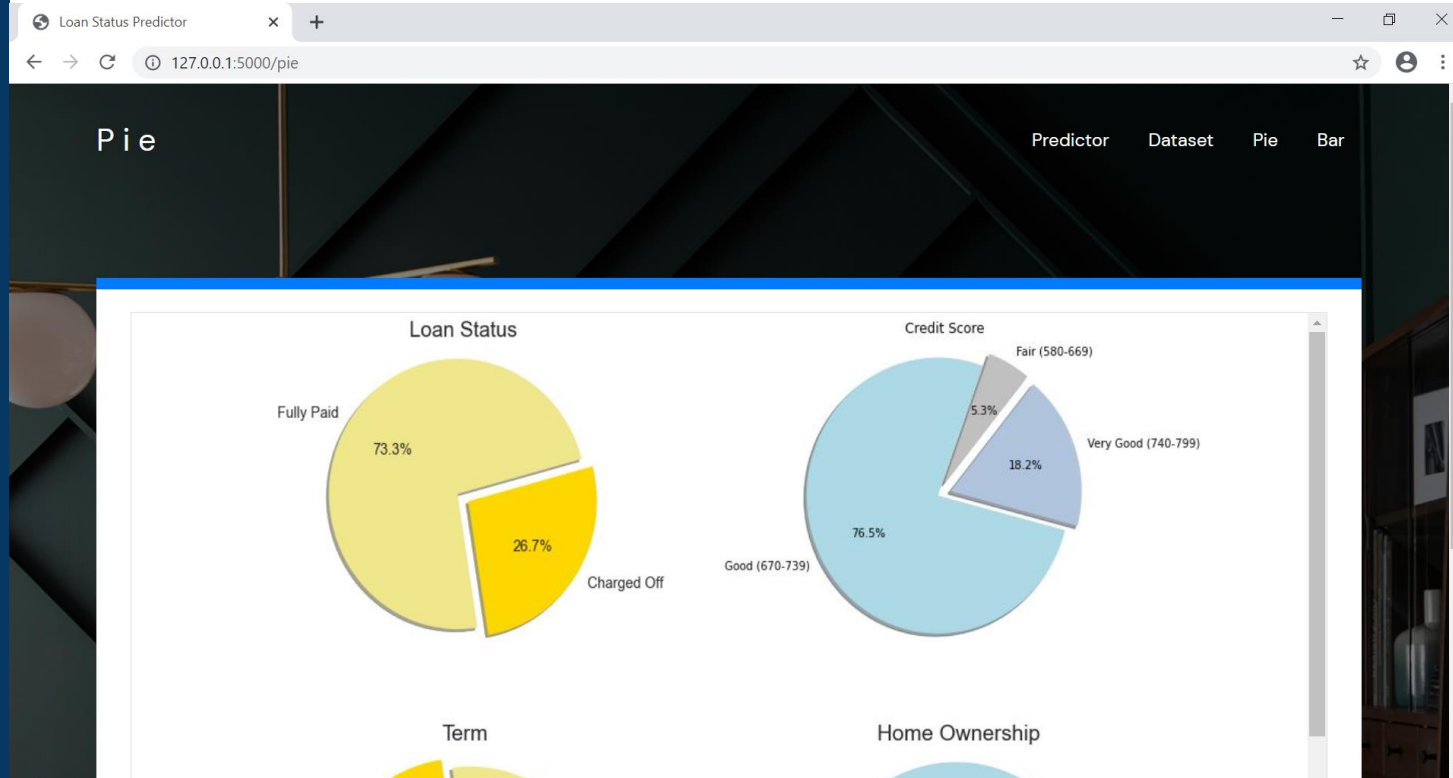
127.0.0.1:5000/data

Dataset

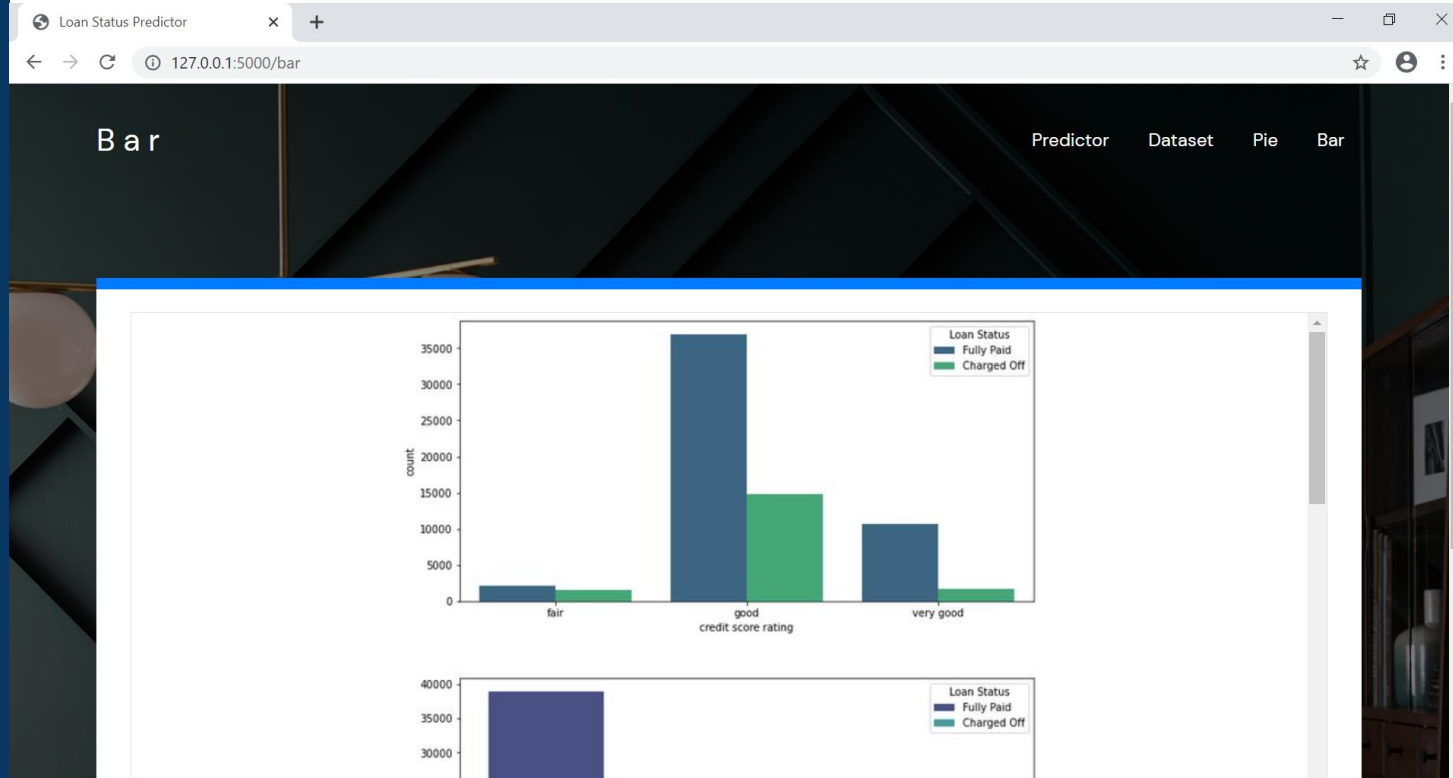
Predictor Dataset Pie Bar

	Loan_Status	Current_Loan_Amount	Term	Credit_Score	Annual_Income	Years_in_current_job	Home_Ownership	Purpose	Month
0	Fully Paid	445412.0	Short Term	709.0	1167493.0	8	Home Mortgage	Home Improvements	5214.0
1	Fully Paid	262328.0	Short Term	726.0	1172718.0	10	Home Mortgage	Debt Consolidation	3329.0
2	Fully Paid	347666.0	Long Term	721.0	806949.0	3	Own Home	Debt Consolidation	8741.0
3	Fully Paid	176220.0	Short Term	726.0	1172718.0	5	Rent	Debt Consolidation	2063.0
4	Fully Paid	217646.0	Short Term	730.0	1184194.0	0	Home Mortgage	Debt Consolidation	1085.0
5	Charged Off	648714.0	Long Term	726.0	1172718.0	0	Home Mortgage	Buy House	1480.0
6	Fully Paid	548746.0	Short Term	678.0	2559110.0	2	Rent	Debt Consolidation	1866.0

Dashboard Visualisasi



Dashboard Visualisasi



A top-down view of a collaborative workspace. In the center is an open laptop with a dark screen and a visible keyboard. To the left, a person's hands are holding a blue pen over a tablet. Another hand is visible at the bottom left, also near a tablet. To the right of the laptop, a hand is resting on the trackpad. A white mug is positioned between the laptop and the left-hand tablets. The entire scene is overlaid with a semi-transparent blue filter. The text "Thank you" is centered over the laptop keyboard in a white, italicized serif font.

Thank you