

PROGRAM VIRTUAL INTERNSHIP



Rakamin
Academy



id/x partners

End-to-end Data Science Project: Credit Risk Modeling

Oleh:
Tyovendi Arisandy



List Of Content

1

Introduction of Project

2

Exploratory Data Analysis (EDA)

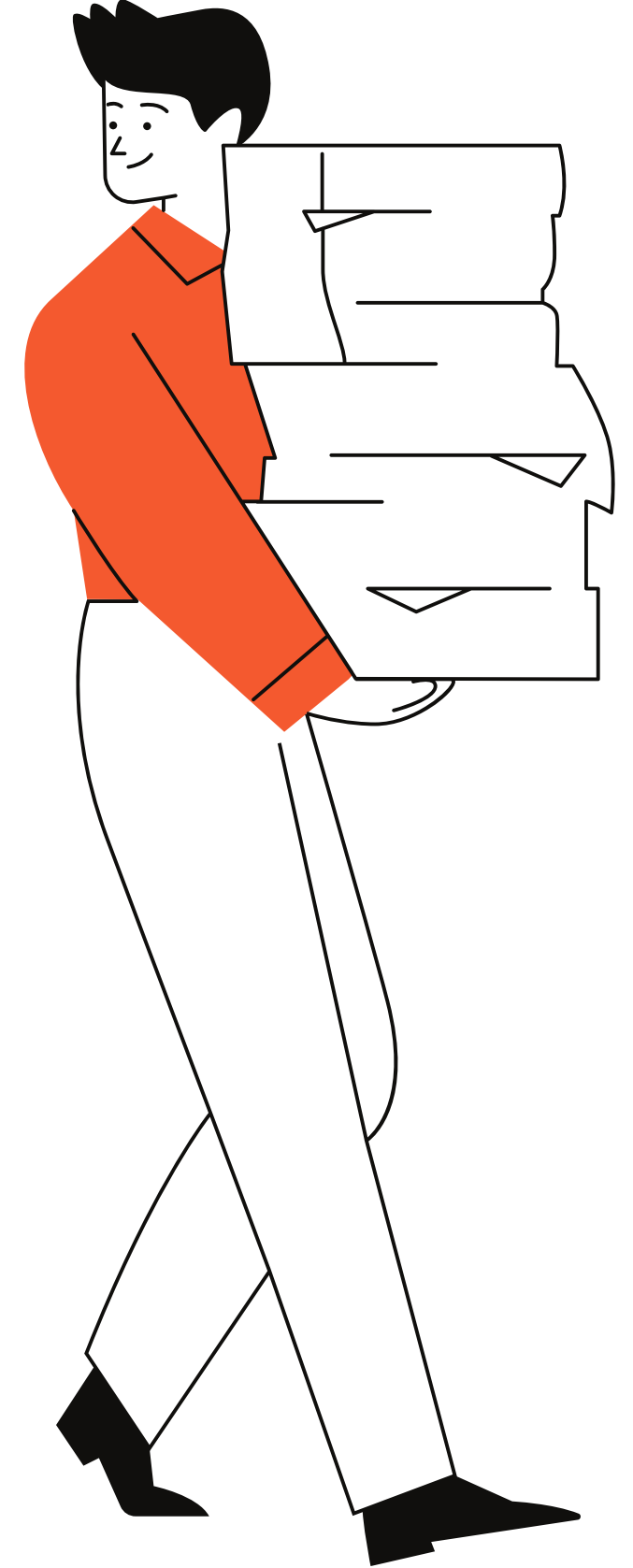
3

Machine Learning Modeling

4

Conclusion/Recommendation

Introduction of Project



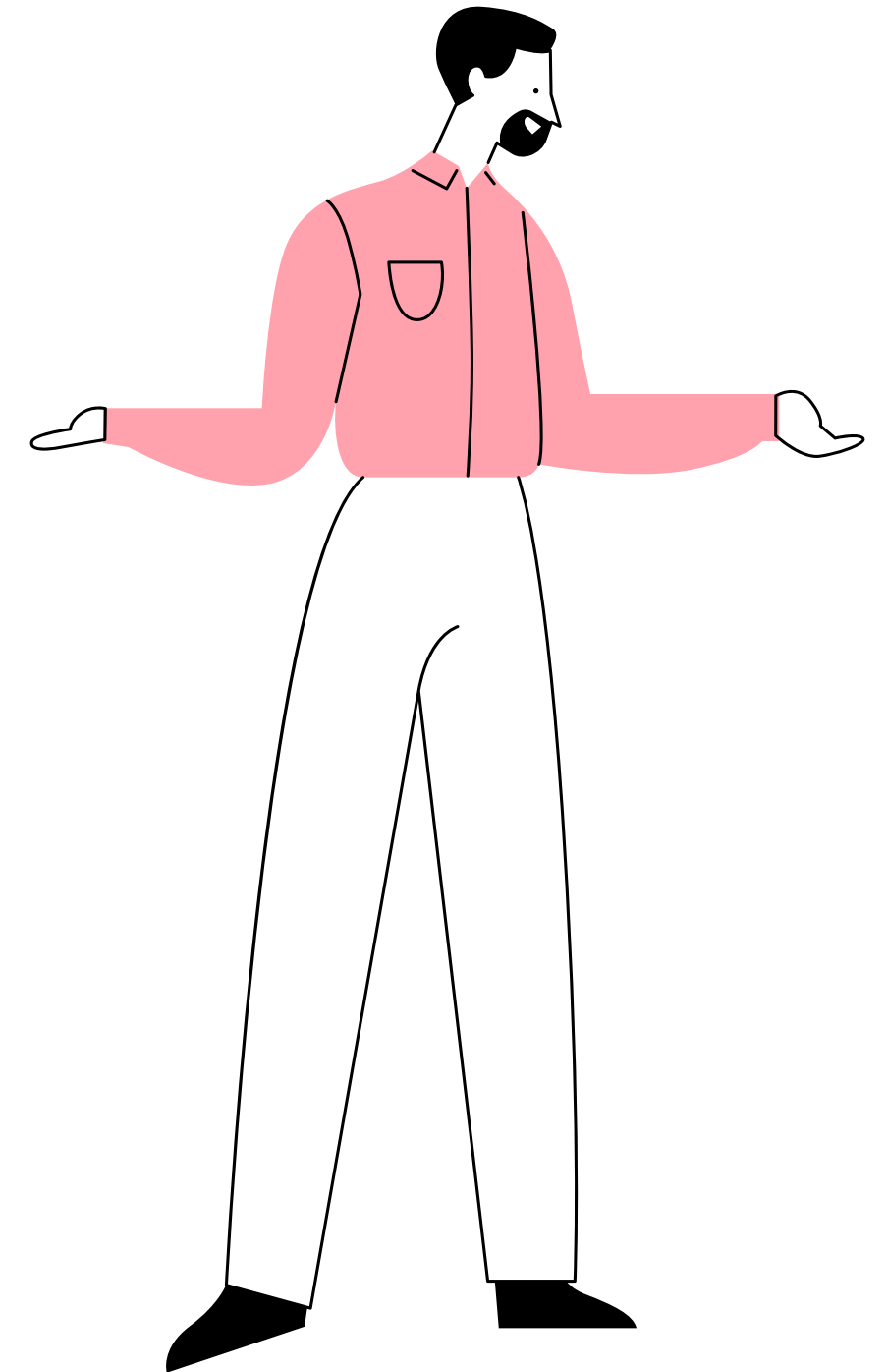
Business Context

Kata kredit berasal dari bahasa Latin Credere yang berarti percaya atau to believe atau to trust. Karenanya dasar pemikiran pemberian kredit oleh suatu perbankan kepada seseorang/lembaga adalah berdasarkan kepercayaan (faith). Sesuai Undang-Undang No. 10 Tahun 1998 tentang Perbankan, Kredit adalah penyediaan uang atau tagihan yang dapat dipersama-kan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. akan tetapi seiring berjalannya waktu, seringkali pihak pemberi kredit/pinjaman mengalami kerugian karena terjadinya kredit yang macet ataupun gagal bayar. untuk itu perlu adanya mekanisme yang lebih efektif untuk mengatasi permasalahan risiko kredit.



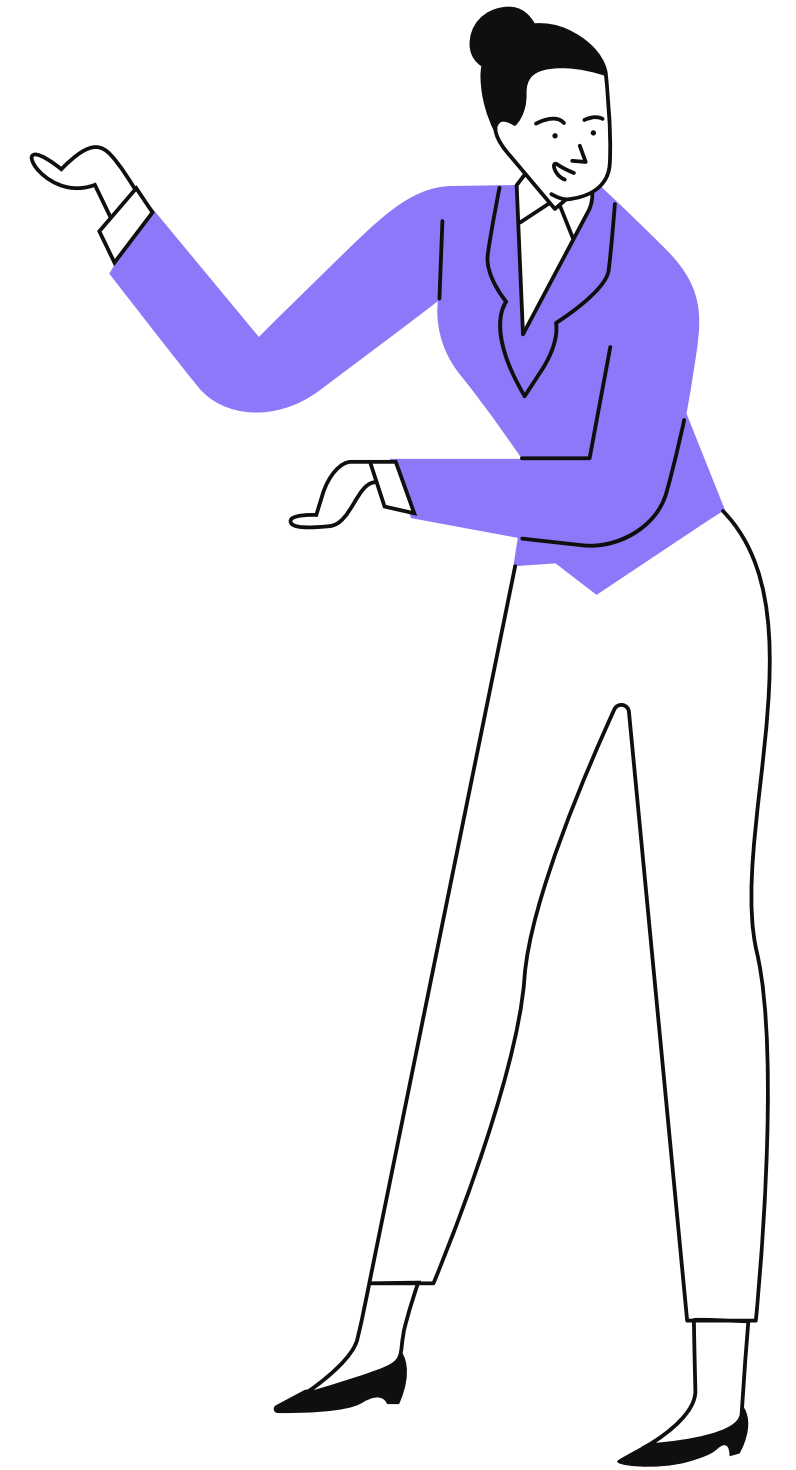
Problem Statement

Risiko kredit (credit risk) didefinisikan sebagai risiko kerugian yang terkait dengan kemungkinan kegagalan counterparty memenuhi kewajibannya; atau risiko bahwa debitur tidak membayar kembali utangnya. Risiko kredit timbul dari adanya kemungkinan bahwa kredit yang diberikan oleh bank, atau obligasi yang dibeli, tidak dapat dibayarkan kembali. Risiko kredit juga timbul dari tidak dipenuhinya berbagai bentuk kewajiban pihak lain kepada bank, seperti kegagalan memenuhi kewajiban pembayaran dalam kontrak derivatif. Untuk sebagian bank, risiko kredit merupakan risiko terbesar yang dihadapi. Pada umumnya, margin yang diperhitungkan untuk mengantisipasi risiko kredit hanyalah merupakan bagian kecil dari total kredit yang diberikan bank dan oleh karenanya kerugian pada kredit dapat menghancurkan modal bank dalam waktu singkat.



Goals

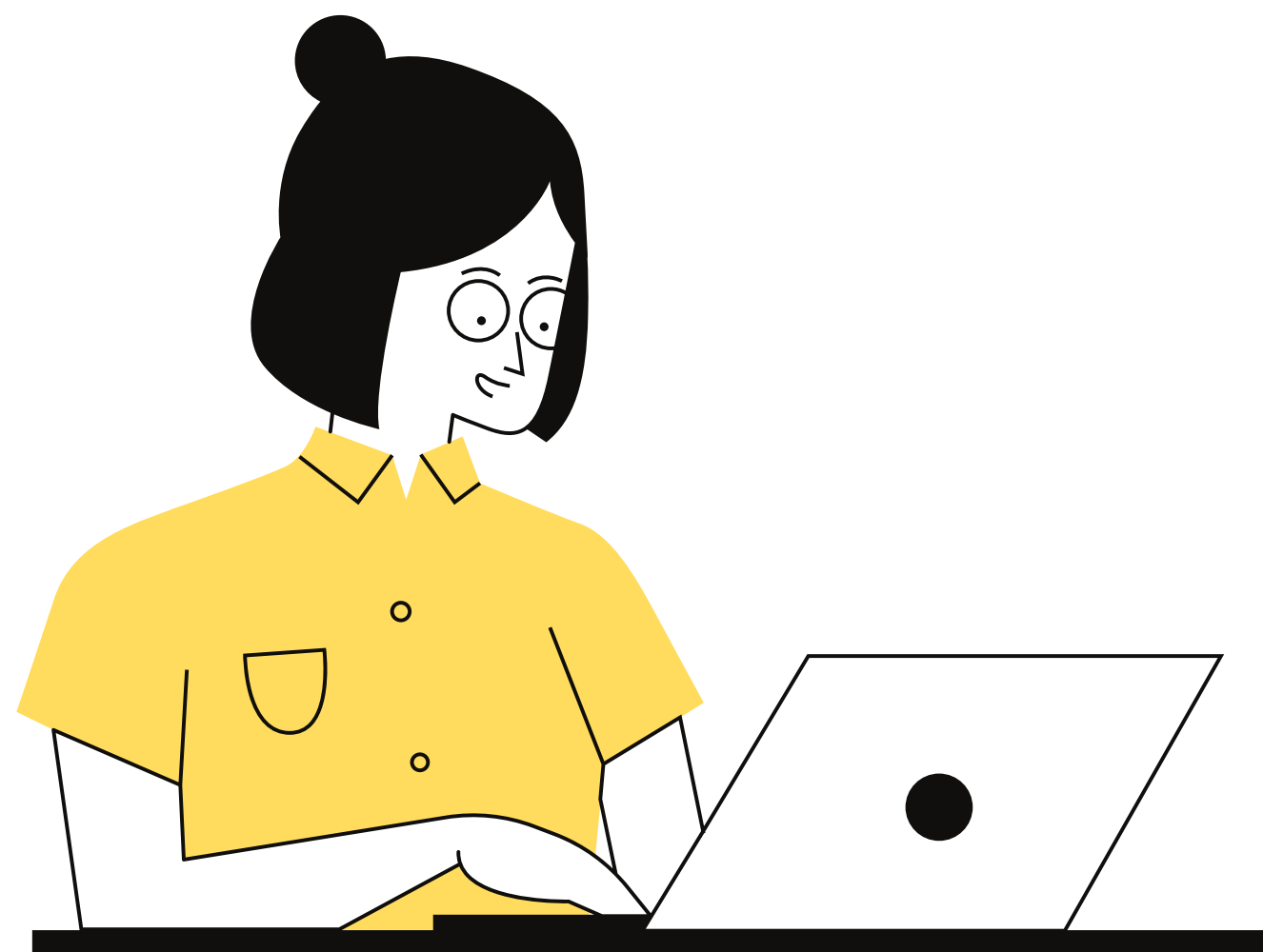
Berdasarkan permasalahan diatas, mengurangi tingkat risiko kredit (credit risk) merupakan tujuan bisnis utama perusahaan. Maka dari itu, perlu adanya sebuah mekanisme yang dapat memprediksi apakah pihak peminjam termasuk kedalam kategori 'Bad' atau 'Good' dalam sistem kredit, serta faktor-faktor yang memberikan informasi tambahan dan mendukung status 'Bad' atau 'Good' pihak peminjam dalam kasus credit risk ini.



Analytical Approach

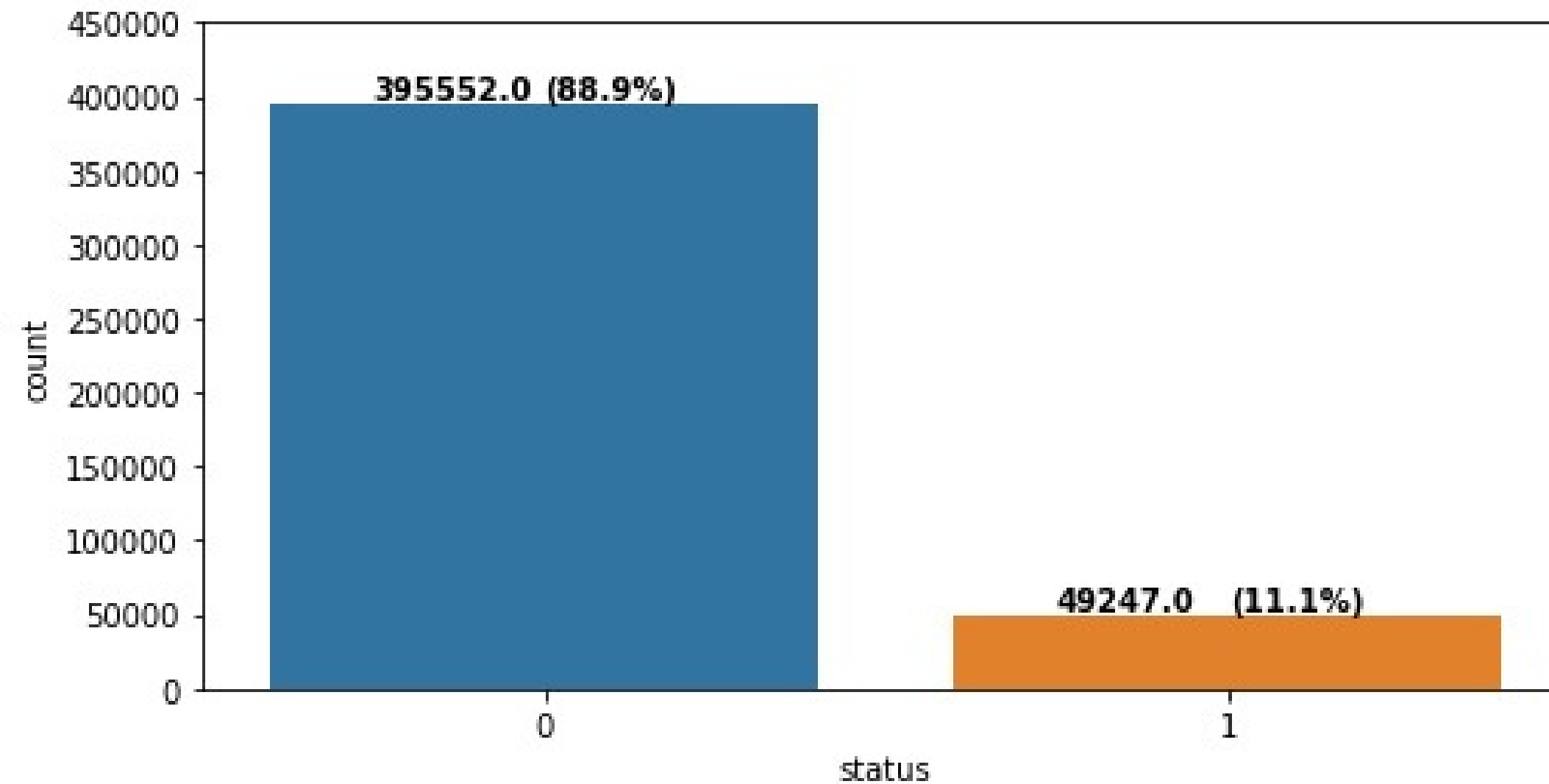
- Menganalisis data untuk dapat menemukan pola dari fitur-fitur yang tersedia dalam dataset, pada akhirnya menemukan karakteristik dari peminjam yang berstatus 'Bad' atau 'Good'.
- Membangun suatu model klasifikasi yang akan membantu perusahaan untuk dapat menyediakan 'tool' yang dapat memprediksi peminjam/debitur memperoleh status 'Bad' atau 'Good'.





Exploratory Data Analysis (EDA)

Loan Status



Keterangan :

- Terdapat 11% (49247) pinjaman dengan status Bad Loans, sedangkan sebanyak 88% (395552) pinjaman berstatus Good Loans.
- 0 = Good Loans
- 1 = Bad Loans

Loan Status

```
[ ] loss_amnt[loss_amnt['status']== 1].sum()
```

```
loan_amnt    19086425
status        1178
total        730682825
dtype: int64
```

```
[ ] loss_amnt[loss_amnt['status']== 0].sum()
```

```
loan_amnt    23750675
status         0
total        5697873550
dtype: int64
```

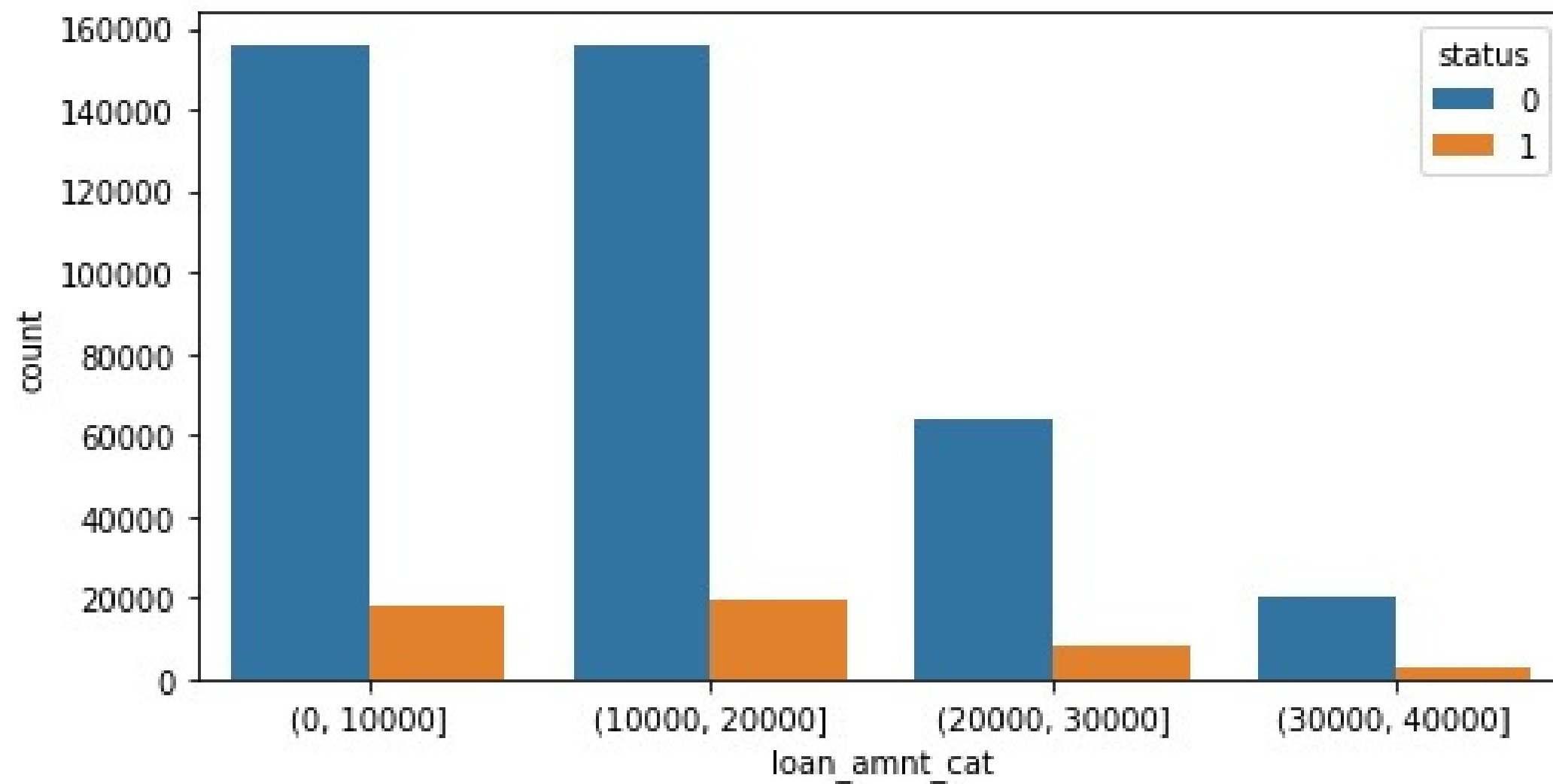
```
[ ] # Total Kredit (Good Loan + Bad Loan)
    5697873550+730682825
```

```
6428556375
```

Keterangan :

- Total seluruh pinjaman yang diterbitkan yaitu \$6.428.556.375 B
- Total pinjaman yang diterbitkan untuk pinjaman yang berstatus Good Loans sebesar \$5.697.873.550 B
- Total pinjaman yang diterbitkan untuk pinjaman yang berstatus Bad Loans sebesar \$730.682.825 M
- Berdasarkan ratio pinjaman antara Good Loans vs Bad Loans, total pinjaman Good Loans > Bad Loans. akan tetapi potensi kerugian sebesar \$730.682.825 M masih tetap akan terjadi jika peminjam/debitur tidak sanggup membayar utangnya/kredit dan berakhir pada kondisi gagal bayar (default)

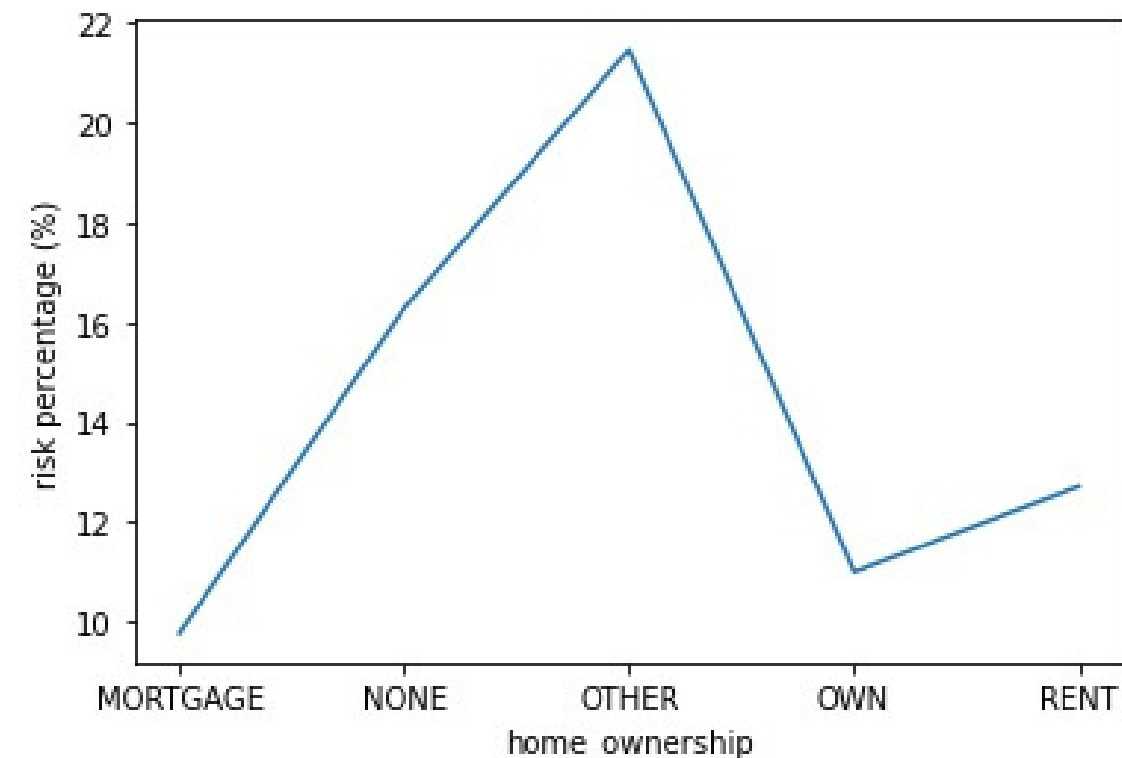
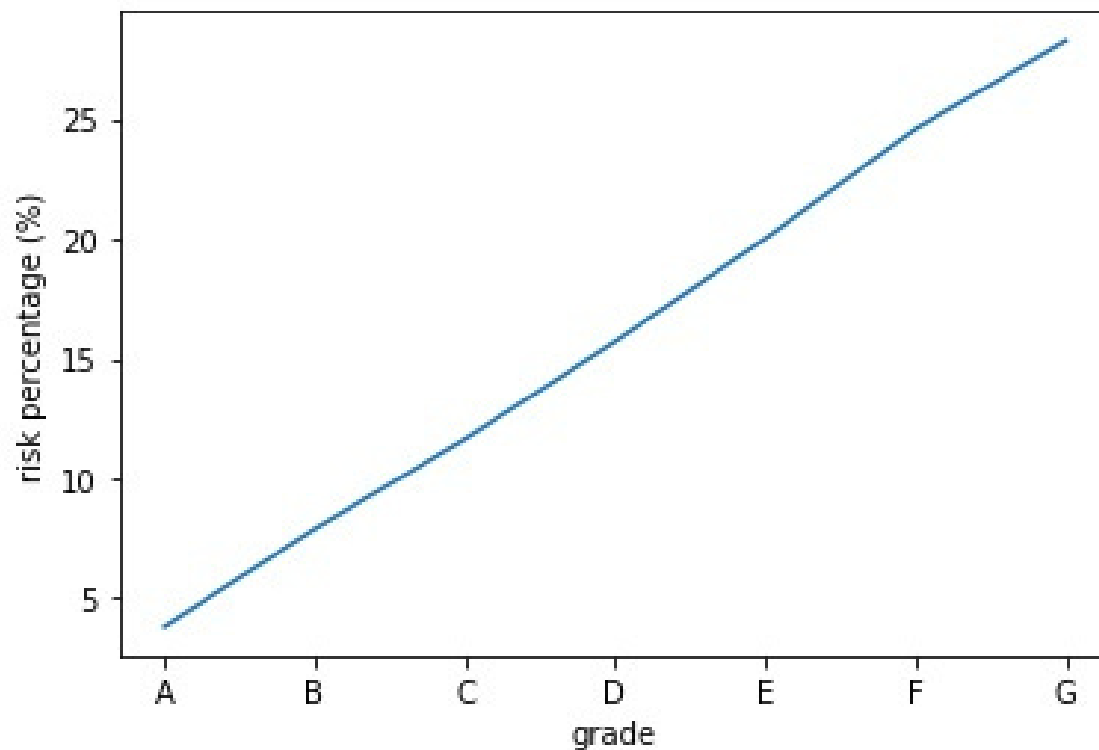
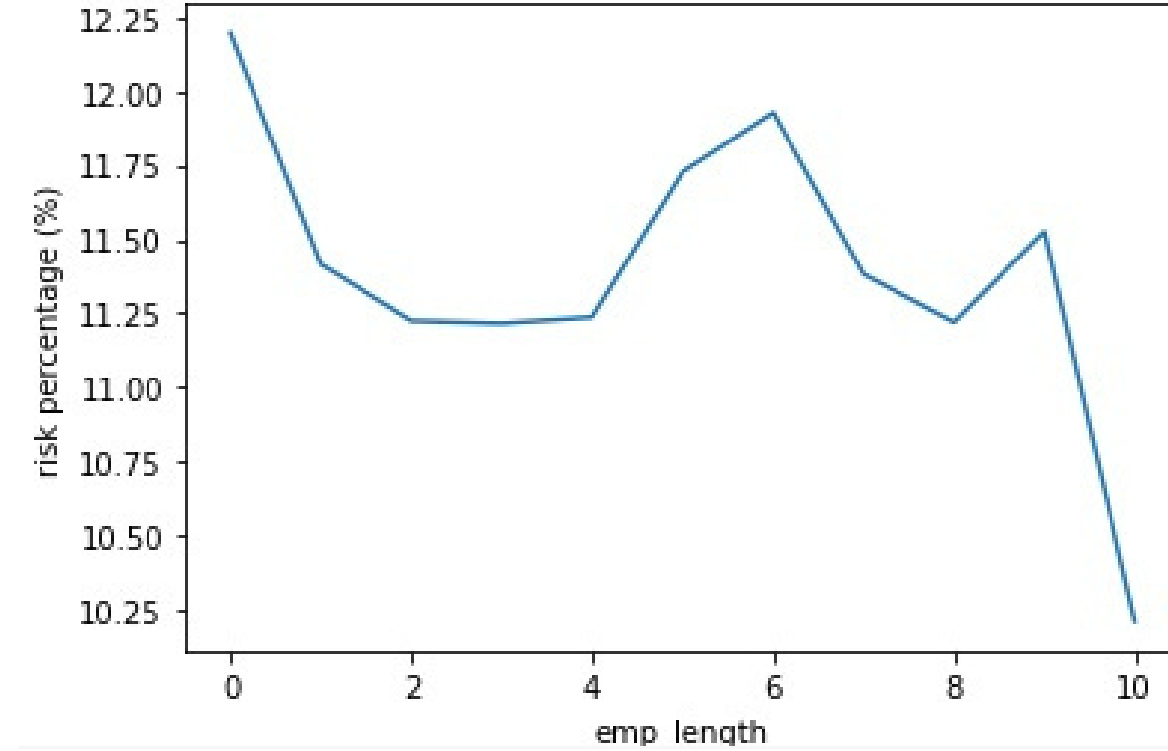
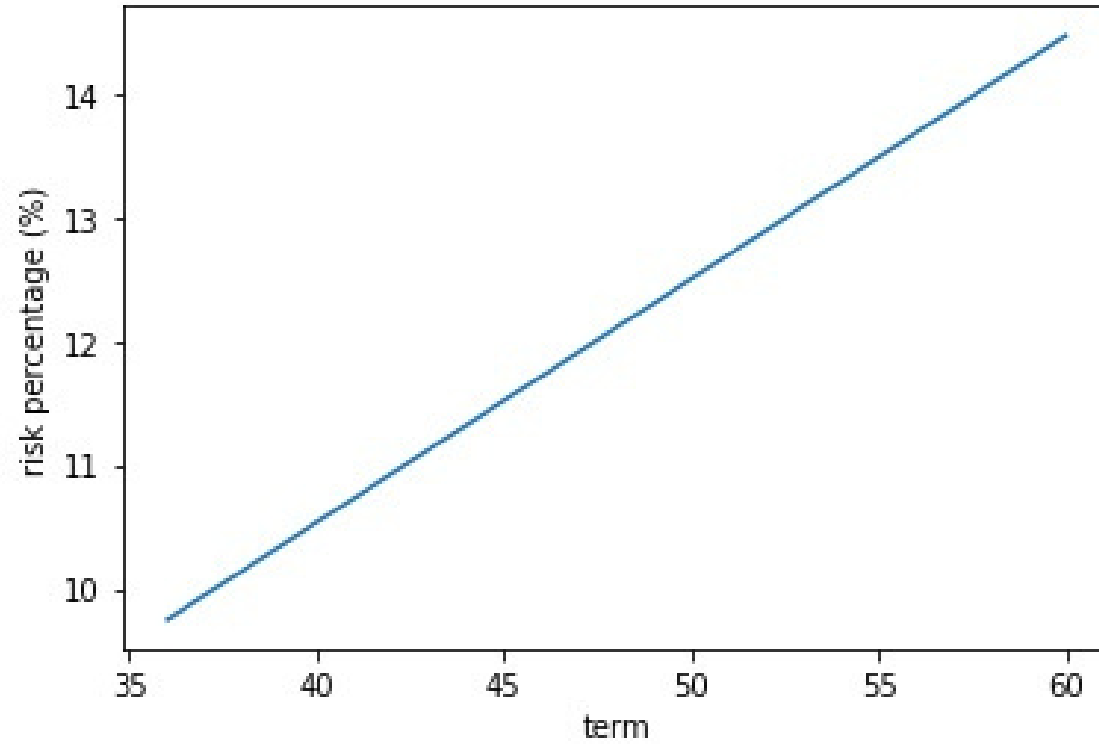
Loan Status



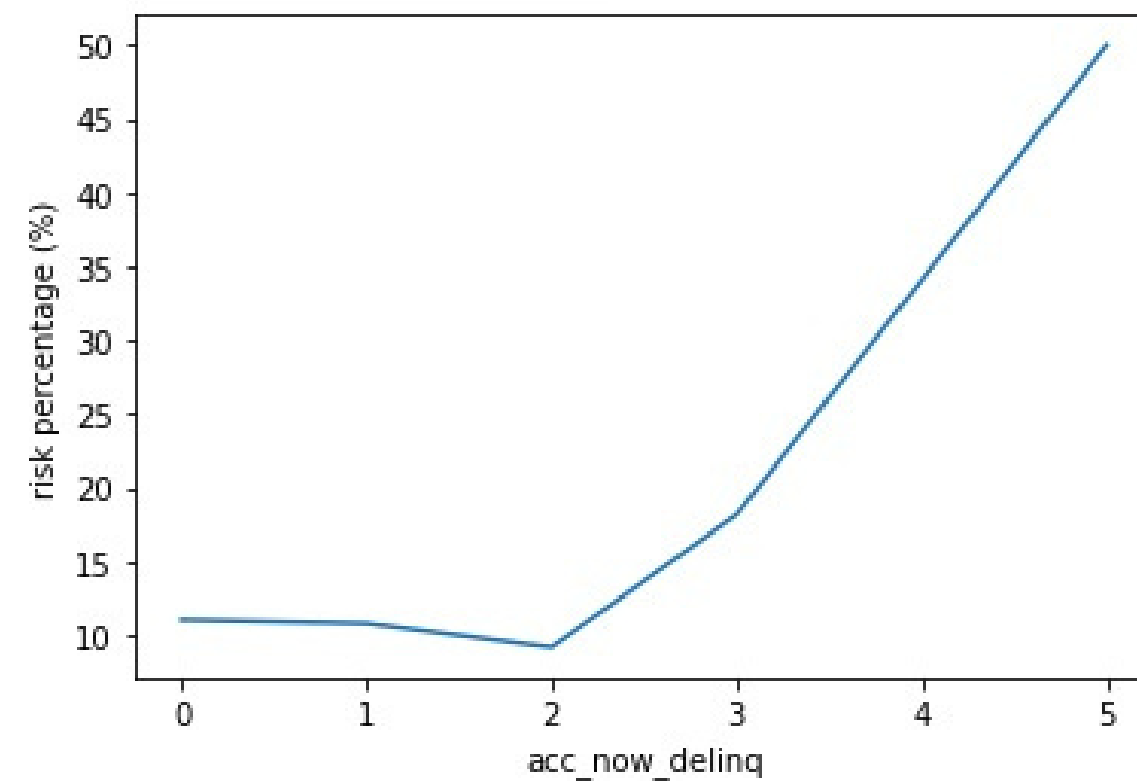
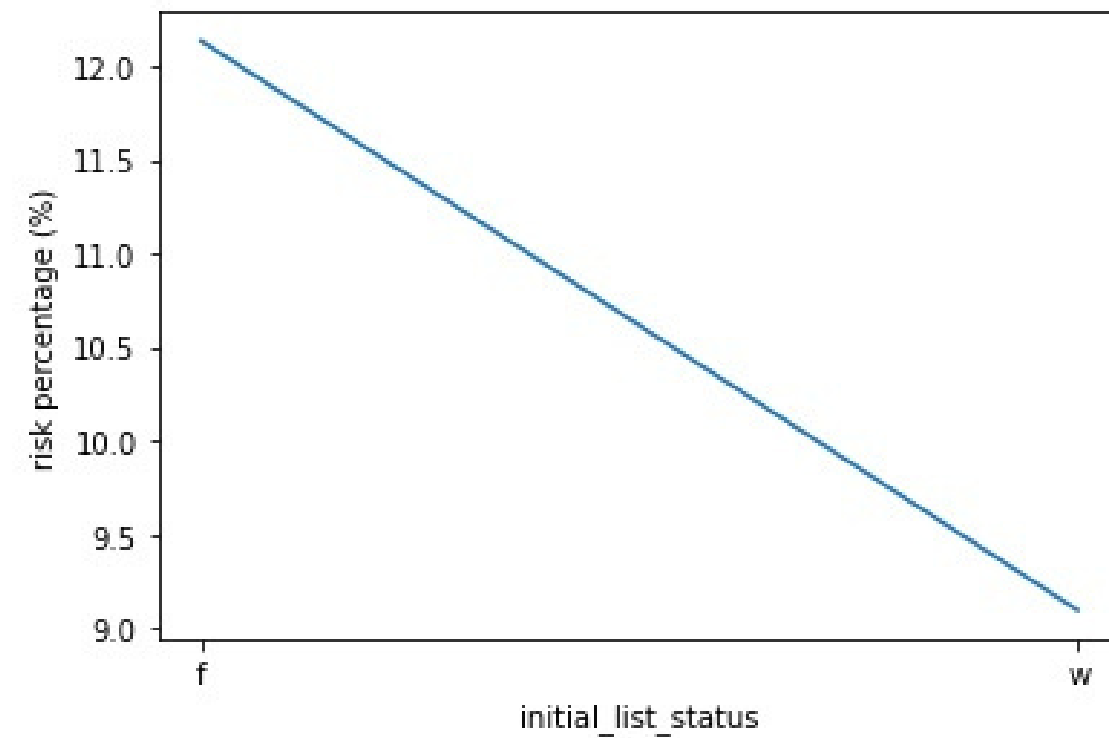
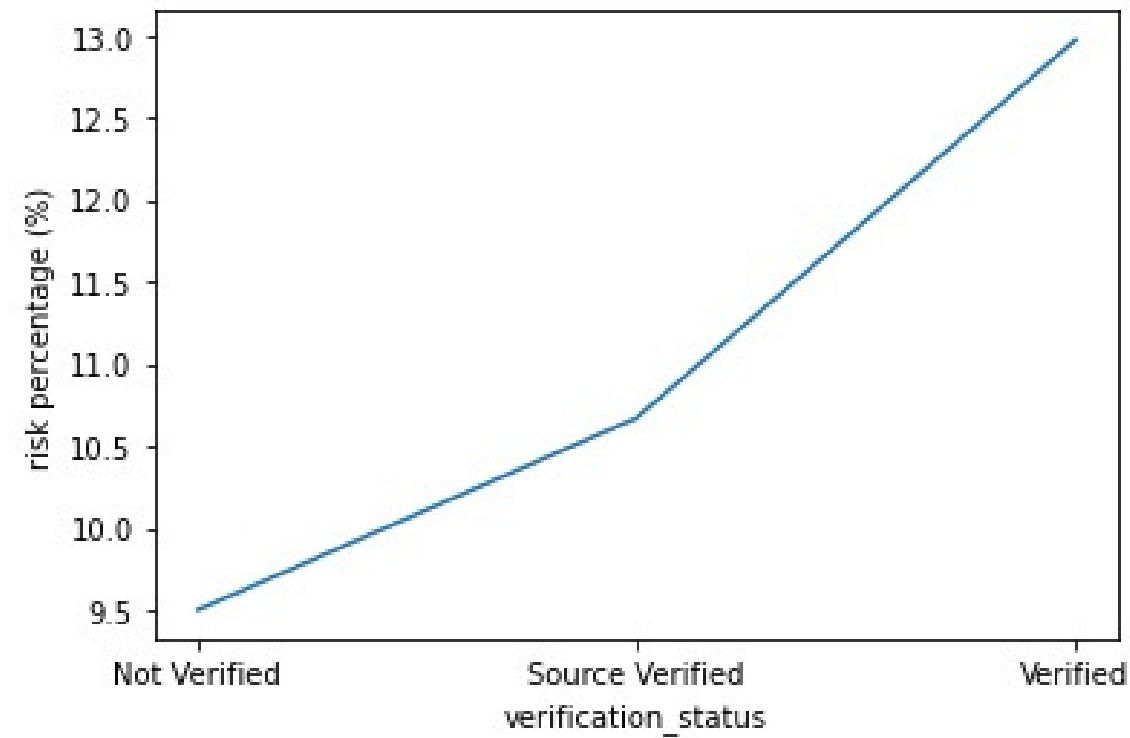
Keterangan :

- Kebanyakan pinjaman berstatus Bad Loans terbanyak berada pada rentang \$10000 - 20000

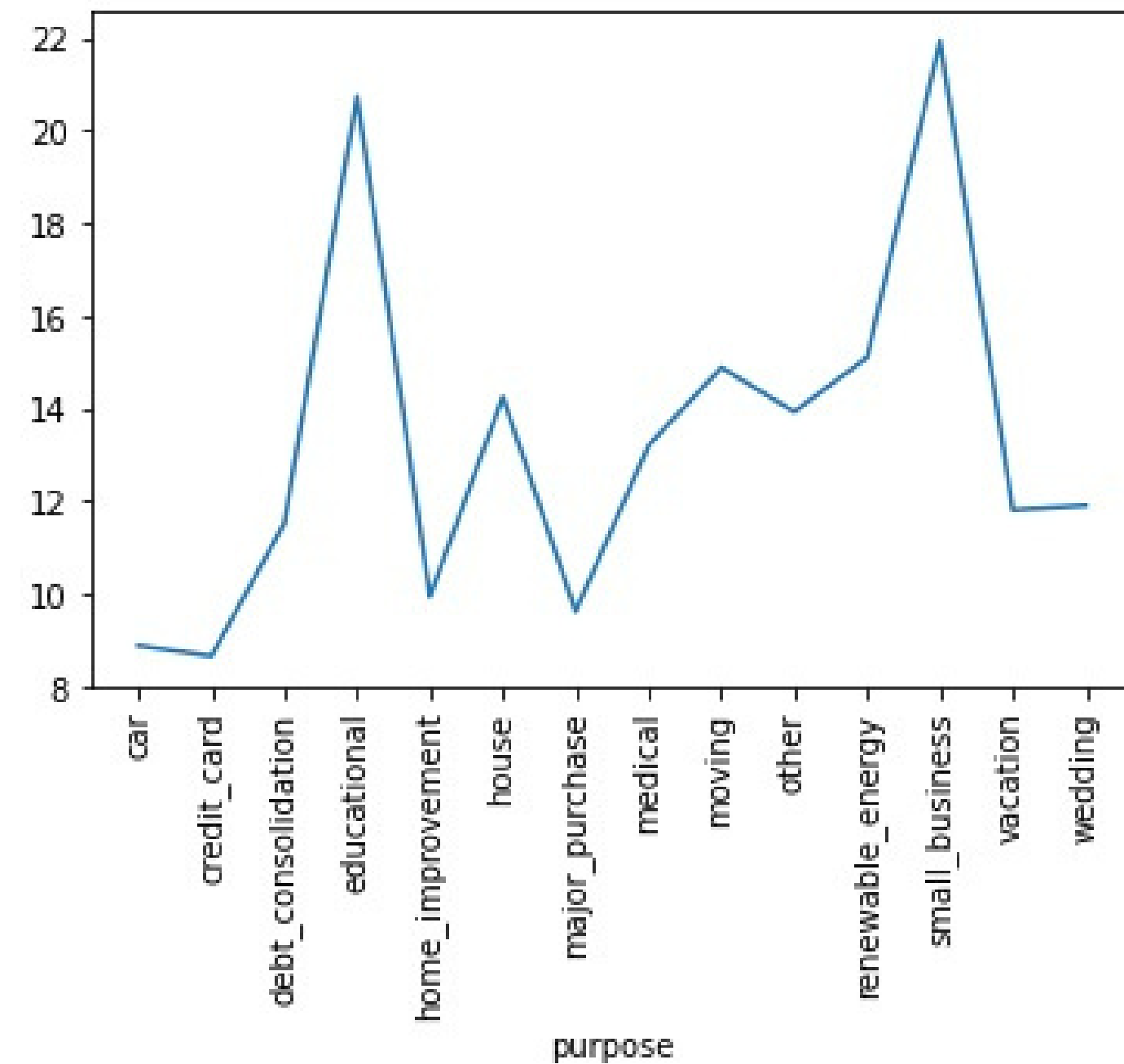
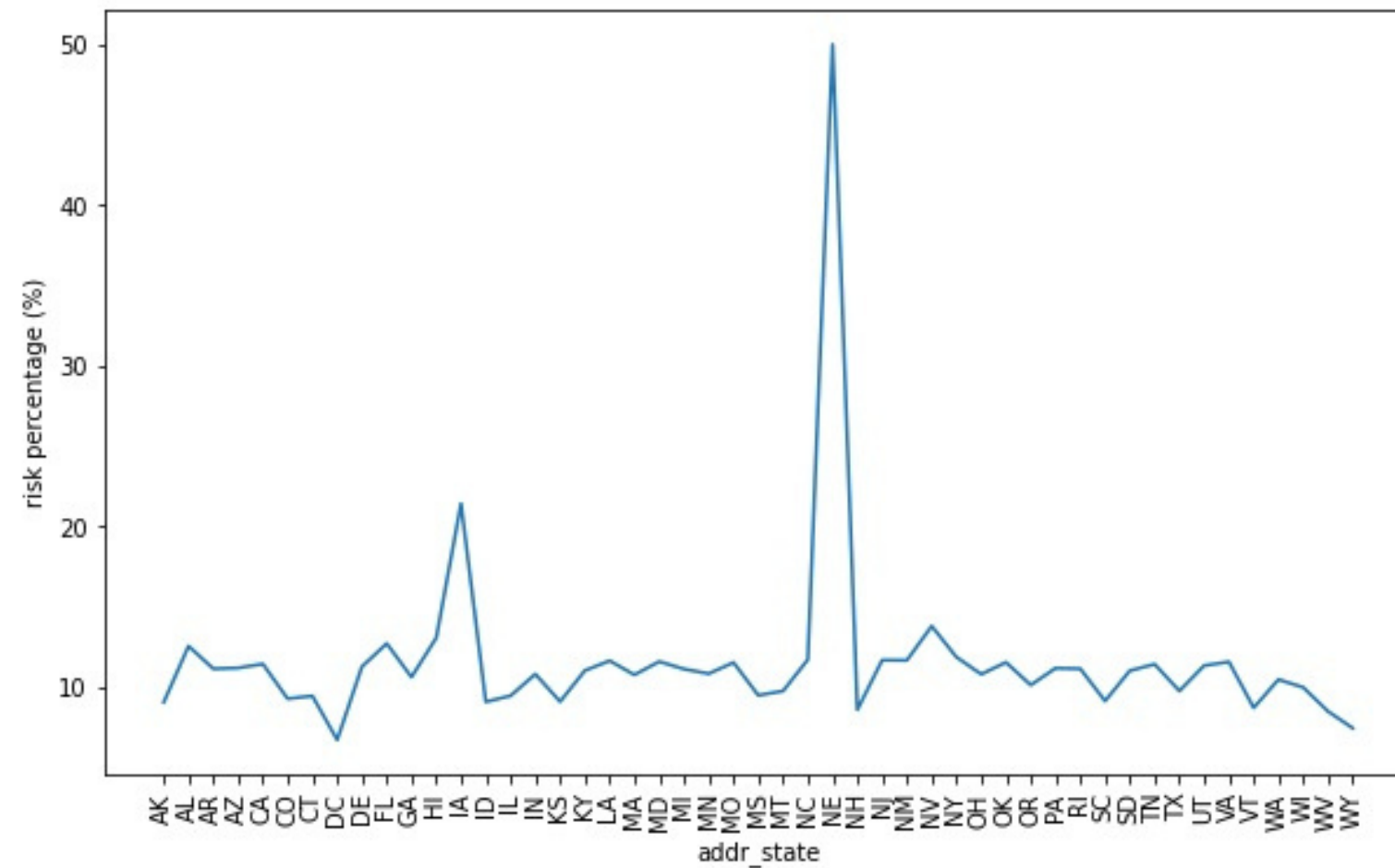
Risk Percentage of Bad Loans



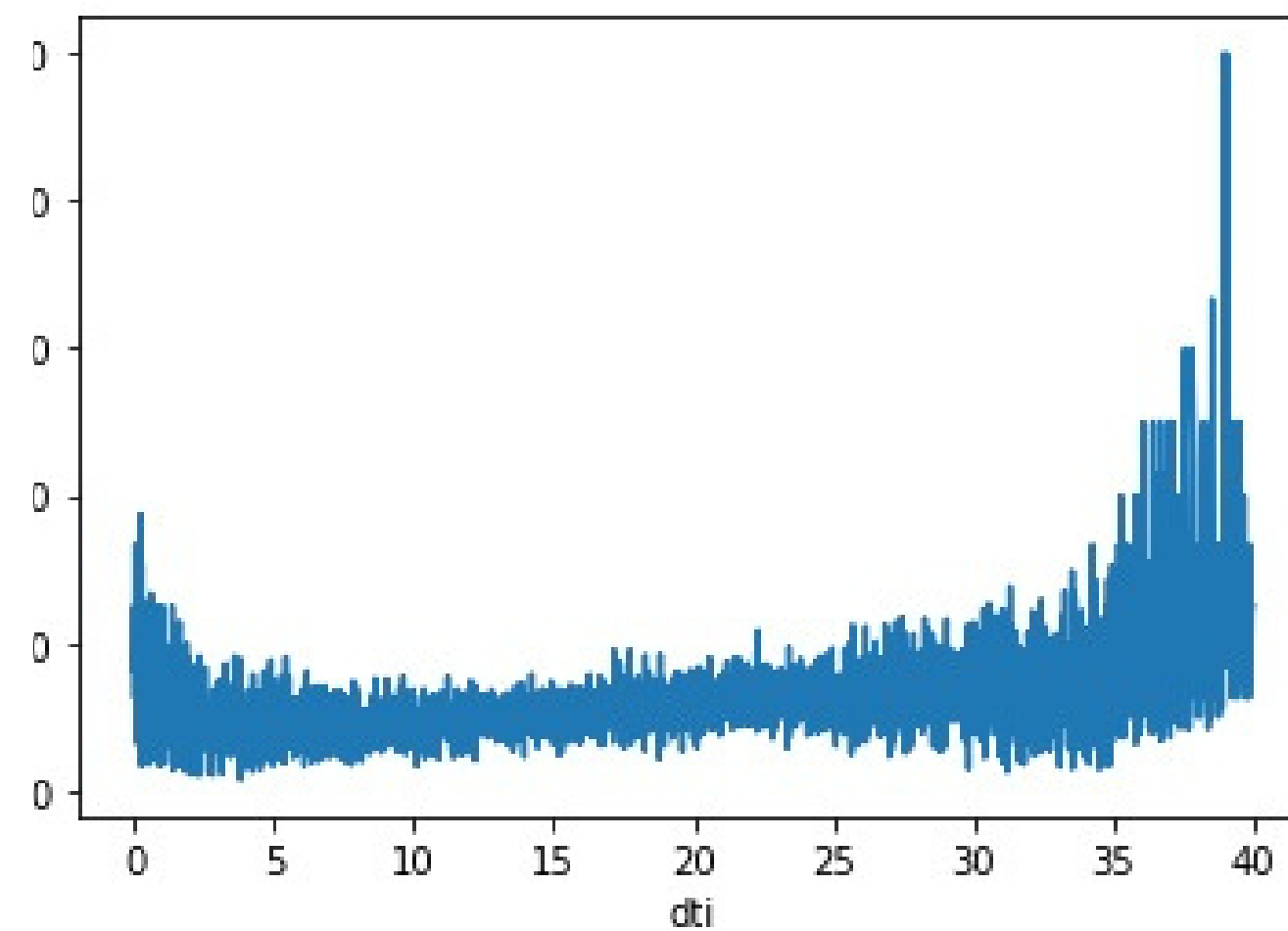
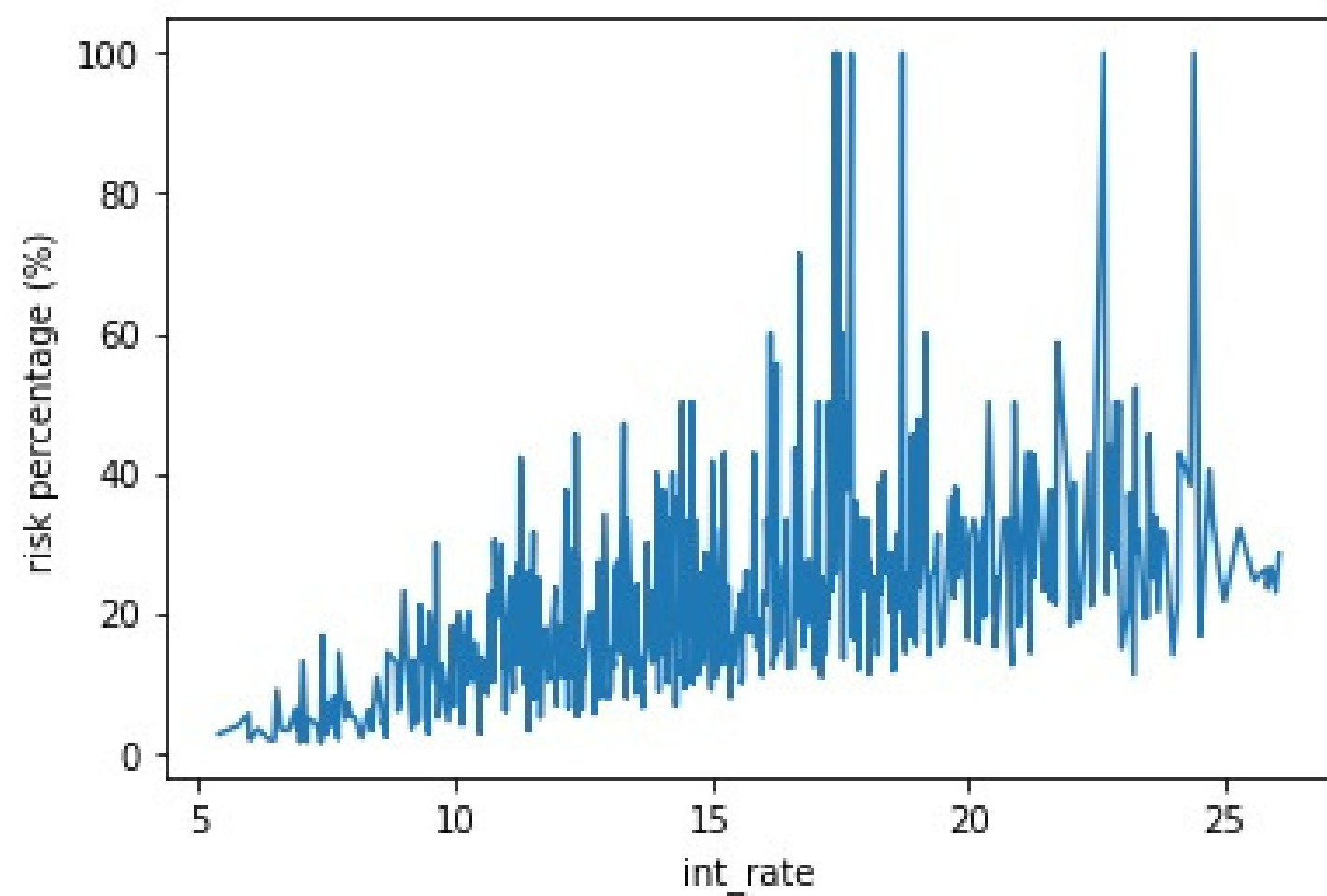
Risk Percentage of Bad Loans



Risk Percentage of Bad Loans



Risk Percentage of Bad Loans



Insight of Bad Loans

- **term**: semakin tinggi term maka semakin tinggi persentase risiko terjadinya Bad Loans
- **emp_length**: semakin lama masa kerja peminjam/debitur maka persentase risiko terjadinya Bad Loans semakin kecil
- **grade**: semakin tinggi grade pinjaman maka persentase risiko terjadinya Bad Loans semakin tinggi
- **home_ownership**: jika peminjam/debitur tidak memiliki kepemilikan rumah (none/any/other) maka persentase risiko terjadinya Bad Loans sangat tinggi, sedangkan yang memiliki hipotik cenderung rendah terjadinya Bad Loans
- **addr_state**: negara bagian NE (Nebraska) menjadi negara paling berisiko terjadinya Bad Loans
- **purpose**: tujuan pinjaman untuk bisnis kecil dan dana pendidikan memiliki persentase risiko terjadi Bad Loans sangat tinggi. sedangkan untuk kebutuhan kartu credit dan kredit kendaraan (mobil) cenderung rendah untuk terjadinya Bad Loans
- **initial_list_status**: jika pinjaman telah memiliki berkas yang lengkap, maka persentase risiko terjadi Bad Loans cenderung rendah. sedangkan yang tidak melengkapi berkas pinjaman cenderung sangat tinggi untuk menaikkan persentase risiko terjadinya Bad Loans
- **int_rate**: kenaikan persentase risiko terjadinya Bad Loans dari suku bunga terlihat bervariasi, akan tetapi jika melihat trendnya, maka akan tampak bahwa semakin tinggi suku bunga maka akan semakin naik persentase risiko terjadinya Bad Loans
- **dti**: jika cicilan/angsuran rasionya diatas 35, maka terlihat adanya kenaikan persentase risiko terjadinya Bad Loans yang signifikan
- **acc_now_delinq**: jika akun rekening kredit > 2 akun, maka persentase risiko terjadinya Bad Loans naik secara signifikan

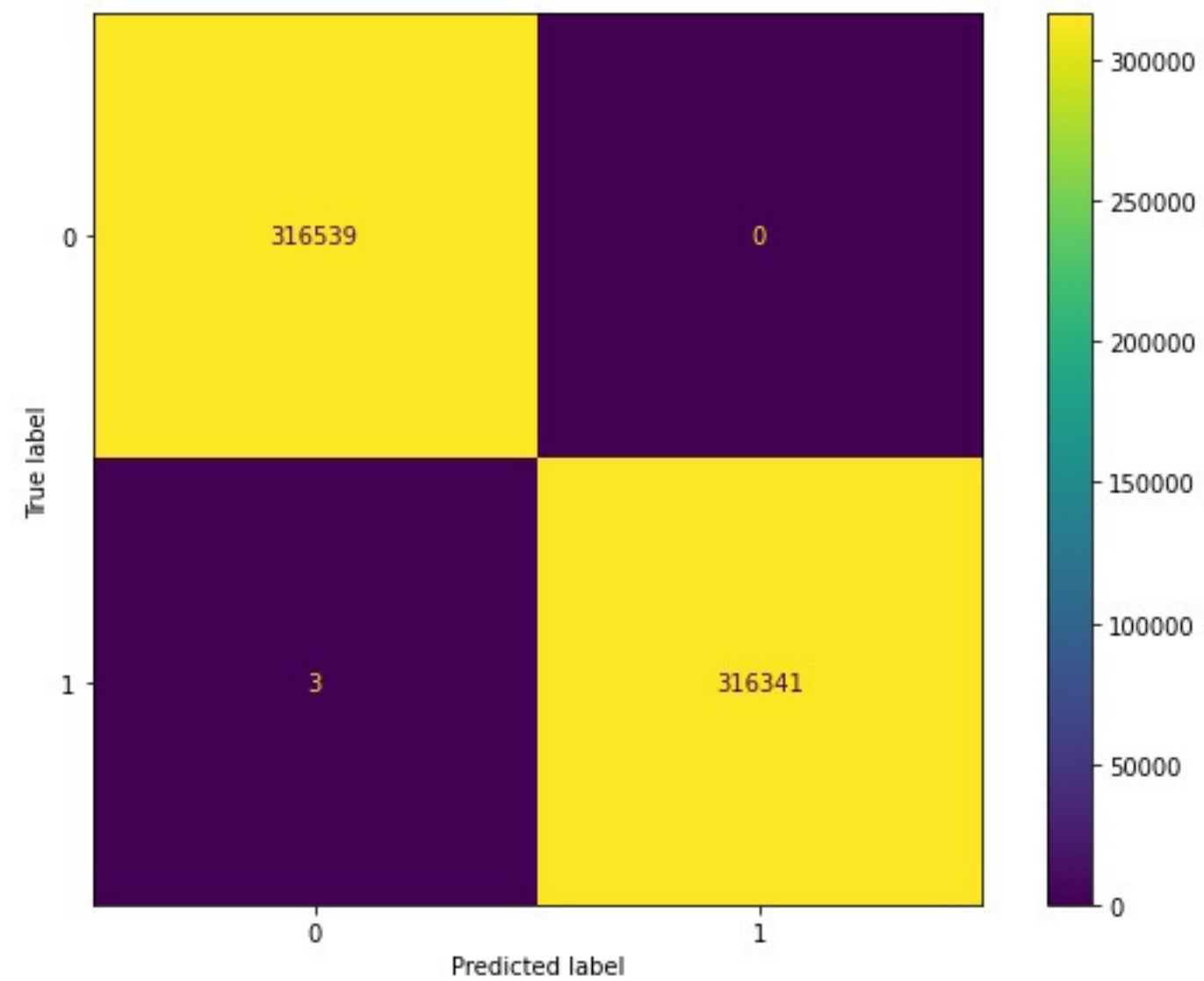
Machine Learning Modeling



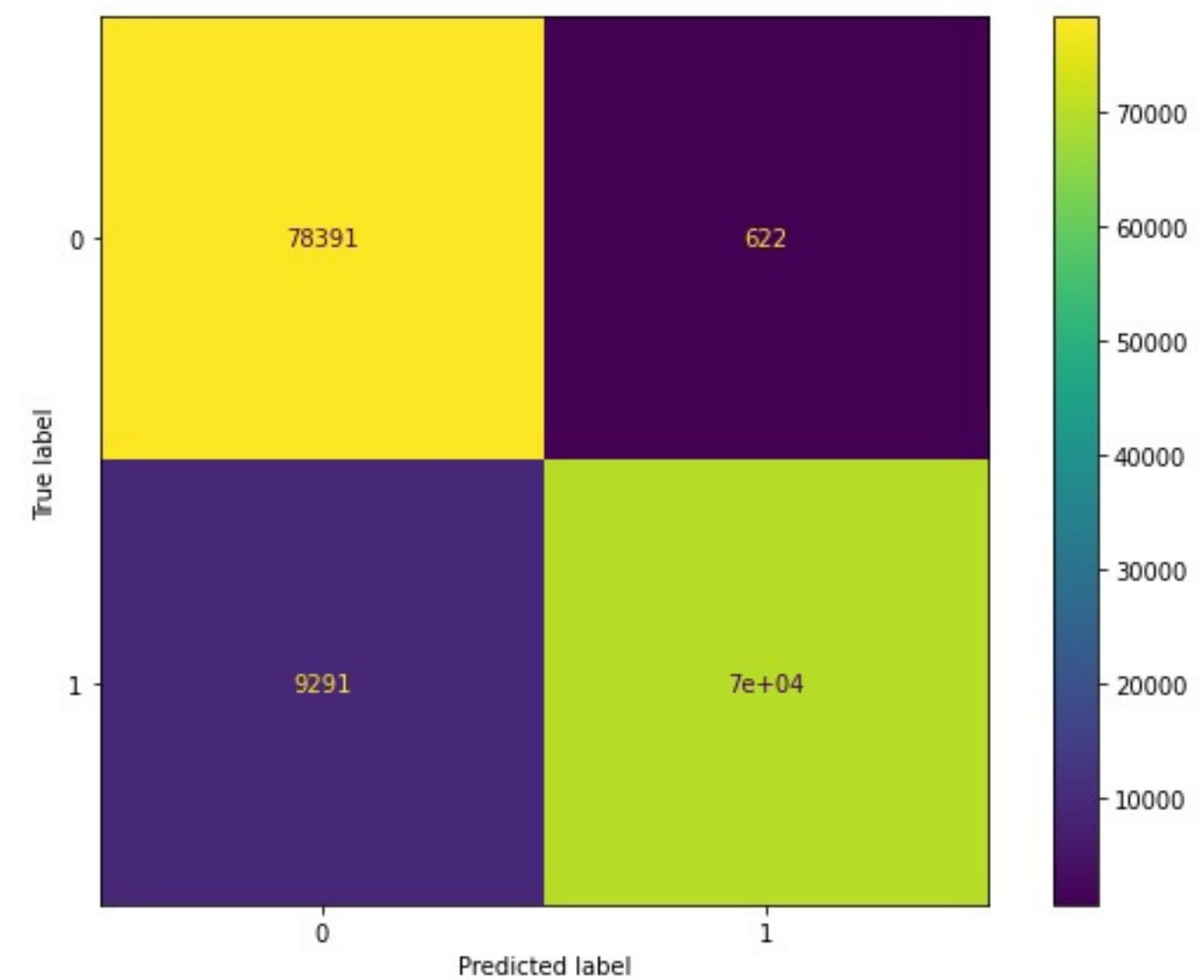
Random Forest

Confusion Matrix
Evaluation

Train Set



Test Set



Random Forest

Confusion Matrix Evaluation

	precision	recall	f1-score	support
0	1.00	1.00	1.00	316539
1	1.00	1.00	1.00	316344
accuracy			1.00	632883
macro avg	1.00	1.00	1.00	632883
weighted avg	1.00	1.00	1.00	632883

Train Set

	precision	recall	f1-score	support
0	0.89	0.99	0.94	79013
1	0.99	0.88	0.93	79208
accuracy			0.94	158221
macro avg	0.94	0.94	0.94	158221
weighted avg	0.94	0.94	0.94	158221

Test Set

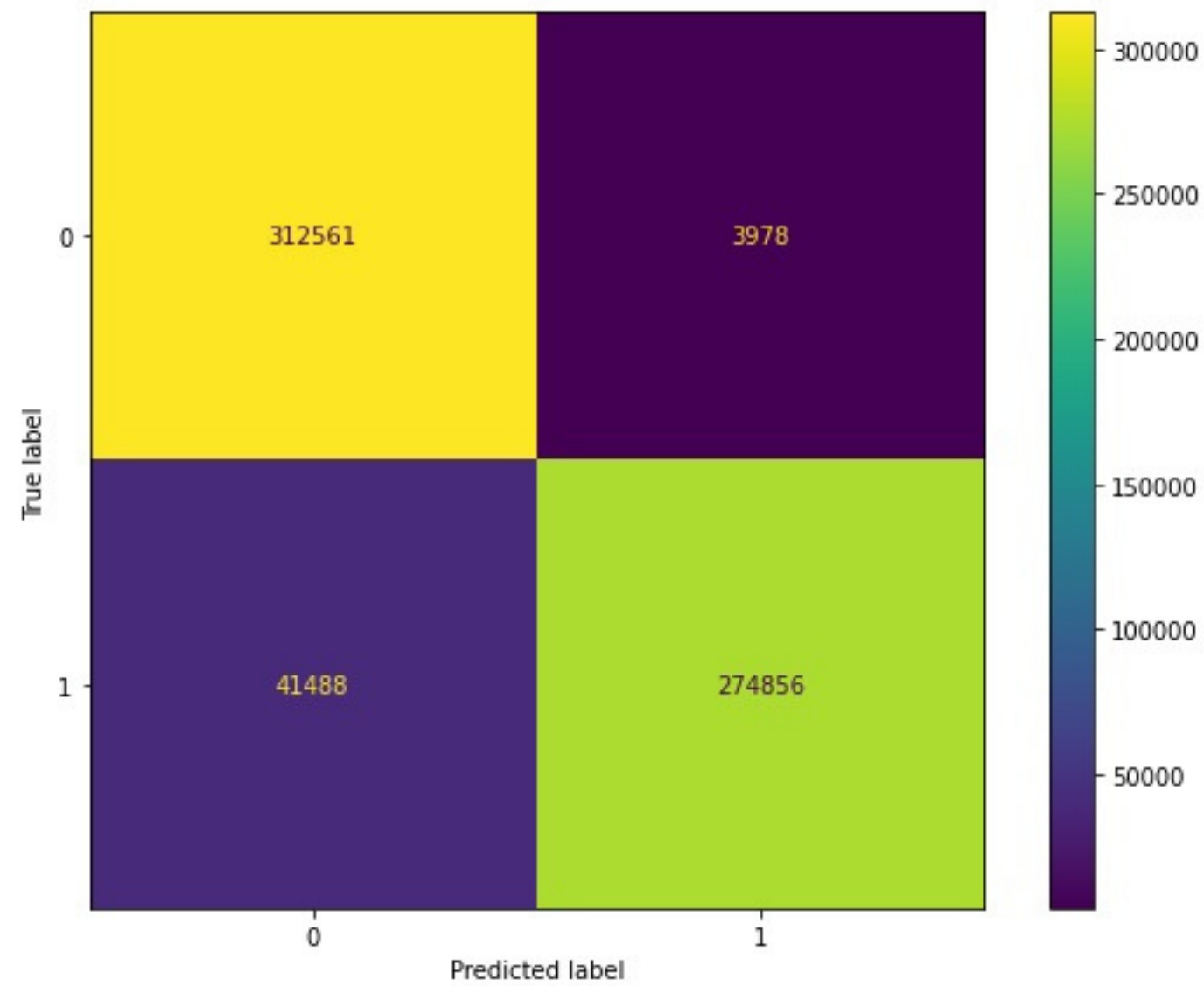
Keterangan:

- terlihat bahwa pada train set terjadi overfitting, sedangkan pada test set jika dilihat dari accuracy, precision, recall, dan f1-score menunjukkan performa yang baik (rata-rata diatas >90%)

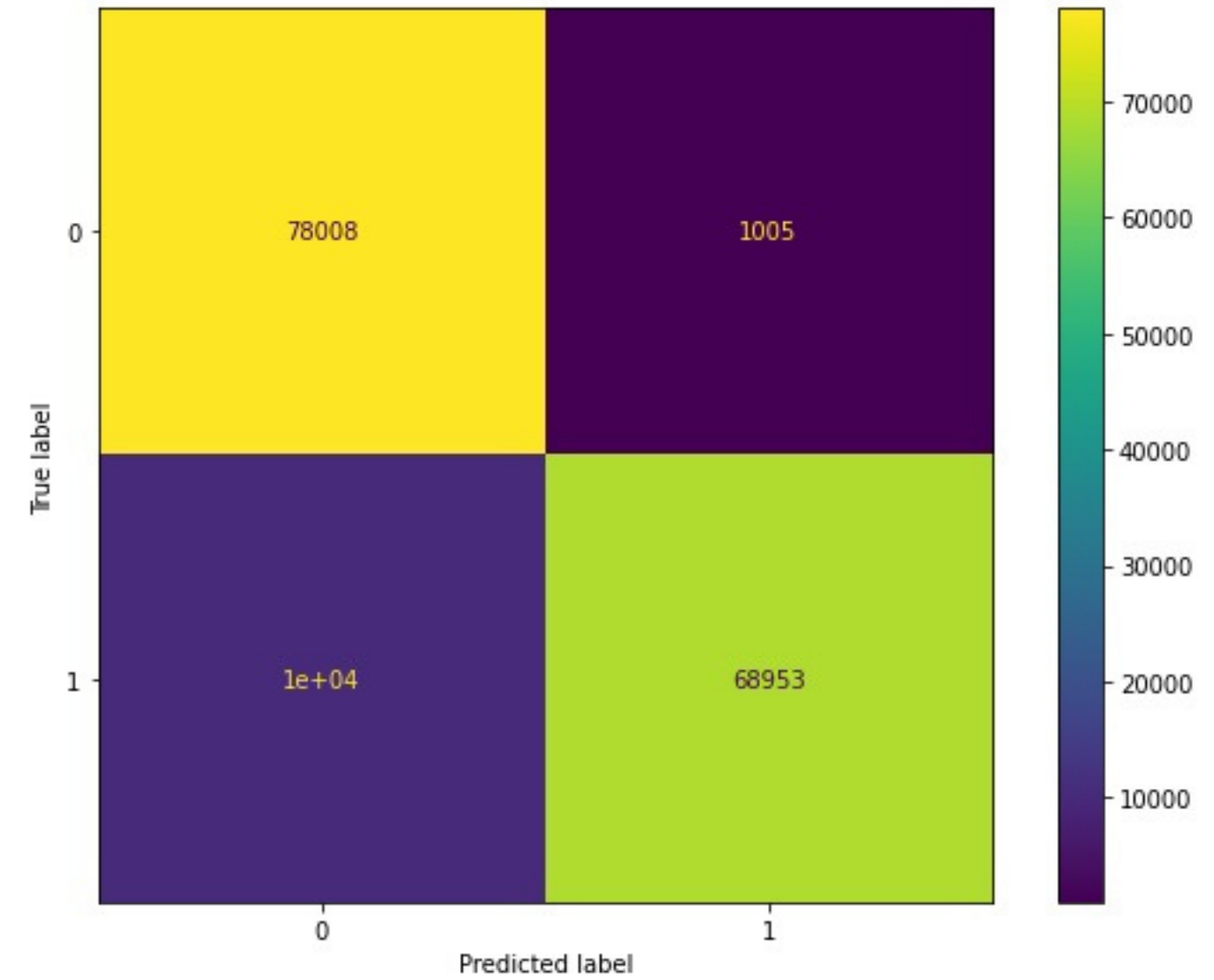
XGBoost

Confusion Matrix Evaluation

Train Set



Test Set



XGBoost

Confusion Matrix Evaluation

	precision	recall	f1-score	support
0	0.88	0.99	0.93	316539
1	0.99	0.87	0.92	316344
accuracy			0.93	632883
macro avg	0.93	0.93	0.93	632883
weighted avg	0.93	0.93	0.93	632883

Train Set

	precision	recall	f1-score	support
0	0.88	0.99	0.93	79013
1	0.99	0.87	0.92	79208
accuracy			0.93	158221
macro avg	0.93	0.93	0.93	158221
weighted avg	0.93	0.93	0.93	158221

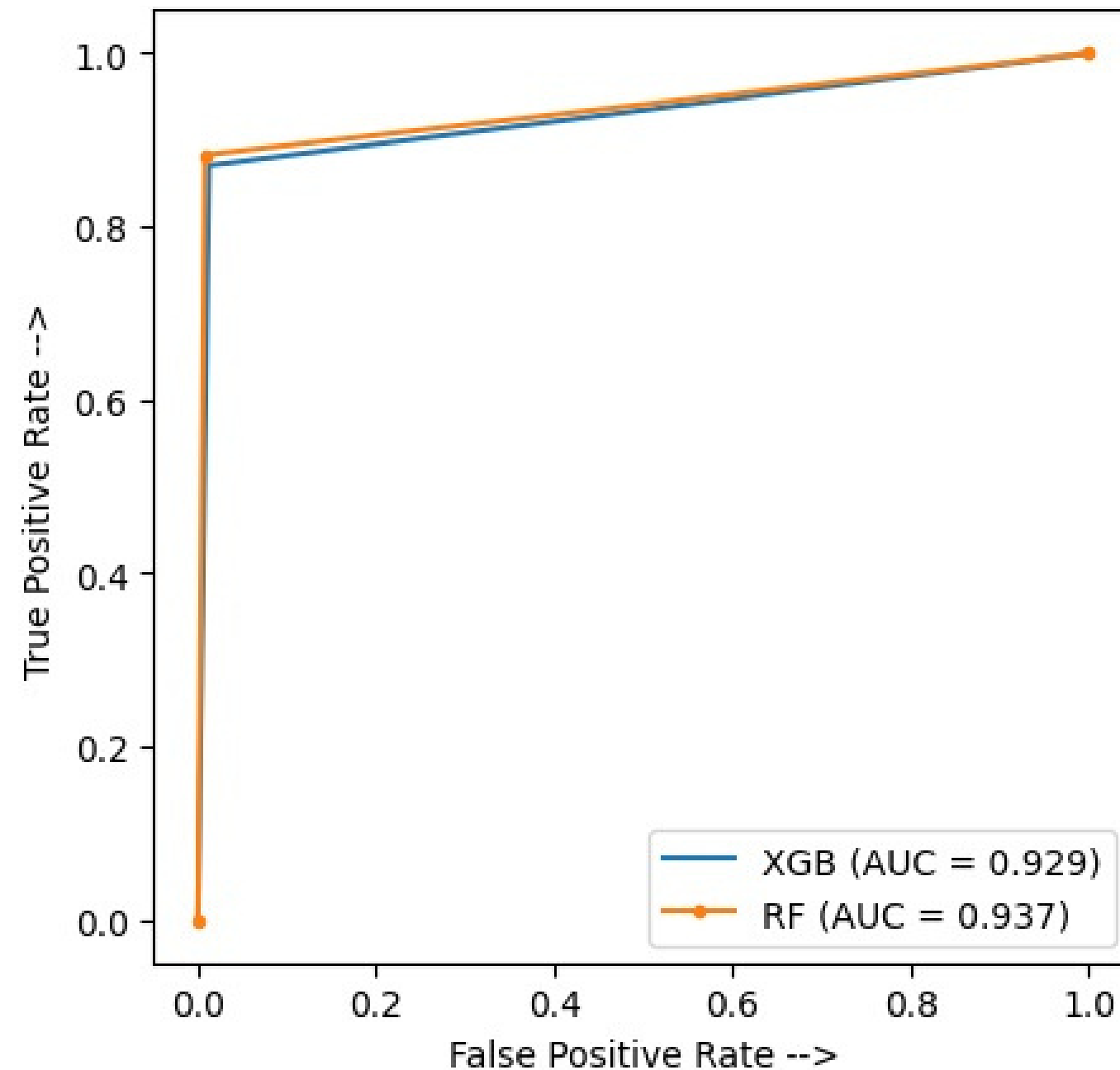
Test Set

Keterangan:

- terlihat bahwa performa XGBoost baik train set dan test set memiliki accuracy, precision, recall, dan f1-score menunjukkan performa yang baik (rata-rata diatas >90%)
- XGBoost lebih disarankan untuk dipilih dalam pemodelan karena tidak terjadi overfitting dan memiliki performa yang baik untuk memprediksi credit risk.
- Model dapat ditingkatkan dengan melakukan Hyperparamter tuning agar performa model menjadi lebih stabil (tuning membutuhkan waktu yang lama untuk proses customisasi performa model).
- Tuning model tidak dilakukan dalam project ini karena run time memakan waktu yang lama (>30 menit)

ROC-AOC Metric

Confusion Matrix Evaluation

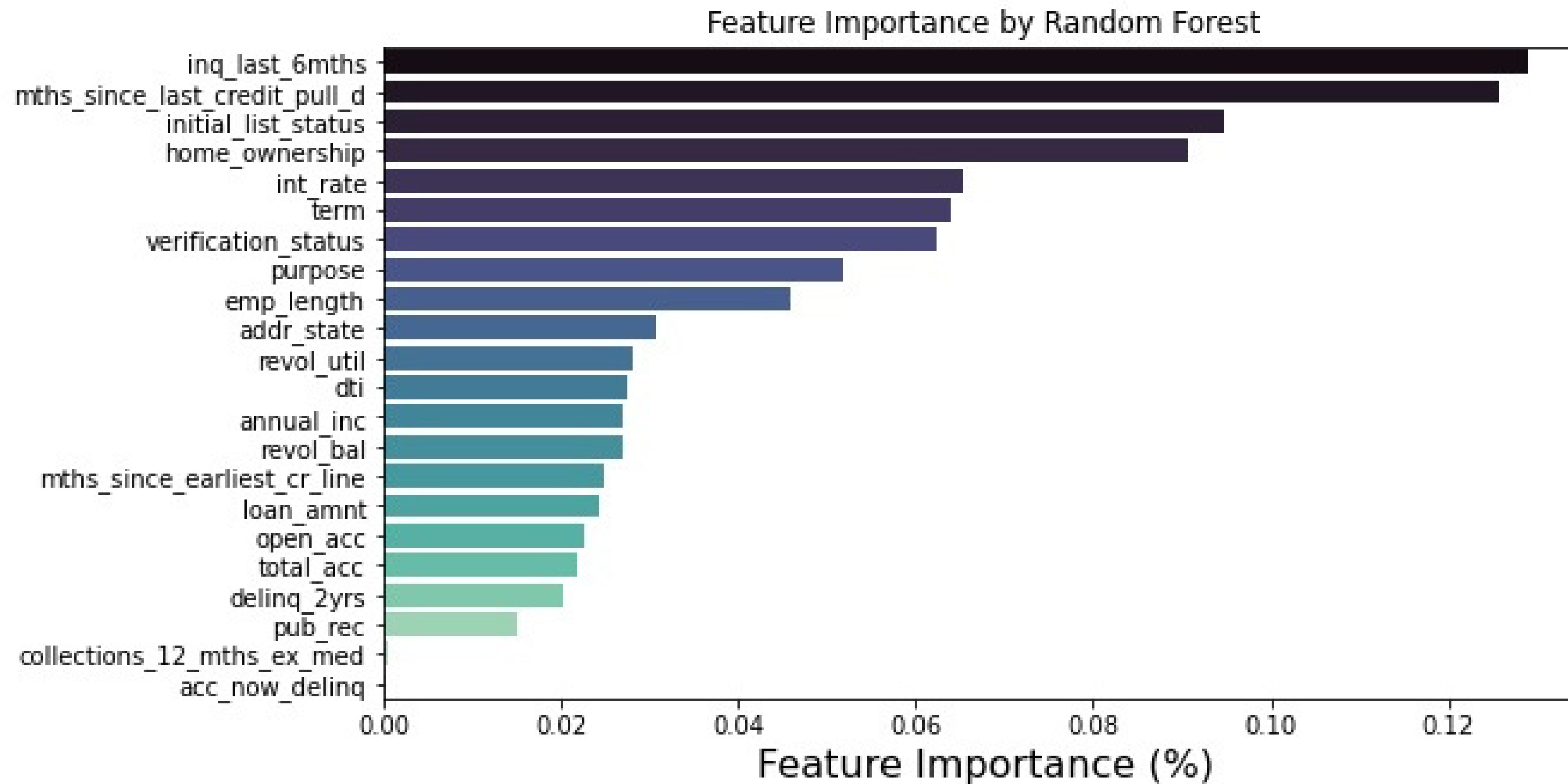


Keterangan:

- Berdasarkan grafik dari ROC-AUC matrix disamping menunjukkan bahwa kedua model (Random Forest dan XGBoost) memiliki score AUC yang baik ($AUC > 70\%$) yang dapat dipahami bahwa performa dari kedua model tersebut sudah proper untuk memprediksi "Bad Loans"
- Score AUC Random Forest (0.93) $>$ XGBoost (0.92)
- Kedua model ini lebih sering digunakan disaat dataset memiliki banyak outliers (rasionalisasi: model ML esemble tree tidak sensitif terhadap outliers, berbeda dengan distance-based model (KNN, K-Means, LogReg, dll) yang sensitif terhadap outliers)

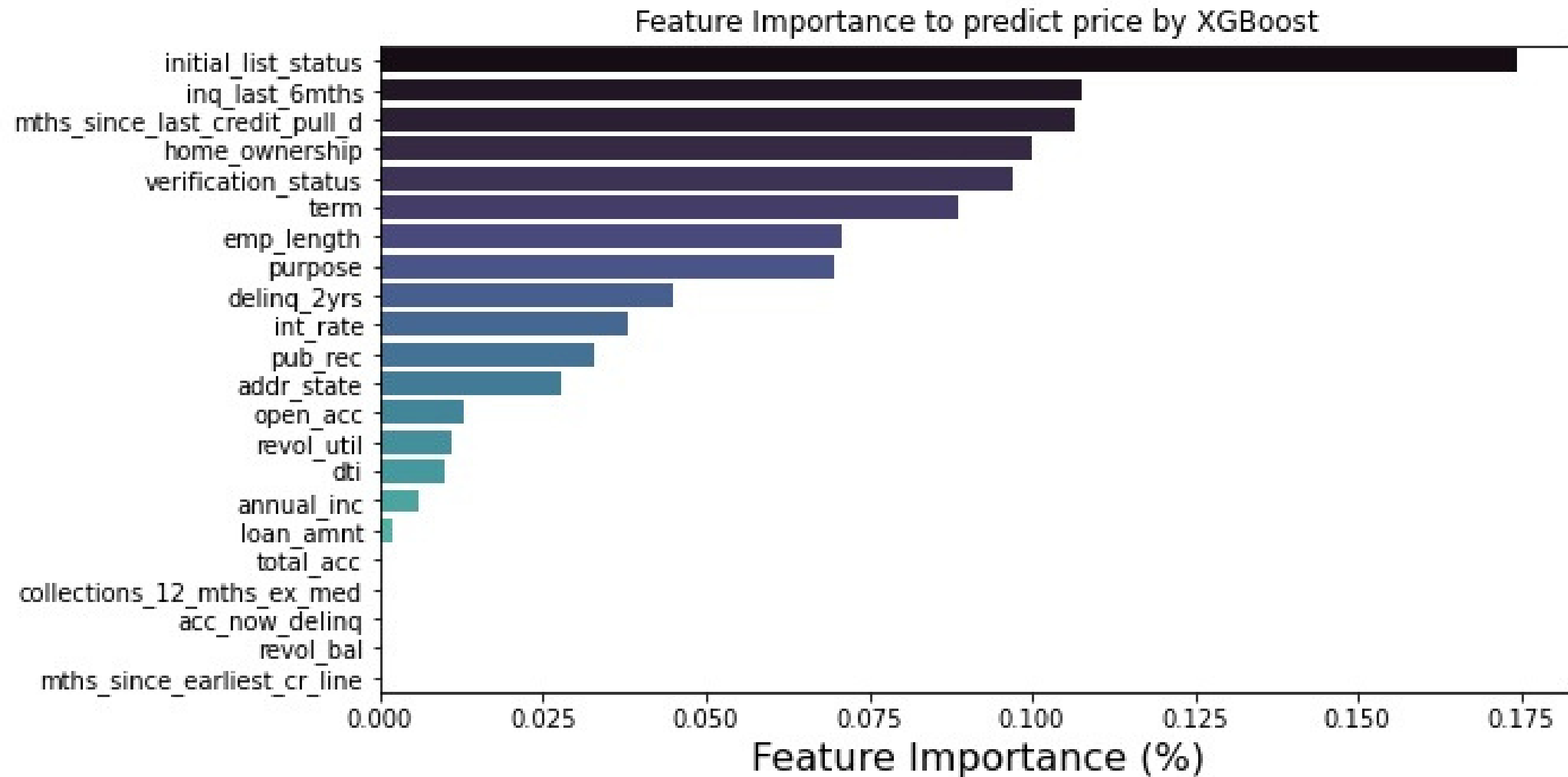
Feature Importance

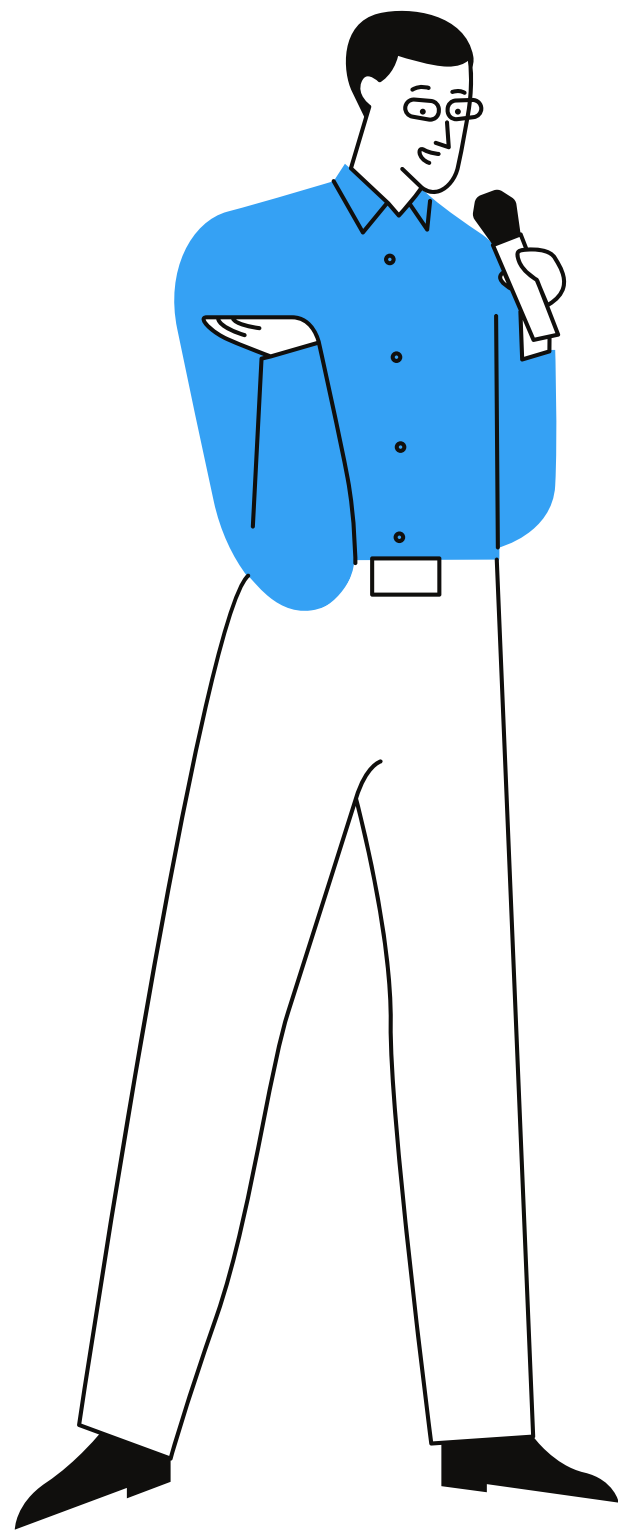
Random Forest



Feature Importance

XGBoost





***Thank
you!***

