

# Credit Risk Prediction With Machine Learning

**ID/X Partners - Data Scientist**

Achmad Wafi Makarim



## Achmad Wafi Makarim

### About

Fresh graduate from Brawijaya University, holding a degree in Information Technology, I recently completed the Data & Business Analytics bootcamp offered by Ruangguru. This intensive program provided hands-on experience and in-depth knowledge in data analytics, allowing me to acquire valuable skills in data cleaning, exploratory data analysis (EDA), and data visualization. The bootcamp also exposed me to real-world applications and business scenarios, reinforcing my belief in the transformative power of analytics. Fueled by this comprehensive education, I am eager to apply my analytical skills in practical settings and contribute effectively to future projects.

### Experience

#### Data & Business Analytics Bootcamp Ruangguru

- Make e-commerce analysis using statistical techniques, then evaluate the business and provide the best recommendations.
- Create a business plan using a business model canvas table.
- Used Python, Tableau, and SQL to analyze data.
- Create reports and dashboards that can be used to understand business performance.

#### Datawarehouse Project University

- Lead the process of database creation flow to define multiple How Might We (HMW) statements and gather solution ideas.
- Creating a mind map based on the flow of database creation that serves to make it more organized and structured.
- Looking for data sources from various sources to be analyzed.
- Processing data that has been analyzed to become a large data set.

# Project Portfolio

**Perusahaan multifinance perlu meningkatkan keakuratan penilaian risiko kredit untuk mengoptimalkan keputusan bisnis dan mengurangi kerugian. Kami mengembangkan model machine learning menggunakan data pinjaman dari Lending Club (2007-2014) untuk memprediksi risiko kredit, dengan fokus pada metrik bisnis seperti kerugian dan margin keuntungan bersih. Analisis data ini bertujuan untuk mengidentifikasi pola yang mengindikasikan pinjaman berpotensi buruk atau berisiko, tanpa asumsi yang kuat, untuk mendukung pengambilan keputusan investasi.**

**Project explanation video [here!](#)**



# About Company

**id/x partners didirikan pada tahun 2002 oleh mantan bankir dan konsultan manajemen yang memiliki pengalaman luas dalam manajemen siklus dan proses kredit, pengembangan scoring, dan manajemen kinerja. Pengalaman gabungan kami telah melayani korporasi di seluruh wilayah Asia dan Australia serta di berbagai industri, khususnya layanan keuangan, telekomunikasi, manufaktur, dan ritel.**



**id/x partners menyediakan layanan konsultasi yang mengkhususkan diri dalam memanfaatkan solusi analitik data dan pengambilan keputusan (DAD) yang dikombinasikan dengan disiplin manajemen risiko dan pemasaran terintegrasi untuk membantu klien mengoptimalkan profitabilitas portofolio dan proses bisnis.**

**Layanan konsultasi yang komprehensif dan solusi teknologi yang ditawarkan oleh id/x partners menjadikannya sebagai penyedia layanan terpadu.**

# Study Case

**id/x partners didirikan pada tahun 2002 oleh mantan bankir dan konsultan manajemen yang memiliki pengalaman luas dalam manajemen siklus dan proses kredit, pengembangan scoring, dan manajemen kinerja. Pengalaman gabungan kami telah melayani korporasi di seluruh wilayah Asia dan Australia serta di berbagai industri, khususnya layanan keuangan, telekomunikasi, manufaktur, dan ritel.**



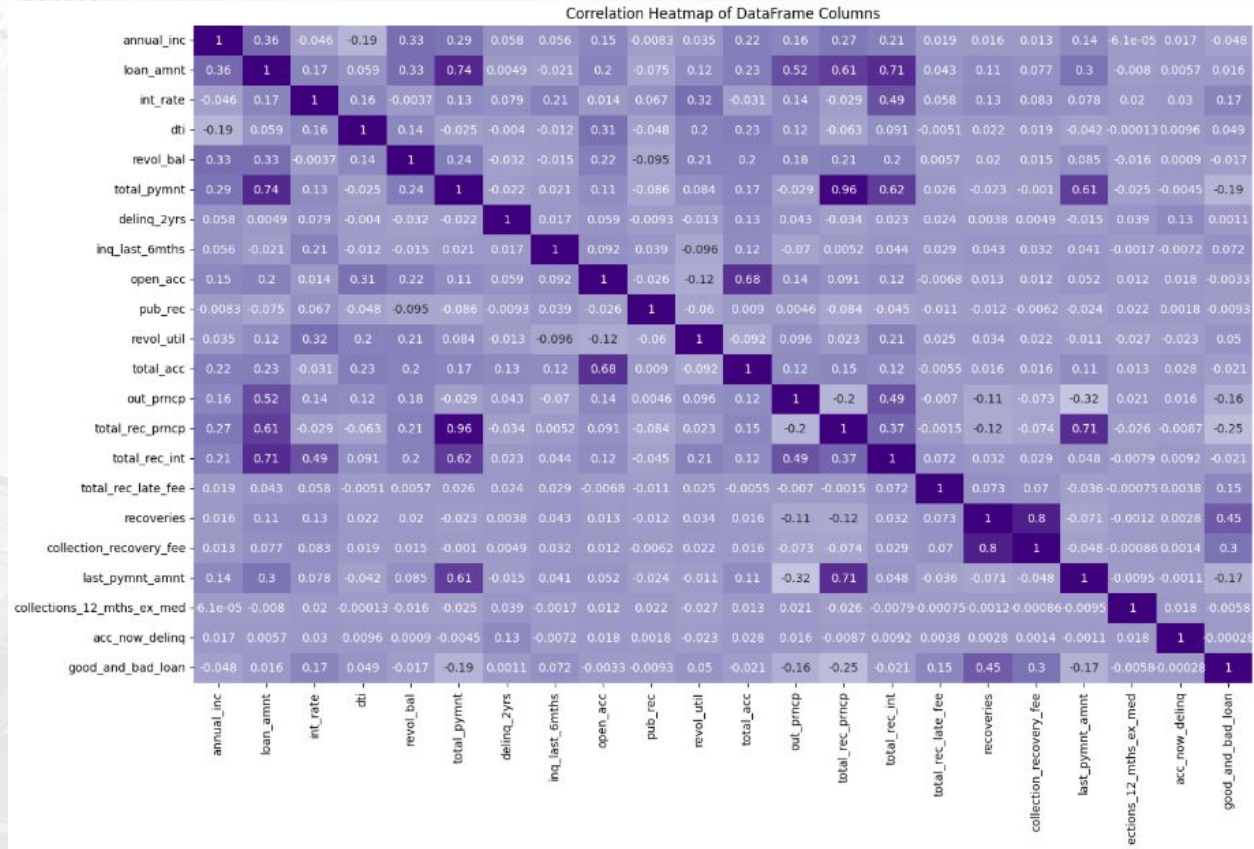
**id/x partners menyediakan layanan konsultasi yang mengkhususkan diri dalam memanfaatkan solusi analitik data dan pengambilan keputusan (DAD) yang dikombinasikan dengan disiplin manajemen risiko dan pemasaran terintegrasi untuk membantu klien mengoptimalkan profitabilitas portofolio dan proses bisnis.**

**Layanan konsultasi yang komprehensif dan solusi teknologi yang ditawarkan oleh id/x partners menjadikannya sebagai penyedia layanan terpadu.**

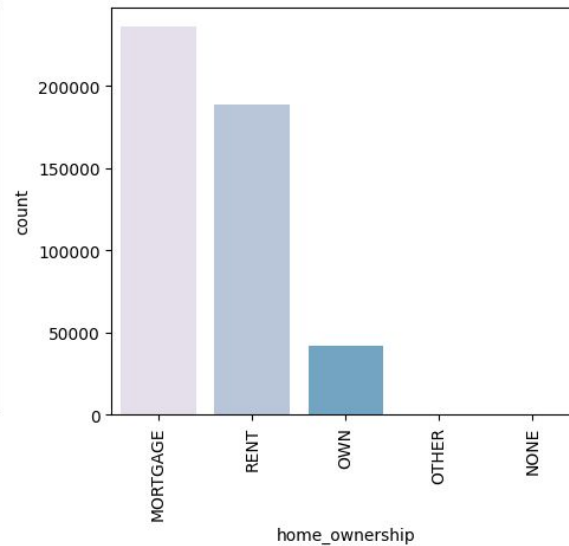
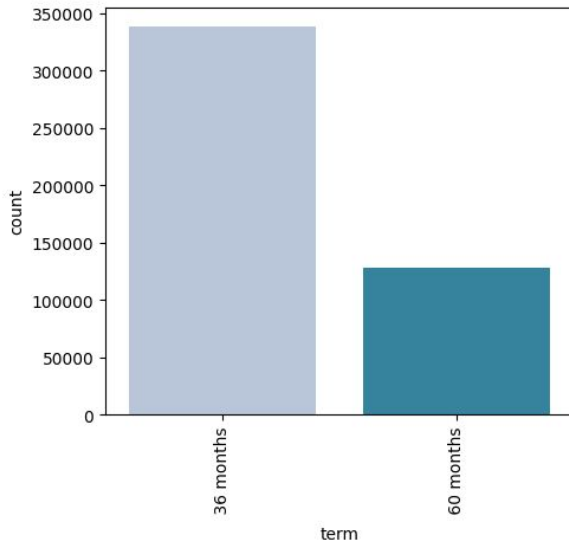
# **Explanatory Data Analysis**



# Check Correlation



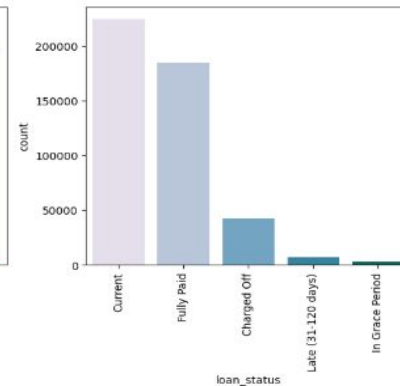
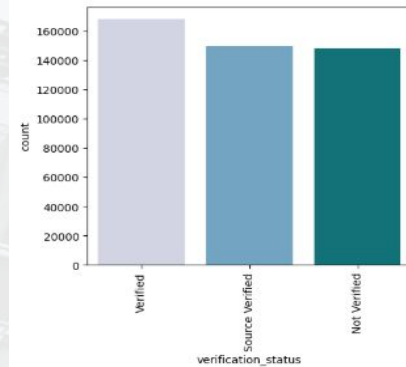
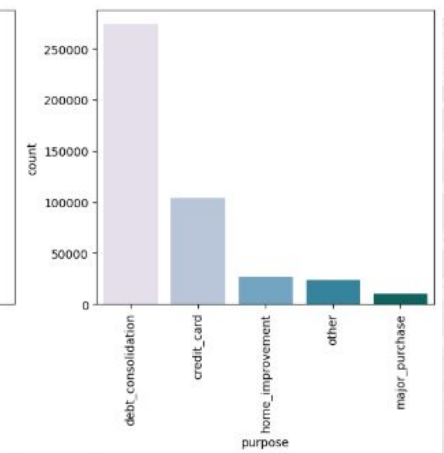
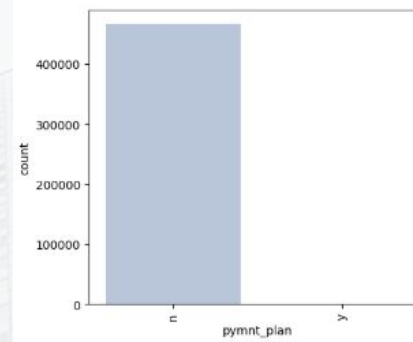
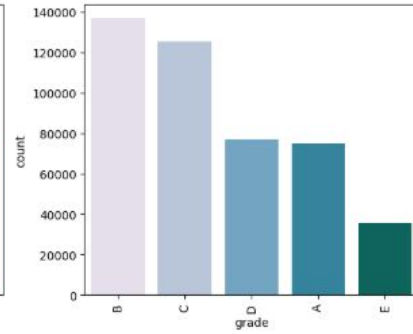
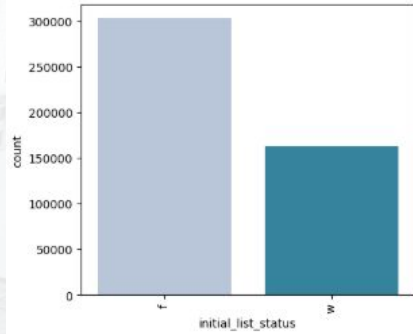
# Univariate Analysis



- Jangka waktu angsuran terbanyak peminjam yaitu pada 36 bulan
- Kebanyakan nasabah melakukan pinjaman untuk membuat bisnis perumahan yang tentunya perlu banyak dana
- Nasabah yang menggunakan uang pinjaman untuk sewa juga termasuk banyak daripada untuk pembelian rumah sendiri



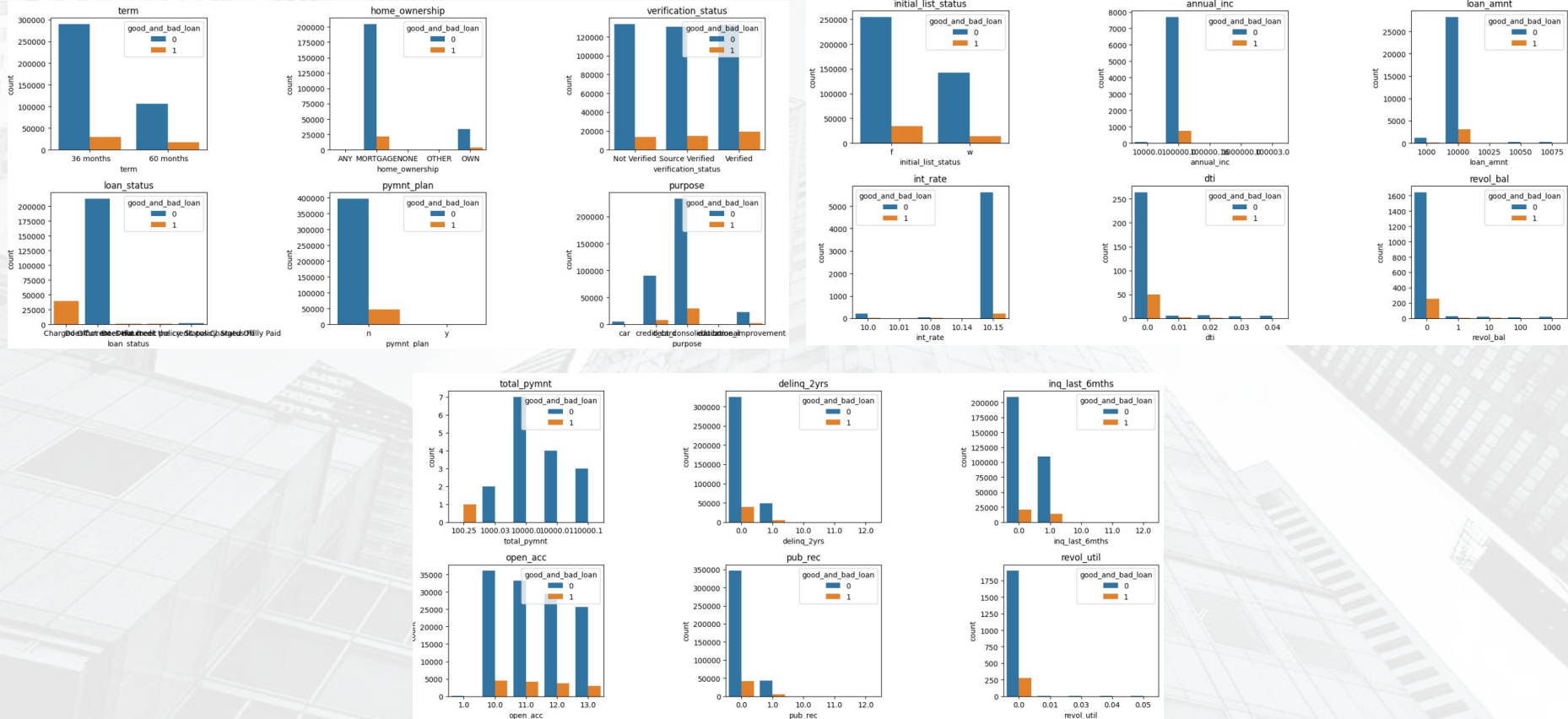
# Univariate Analysis



# Univariate Analysis

- Proses peminjaman kepada nasabah masih terbilang cukup baik dimana masih banyak yang melakukan peminjaman dan banyak juga yang telah melunasi pinjaman
- Tetapi perlu diperhatikan untuk nasabah yang memiliki status Charged Off yang membuat kerugian dan hal ini perlu diatasi agar tidak terjadi kerugian pada masa mendatang
- Terdapat berbagai kategori yang disediakan oleh peminjam dan kebanyakan nasabah termasuk kedalam kategori debt\_consolidation
- Perlu diperhatikan untuk bagian verifikasi masih banyak yang belum terverifikasi yang harusnya nasabah melakukan verifikasi terlebih dahulu agar tidak adanya kerugian karena kurangnya data nasabah diakibatkan nasabah yang tidak terverifikasi sebelumnya
- Sebagian besar pinjaman berada pada grade B dan C yang artinya mereka lebih memilih pinjaman dengan resiko sedang

# Multivariate Analysis

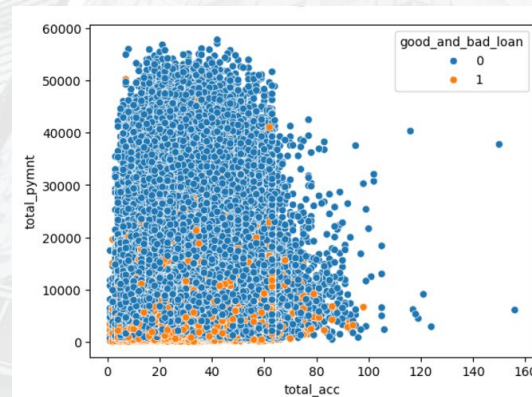
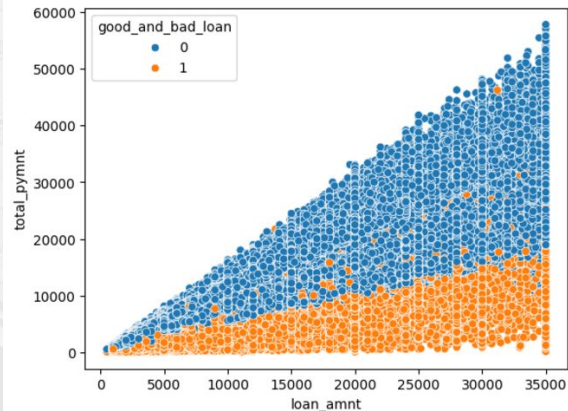
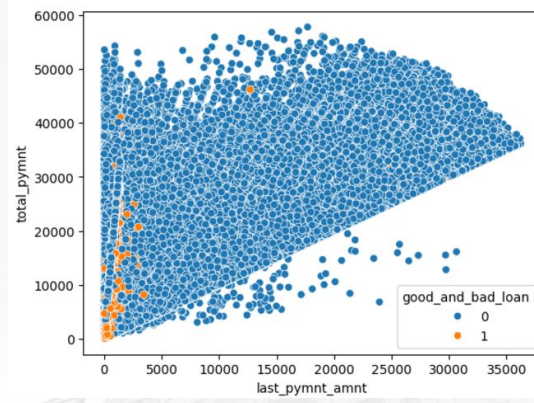
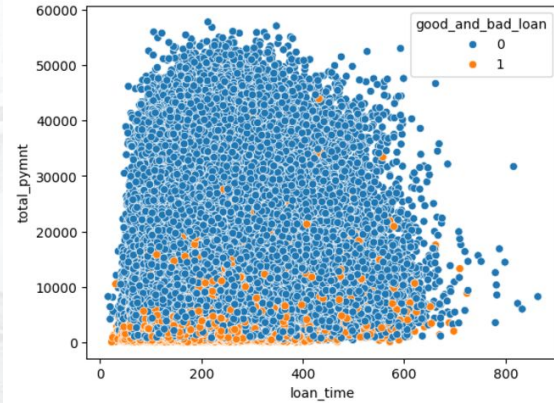




# Univariate Analysis

- Jangka waktu angsuran terbanyak peminjam yaitu pada 36 bulan yang memiliki status good loan atau pinjaman yang baik
- Meskipun banyaknya pinjaman untuk keperluan bisnis perumahan tetapi masih terbilang pinjaman yang baik
- Pada pilihan kategori yang disediakan oleh peminjam dan kebanyakan nasabah termasuk kedalam kategori debt\_consolidation juga masih terbilang good loan
- Sebagian besar good load pinjaman berada pada grade B dan C yang merupakan pinjaman dengan resiko sedang

# Multivariate Analysis

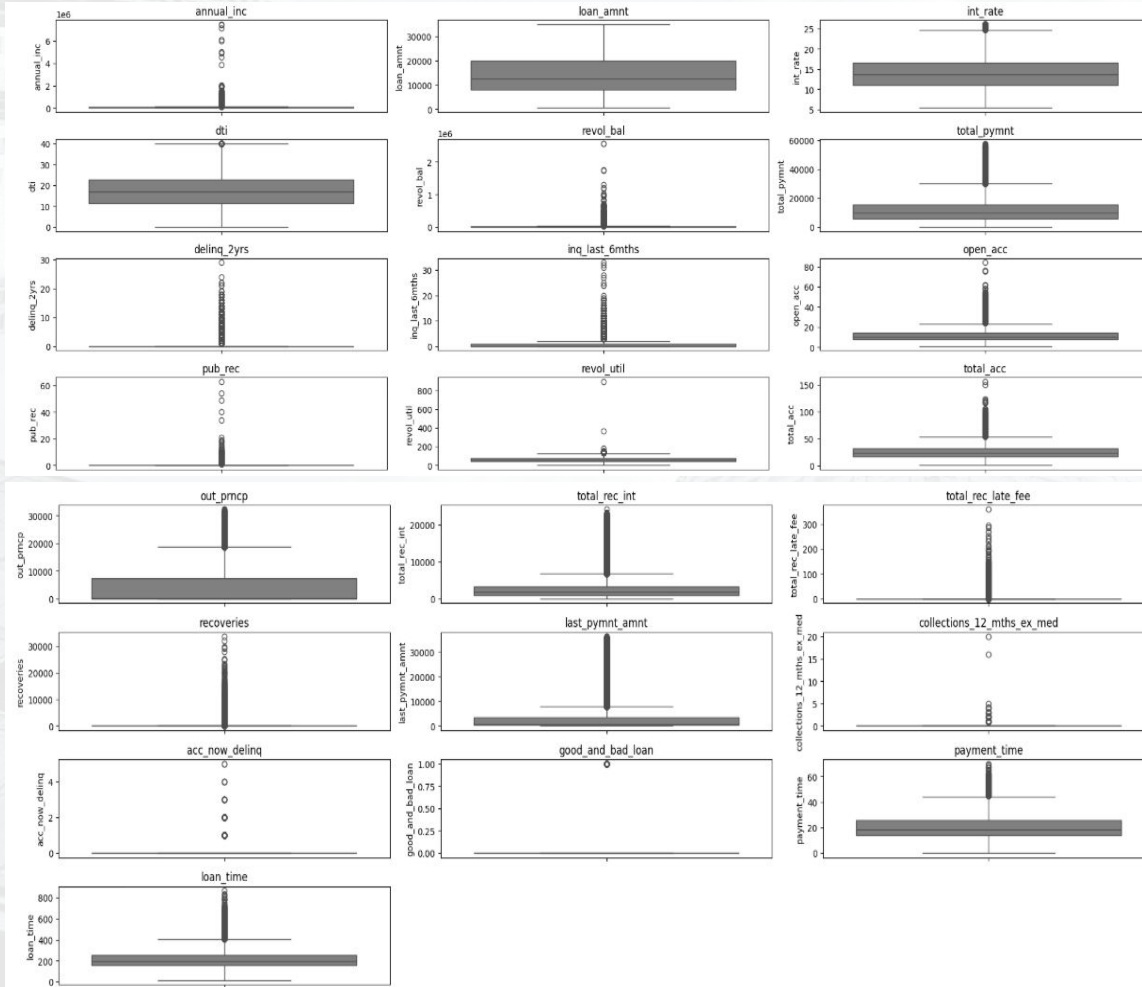


- Total pembayaran yang dilakukan oleh nasabah masih memiliki waktu pinjaman bagus yang artinya nasabah cenderung membayar pinjaman dengan tepat waktu
- Nasabah melakukan pinjaman dengan pembayaran yang sesuai dengan total yang dipinjam
- Meskipun nasabah memiliki batas kredit yang kecil tetapi masih membayar kredit dengan jumlah yang sesuai juga



# **Data** **Preprocessing**





# Handling Outliers

	Feature	Outlier_Percentage
0	annual_inc	4.224088
1	loan_amnt	0.000000
2	int_rate	1.054109
3	dti	0.032175
4	revol_bal	4.906053
5	total_pymnt	4.011016
6	delinq_2yrs	17.970260
7	inq_last_6mths	7.907508
8	open_acc	2.285966
9	pub_rec	12.664388
10	revol_util	0.004050
11	total_acc	1.875572
12	out_prncp	4.878603
13	total_rec_int	6.644152
14	total_rec_late_fee	2.301941
15	recoveries	5.171324
16	last_pymnt_amnt	14.686208
17	collections_12_mths_ex_med	0.832488
18	acc_now_delinq	0.372370
19	good_and_bad_loan	10.743817

## Before

- Melihat persentase outlier pada tiap kolom, dapat dilihat bahwa sebagian besar kolom memiliki outlier, yang menunjukkan adanya nilai ekstrim atau tidak biasa dalam data.
- Namun ada beberapa kolom seperti 'loan\_amnt' yang tidak menunjukkan adanya outlier signifikan, menandakan distribusi data yang lebih teratur.
- Penghapusan outlier dilakukan dengan pertimbangan maksimal outlier yang dapat dihapus adalah 5% dikarenakan apabila melebihi dari 5% akan dapat mempengaruhi hasil pemodelan

	Feature	Outlier_Percentage
0	annual_inc	2.147095
1	loan_amnt	0.962833
2	int_rate	0.387353
3	dti	0.057548
4	revol_bal	2.508158
5	total_pymnt	1.625366
6	delinq_2yrs	16.838191
7	inq_last_6mths	7.319695
8	open_acc	0.490472
9	pub_rec	13.764193
10	revol_util	0.003213
11	total_acc	0.000000
12	out_prncp	3.470114
13	total_rec_int	5.287696
14	total_rec_late_fee	0.000000
15	recoveries	5.301718
16	last_pymnt_amnt	13.360481
17	collections_12_mths_ex_med	0.000000
18	acc_now_delinq	0.000000
19	good_and_bad_loan	10.727003

## After

# Modelling



# Penggunaan Metrik

- Penggunaan metrik berfokus pada nilai akurasi untuk mengukur seberapa banyak hasil prediksi yang benar yang telah dibuat oleh model
- Nilai akurasi juga memberikan gambaran seberapa baik machine learning yang digunakan

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

## Logistic Regression

Accuracy (Test Set): 0.981  
Accuracy (Train Set): 0.981  
Precision (Test Set): 0.985  
Precision (Train Set): 0.985  
Recall (Test Set): 0.835  
Recall (Train Set): 0.840  
F1-Score (Test Set): 0.904  
F1-Score (Train Set): 0.907  
ROC AUC (Test Set): 0.966  
ROC AUC (Train Set): 0.966  
Recall (Crossval Train): 0.842  
Recall (Crossval Test): 0.840

## Naive Bayes

Accuracy (Test Set): 0.946  
Accuracy (Train Set): 0.946  
Precision (Test Set): 0.999  
Precision (Train Set): 0.999  
Recall (Test Set): 0.501  
Recall (Train Set): 0.495  
F1-Score (Test Set): 0.667  
F1-Score (Train Set): 0.662  
ROC AUC (Test Set): 0.943  
ROC AUC (Train Set): 0.942  
Recall (Crossval Train): 0.497  
Recall (Crossval Test): 0.498

- Pengujian dilakukan dengan menggunakan 2 metode yaitu logistic regression dan naive bayes
- Model logistic memiliki nilai akurasi yang lebih baik daripada naive bayes dengan nilai 98.1% dibandingkan naive bayes yang memiliki nilai 94.6%



# Kesimpulan

1. **Melakukan Oversampling** : Alasannya adalah karena data tersebut memiliki hasil yang tidak seimbang, hanya 22% data yang merupakan pinjaman berisiko.
2. **Memberikan tambahan kolom** : Untuk mendapatkan informasi yang lebih lengkap perlunya kolom lebih banyak. Peminjam yang memberikan informasi lebih lengkap cenderung memiliki motivasi lebih besar untuk membayar pinjaman tepat waktu. Terutama untuk kategori pinjaman usaha kecil, yang secara statistik memiliki risiko lebih tinggi. Dengan adanya deskripsi atau informasi tambahan dapat menjadi indikator bahwa peminjam memiliki rencana atau komitmen yang lebih jelas terhadap pengelolaan pinjaman mereka. Hal ini dapat menjadi pertimbangan penting bagi tim bisnis dalam mengevaluasi kelayakan pinjaman.



# Rekomendasi

## 1. Mengidentifikasi Pola Risiko Berdasarkan Total Pembayaran

- Data menunjukkan bahwa peminjam yang memenuhi kewajiban pembayaran cenderung memiliki total pembayaran yang stabil dan konsisten sejak awal hingga akhir masa pinjaman.
- **Rekomendasi** : Tetap lakukan pemantauan terhadap pola pembayaran untuk memastikan stabilitas. Selain itu, tawarkan program insentif, seperti diskon bunga atau bonus loyalitas, bagi peminjam yang melunasi pinjaman lebih awal atau tepat waktu.

## 2. Segmen Berdasarkan Tujuan Pinjaman

- Sebagian besar peminjam menggunakan pinjaman untuk melunasi utang sebelumnya, dan kelompok ini menunjukkan kinerja pembayaran yang konsisten dan baik.
- **Rekomendasi** : Tambahkan analisis mendalam untuk memahami pola pengelolaan keuangan peminjam yang baik ini. Gunakan temuan tersebut untuk mengembangkan program insentif yang dapat mendorong peminjam melunasi pinjaman lebih awal atau meningkatkan loyalitas.

# Thank You



**Rakamin**  
Academy



id/x

partners