



LOAN APPROVAL ANALYSIS & PREDICTION

20 MAY 2023



OUR TEAM



AZRIEL HUTAGALUNG
OUR BEST MENTOR



RIAN DWI HARYONO
DATA SCIENTIST



NICO BERLINSON F.
DATA SCIENTIST



M. RIFKI OSKAR
DATA SCIENTIST



TAUFIQUR R
DATA SCIENTIST



MITHA PUTRI O.
DATA SCIENTIST



GRACE WULANDARI
DATA SCIENTIST



EGA SAHERTI
DATA SCIENTIST



GUSTI AYU S
DATA SCIENTIST

WHO WE ARE?

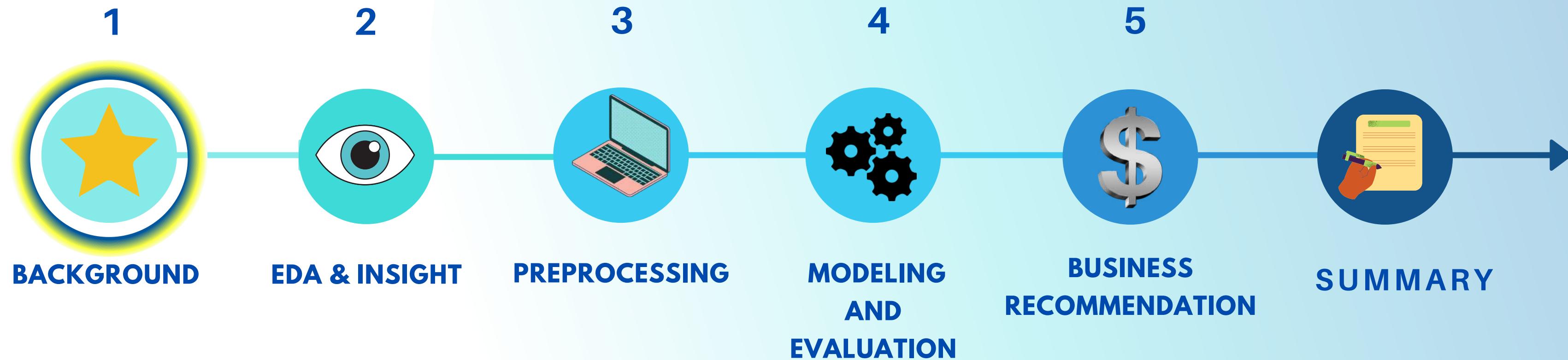


Jupyter Group

Jupyter.

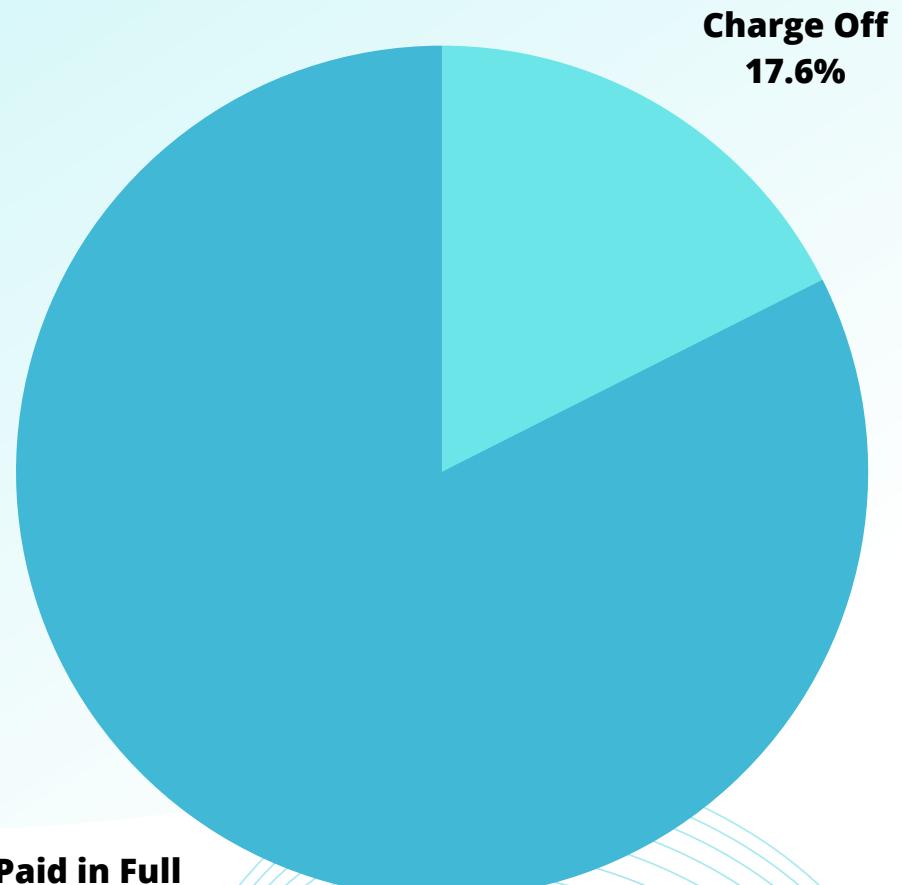
- **Data Science Team** dari perusahaan layanan **konsultasi data analytic and decisioning (DAD)**
- **Bermitra dengan SBA (Small Business Administration)** memberikan rekomendasi **nasabah yang berpotensi gagal bayar** dan **menganalisa faktor-faktor penyebabnya** berdasarkan **dataset perusahaan yang tersedia** guna **menurunkan rasio nasabah yang gagal bayar.**
- **SBA (Small Business Administration)** merupakan pihak ke 3 yang menjembatani bank sebagai pemberi pinjaman dengan peminjam sebagai nasabah sebagai peminjam

TABLE OF CONTENTS



PROBLEM STATEMENT

Terdapat 17.6% jumlah peminjam yang gagal bayar diantara rentang waktu 1988 - 2010



* Years : 1988 - 2010

GOAL

Menurunkan jumlah nasabah yang gagal bayar (Charge Off)

OBJECTIVE

Membuat model machine learning untuk melakukan profiling atau memprediksi profile nasabah yang berpotensi gagal bayar sehingga perusahaan mampu melakukan assessment nasabah lebih akurat dan meningkatkan performa kerja perusahaan

BUSINESS METRIC

↓ Default Rates (Low or High)



<https://www.sba.gov/>



1-800-827-5722

U.S. Small Business Administration (SBA)
409 3rd St SW, Washington, DC 20416, Amerika Serikat

Numerical Features

| | Skew | Outlier |
|-------------------|------|---------|
| Term | ✓ | ✓ |
| NoEmp | ✓ | ✓ |
| DisbursementGross | ✓ | ✓ |
| CreateJob | ✓ | ✓ |
| RetainedJob | ✓ | ✓ |
| ChgOffPrinGr | ✓ | ✓ |
| GrAppv | ✓ | ✓ |
| SBA_Appv | ✓ | ✓ |
| BalanceGross | ✓ | ✓ |

Categorical Features

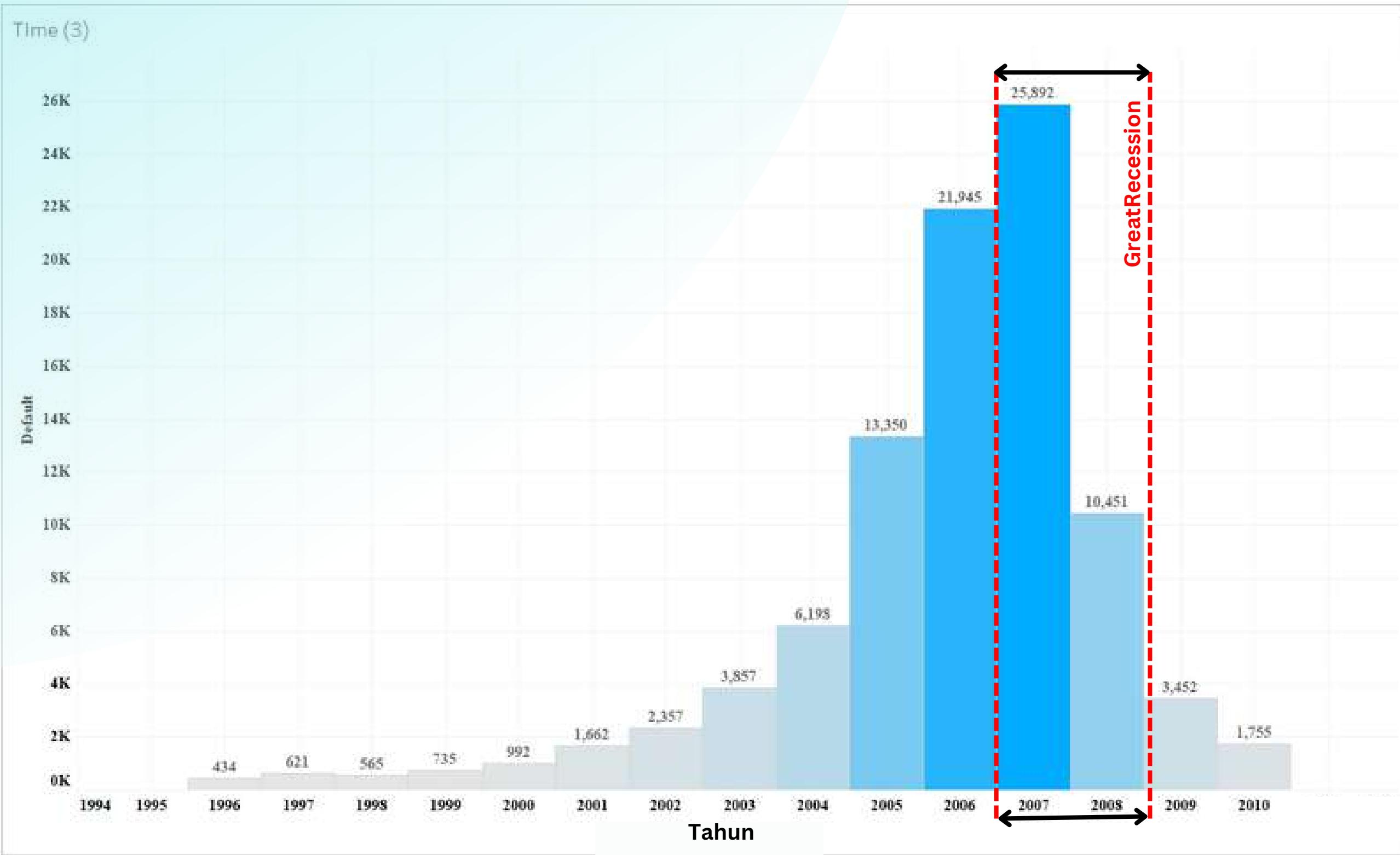
| | Null Value | #Unique |
|------------------|------------|---------|
| Name | ✓ | 779583 |
| City | ✓ | 32581 |
| State | ✓ | 51 |
| Bank | ✓ | 5802 |
| BankState | ✓ | 56 |
| ApprovalDate | □ | 9859 |
| ApprovalFY | □ | 70 |
| NewExist | ✓ | 2 |
| RevLineCr | ✓ | 18 |
| UrbanRural | □ | 3 |
| LoanNr_ChkDgt | □ | Nan |
| ZIP | □ | Nan |
| NAICS | □ | Nan |
| MIS_Status | ✓ | 2 |
| LowDoc | ✓ | 8 |
| ChgOffDate | ✓ | 6448 |
| DisbursementDate | ✓ | 8472 |
| FranchiseCode | □ | Nan |

- Terdapat 28 kolom, terdiri dari 9 kolom numerical, dan 19 kolom categorical, 899.163 baris (tanpa duplikat).
- Terdapat missing value pada beberapa kolom categorical.
- Terdapat data type yang tidak sesuai.
- mayoritas data ditemukan outlier dan skew ke kanan

Target:
MIS Status



Default dlm Tahun

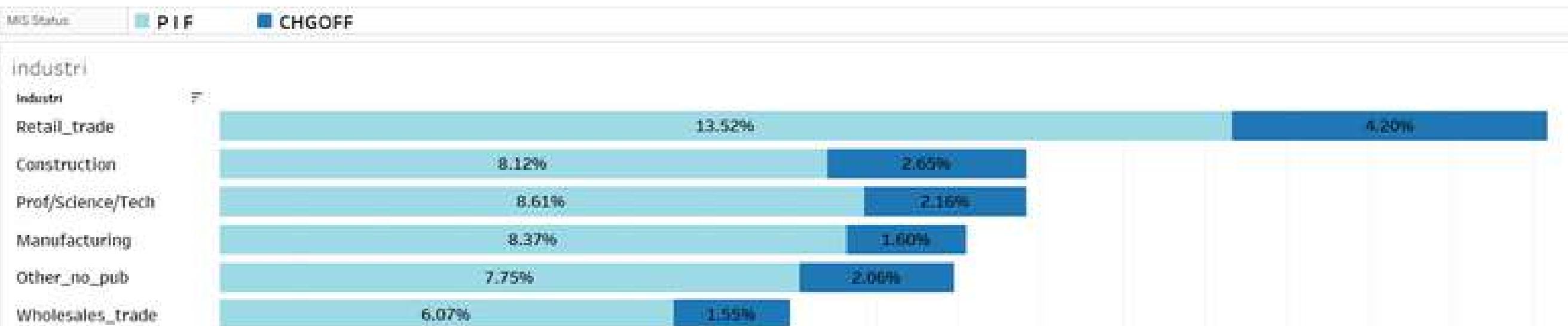


Default rate meningkat saat terjadi recession di USA pada akhir tahun 2007 - akhir tahun 2008

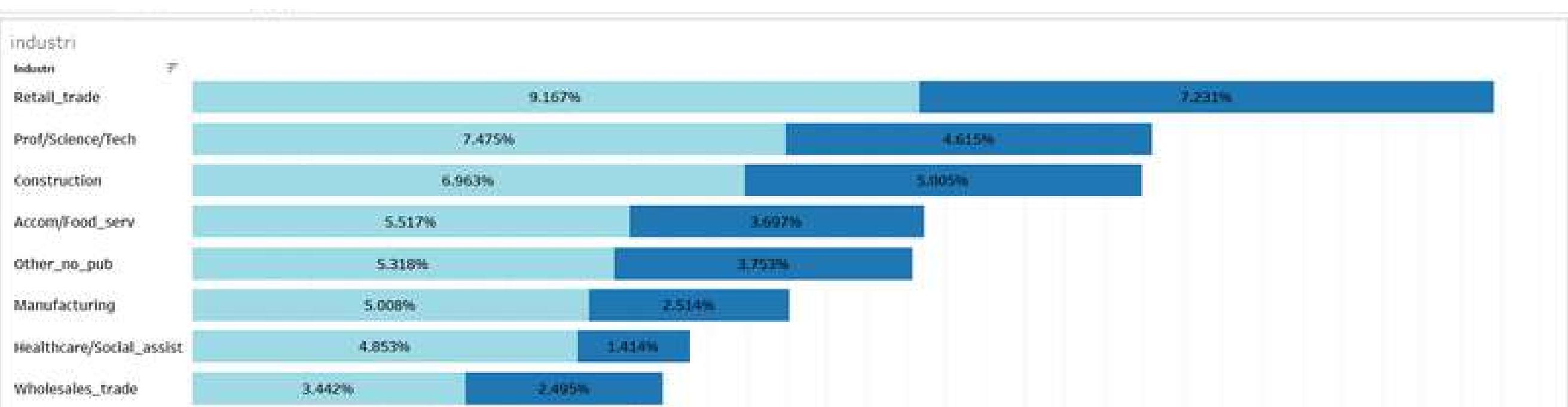
ref: <https://www.thestreet.com/dictionary/f/financial-crisis-2007-2008>



TREN INDUSTRI DEFAULT 1988 - 2010

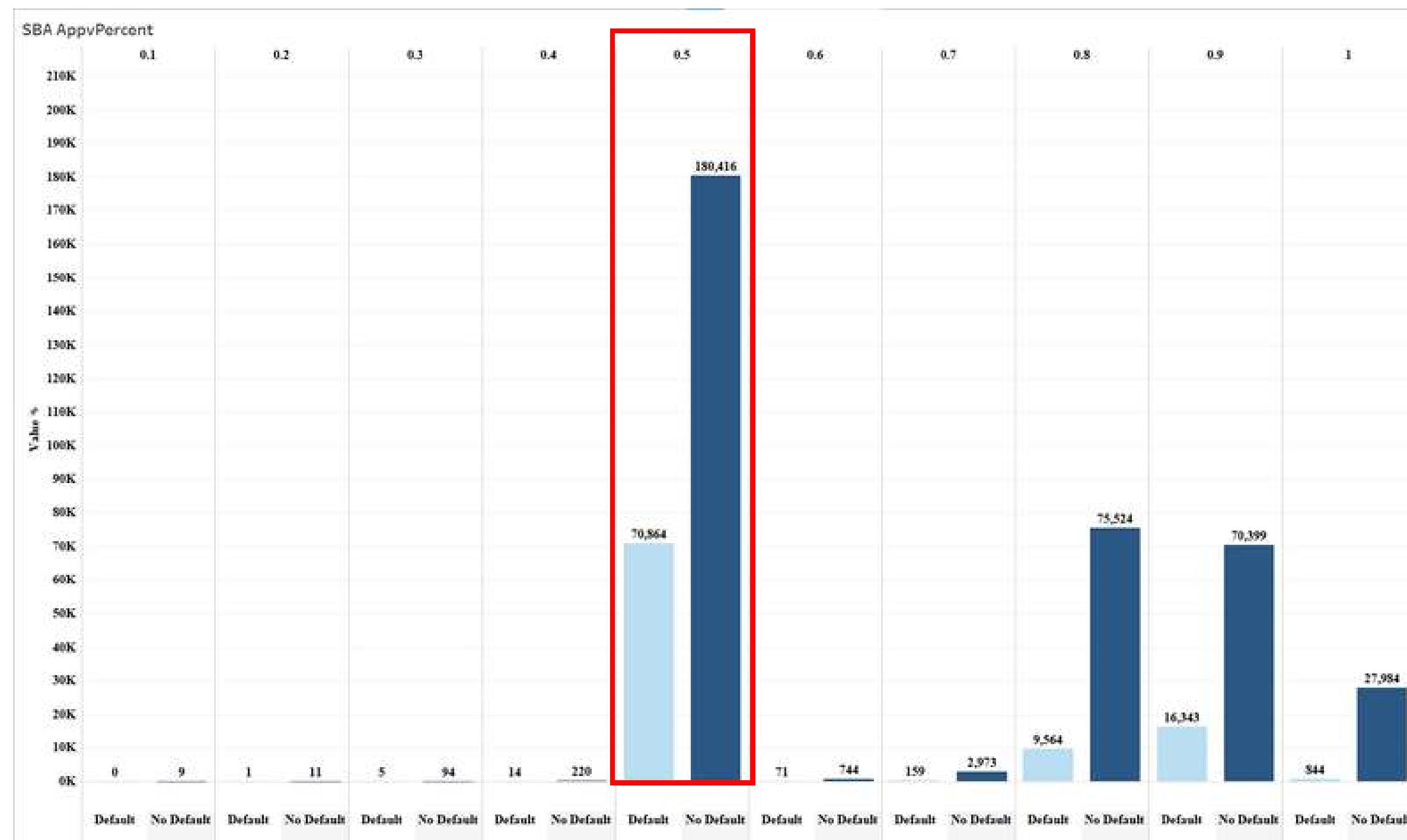


TREN INDUSTRI DEFAULT 2007-2009 (ON RECESSION)



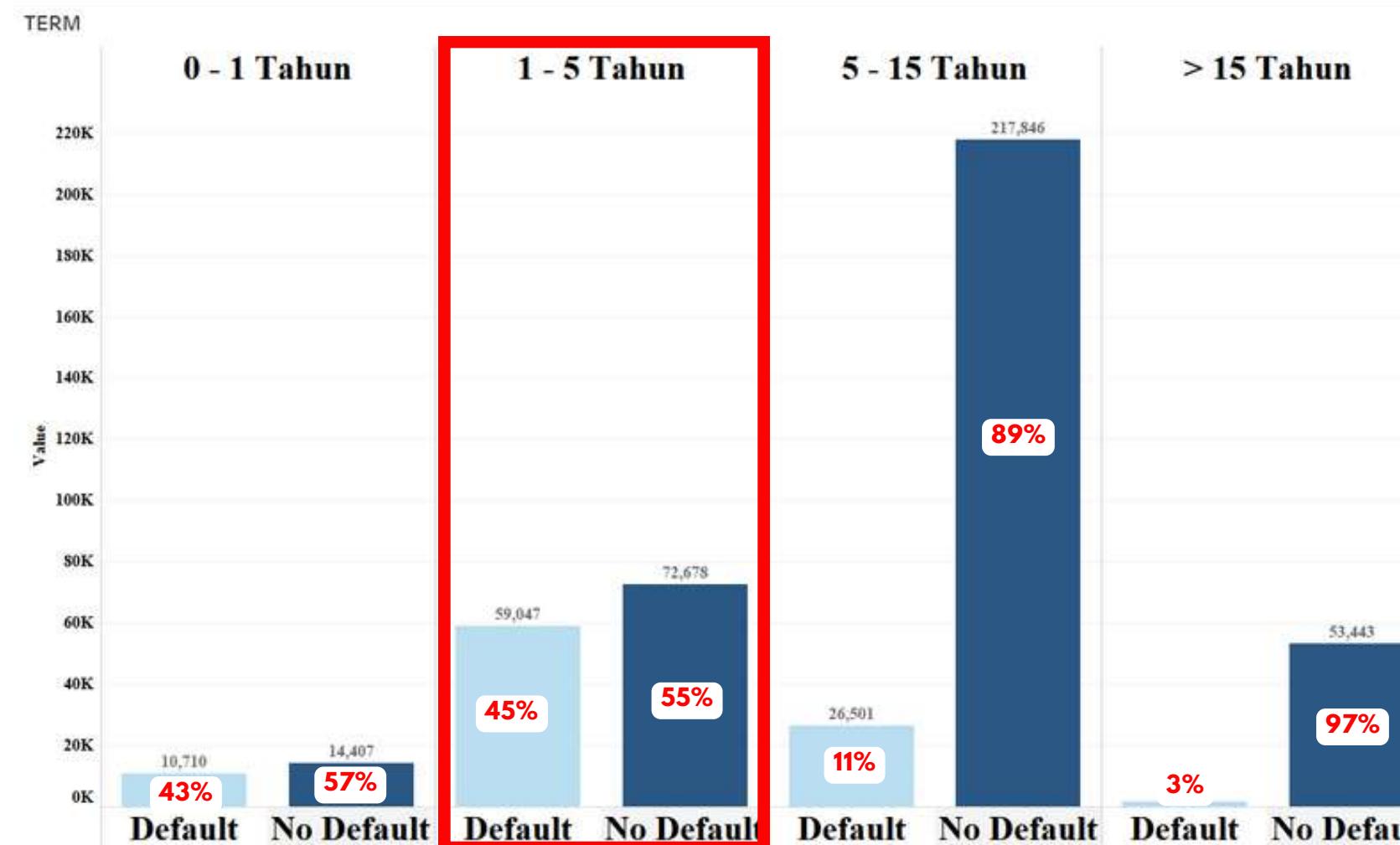
Retail Trade dan construction memiliki tingkat default yang cukup tinggi





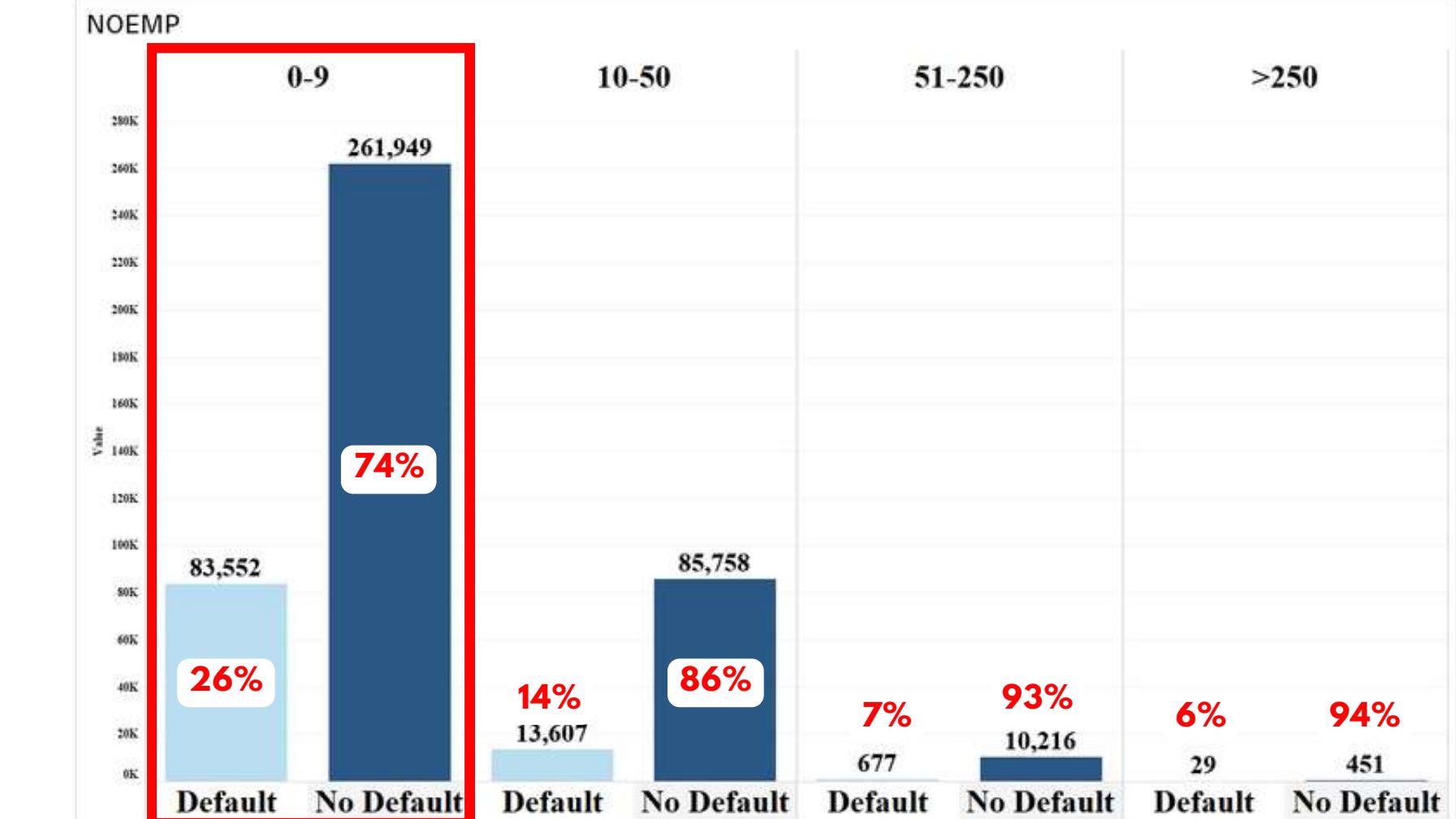
SBA Approve Percentage adalah presentasi jaminan yang diberikan oleh SBA, pinjaman yang memperoleh jaminan SBA mencapai 50% relatif memiliki tingkat gagal bayar tertinggi.

TENOR PINJAMAN



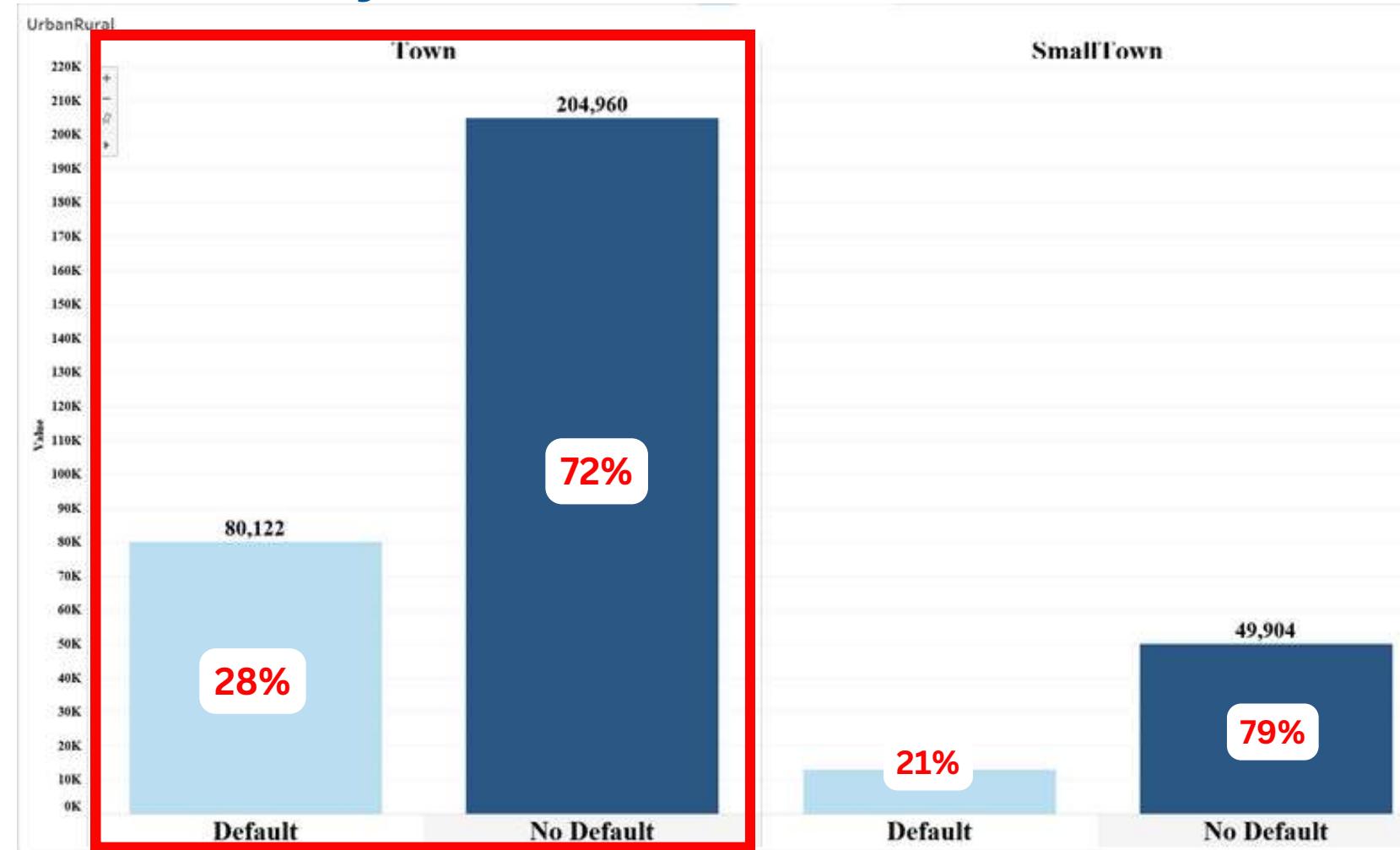
[Term] : tingkat gagal bayar relatif tinggi pada perusahaan dengan rentang tenor pinjaman antara 0-5 tahun.

JUMLAH KARYAWAN



[NoEmp] : tingkat gagal bayar relatif tinggi pada perusahaan yang memiliki jumlah karyawan <9 org atau yang masuk dalam kategori usaha mikro.

LOKASI PEMINJAM



[UrbanRural] : mayoritas peminjam berada di area perkotaan dengan % sebesar 81% dibanding area lainnya, dan dengan tingkat gagal bayar sebesar 28% dari semua peminjam yang berada di area perkotaan

KESAMAAN LOKASI PEMINJAM DENGAN BANK



[StateSame] : tingkat gagal bayar 2x lebih tinggi pada lokasi yang berbeda antara peminjam & bank.



PREPROCESSING

DATA CLEANSING



Handling Missing Value

- Drop null values
- Handling error values

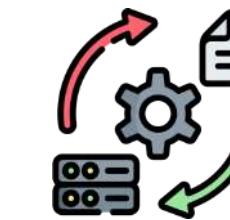


Handling duplicated data

- tidak terdapat data yang duplicate

Handling Outliers

- Menggunakan metode transform



Feature transformation

- melakukan standarization
- melakukan transformation menggunakan metode boxcox dikarenakan nilai skew yg dihasilkan lebih kecil dibandingkan log



Feature encoding

- mengubah feature categorical (NewExist, LowDoc, RevLineCr, Default, Franchise) menjadi numeric, 0 dan 1 secara manual



Handle class imbalance

- Dataset tidak dilakukan handling karena imbalance tidak terlalu ekstrem



PREPROCESSING FEATURE ENGINEERING



Feature selection

- Membuang feature-feature yang tidak relevan
- Membuang feature-feature yang redundant



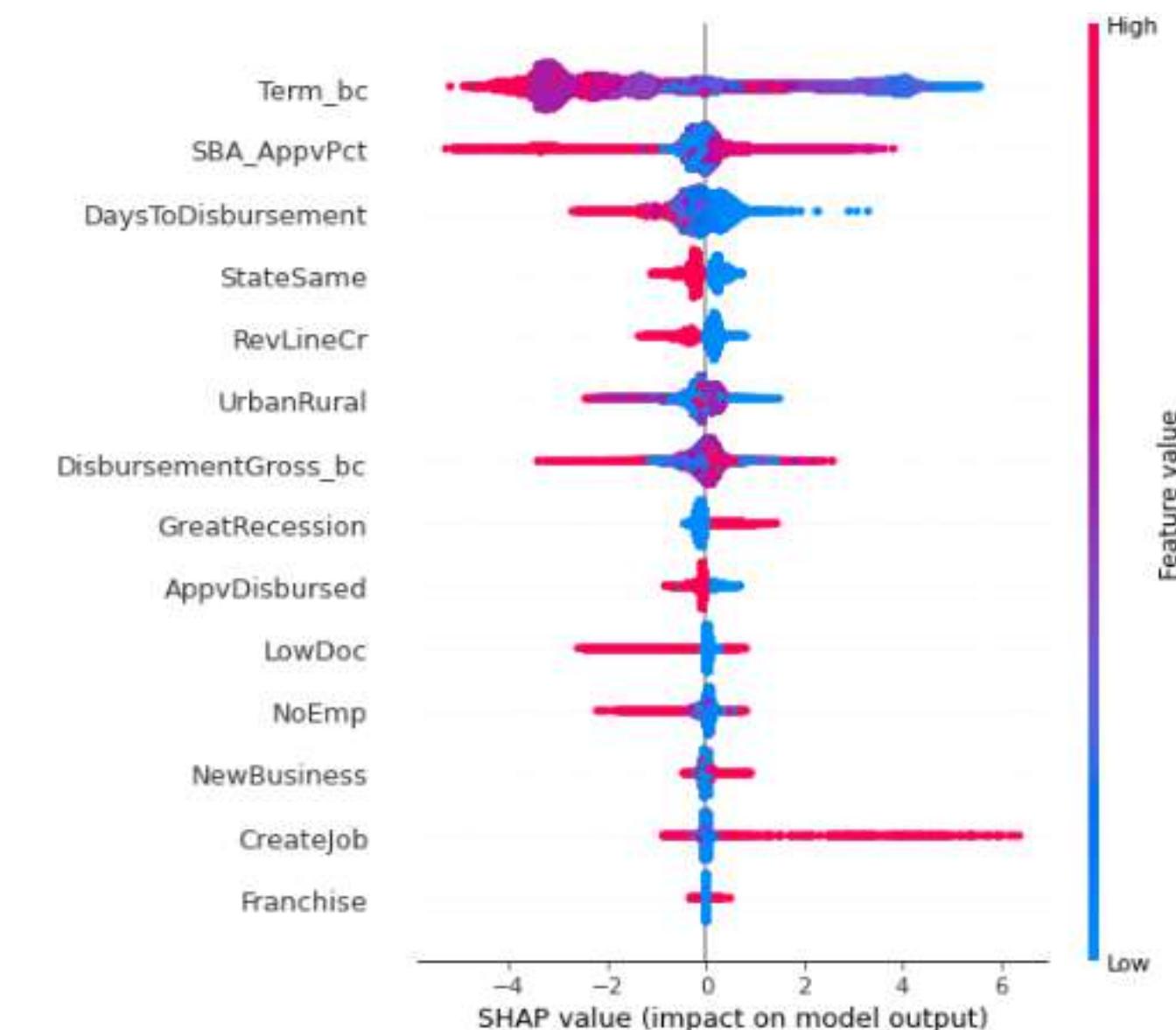
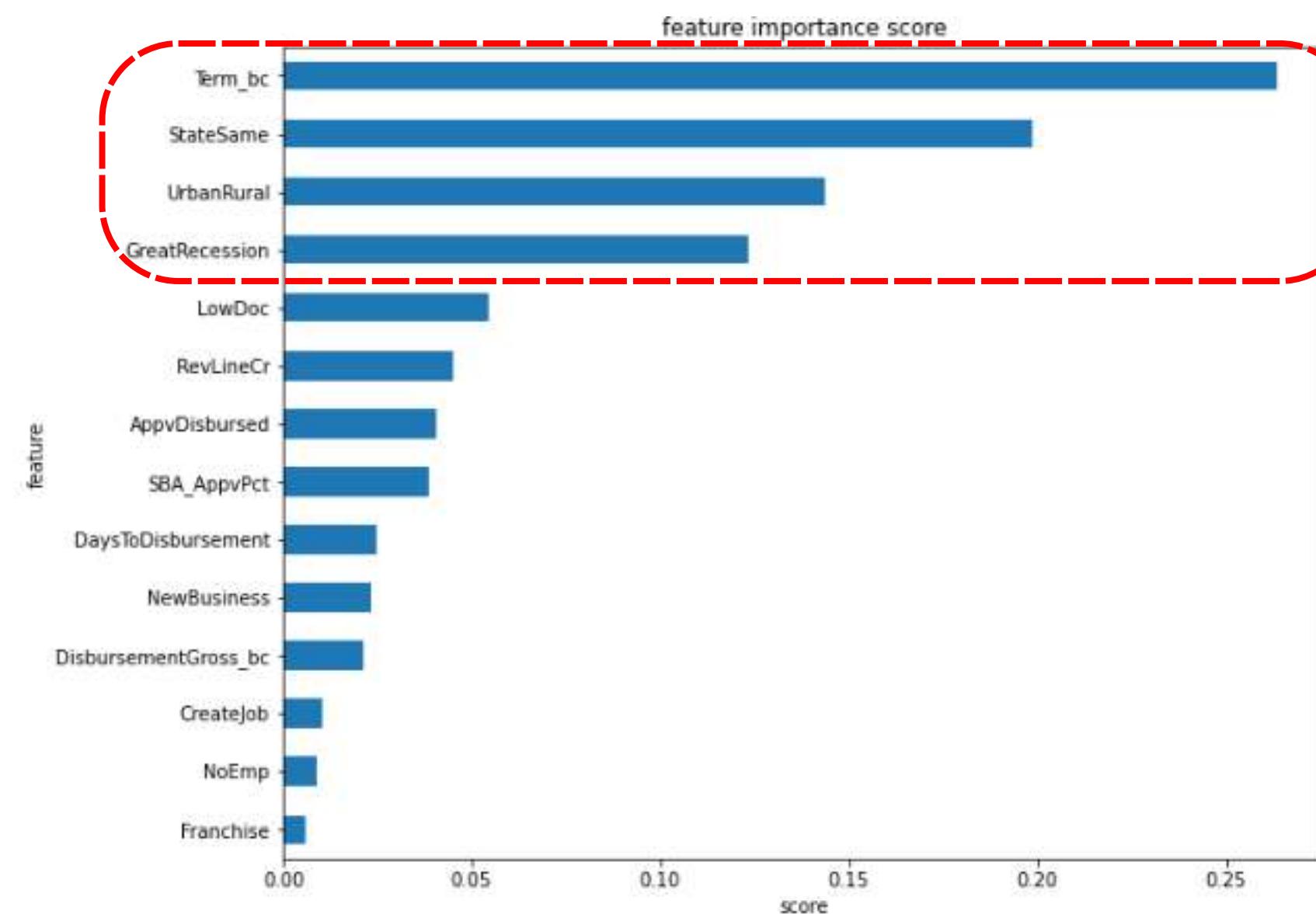
Feature extraction

- Membuat feature baru dari feature yang sudah ada seperti feature GreatRecession. SBA_AppvPct
- Mengambil saripati dari feature yang ada

| Algorithm | Evaluation Model | | | | | |
|---------------------|------------------|-------------|-------------|-------------|-------------|-------------------|
| | Accuracy | Precision | Recall | F-1 Score | AUC | Recall (Crossval) |
| Decision Tree | 0.92 | 0.79 | 0.74 | 0.76 | 0.86 | 0.70 |
| Random Forest | 0.93 | 0.82 | 0.75 | 0.78 | 0.95 | 0.71 |
| Logistic Regression | 0.85 | 0.67 | 0.30 | 0.41 | 0.84 | 0.29 |
| XGBoost | 0.94 | 0.87 | 0.81 | 0.84 | 0.97 | 0.76 |
| Adaboost | 0.90 | 0.76 | 0.67 | 0.71 | 0.94 | 0.66 |
| Gradient Boost | 0.93 | 0.84 | 0.73 | 0.78 | 0.95 | 0.70 |

- Metrics evaluasi yang digunakan adalah **Recall**.
- **XG Boost** memiliki algoritma yang **paling baik** dengan gap antara AUC train dan test yang sangat kecil dengan hasil cross validation yang paling tinggi dari yang lainnya.

TOP Feature Importance From XG Boost



| | | Actual label | |
|-----------------|---------|--------------|--------------|
| | | Default | PIF |
| Predicted label | Default | 26803 TP | 6219 FP |
| | PIF | 4096 FN | 145768 TN |

- True Positives (**TP**): Jumlah debitur yang diprediksi dengan benar sebagai default.
(default aktual, prediksi default).
- True Negatives (**TN**): Jumlah debitur yang diprediksi dengan benar sebagai no default.
(no default aktual, prediksi no default).
- Positif Palsu (**FP**): Jumlah debitur yang diprediksi salah sebagai default.
(no default aktual, prediksi default).
- False Negatives (**FN**): Jumlah debitur yang diprediksi salah sebagai no default.
(default aktual, prediksi no default).

Recall 81% = Prediksi benar default 81%

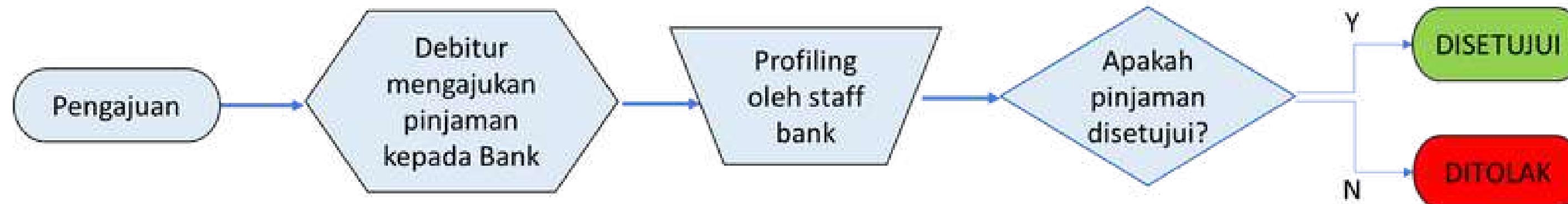
FN = Prediksi bisa bayar ternyata gagal bayar

Terdapat 4096 kasus default dari seluruh data test (182886) yg salah prediksi gagal bayar.

Default setelah menggunakan model 2,2 %

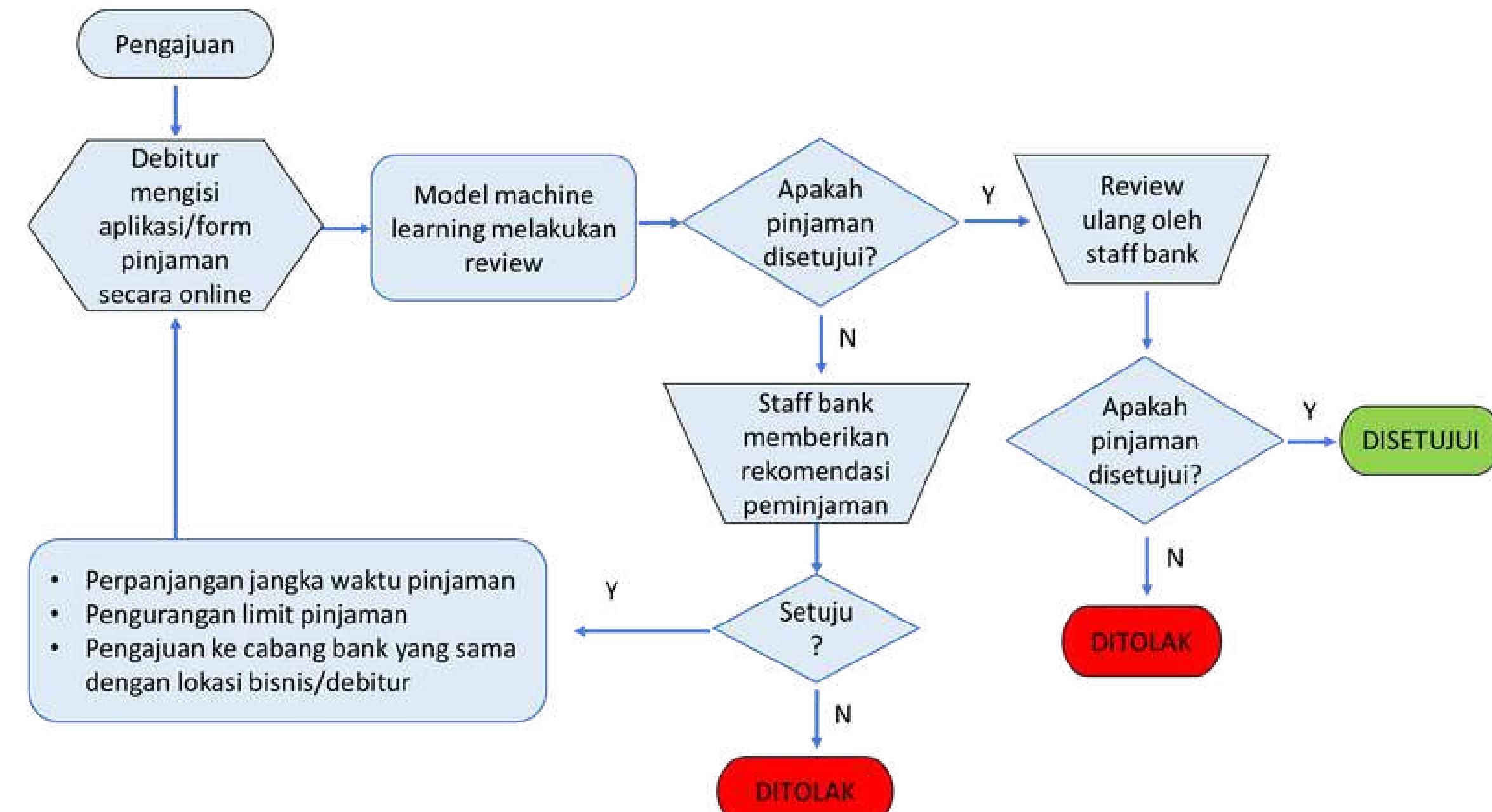


Sebelum penerapan modelling



Setelah penerapan modelling

- Pengajuan pinjaman dapat dilakukan secara online untuk memudahkan bank dalam efisiensi waktu
- Profiling oleh model machine learning dilakukan di awal secara otomatis sehingga akan banyak pengajuan yang di proses dengan cepat
- Pemberian rekomendasi kepada debitur sehingga tidak kehilangan debitur
- Pemberian pelatihan kepada usaha dengan kategori mikro.



NEW REGULATORY ADVICE FOR COMPANIES

IMPACT ML MODELS

| Category of Prospective Borrowers | Loan Amount | Term Period | Interest rate (per year) | Fees | SBA percentage | Collateral |
|-----------------------------------|----------------------|-------------|--------------------------|------|----------------|--------------------|
| Approved | \$ 25,000 or less | < 5 years | 4.75% | 2% | 75% (maximum) | 25% of loan amount |
| | \$25,000 to \$50,000 | < 15 years | 3.75% | 3% | 75% (maximum) | 30% of loan amount |
| | > \$50,000 | < 25 years | 2.75% | 4% | 75% (maximum) | 35% of loan amount |

Semakin besar jumlah pinjaman yang di setujui maka jaminan yang diberikan meningkat. Semakin besar jumlah pinjaman yang di setujui maka ditawarkan untuk memperpanjang periode pembayaran.

PROJECT EVALUATION

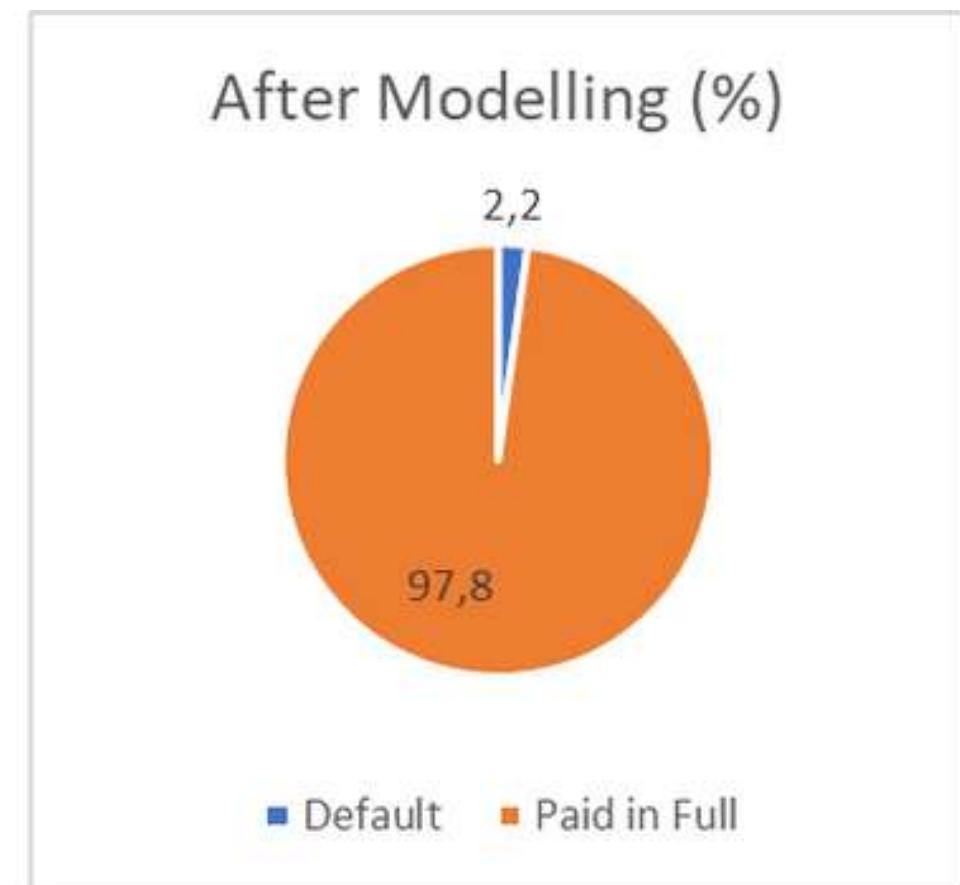
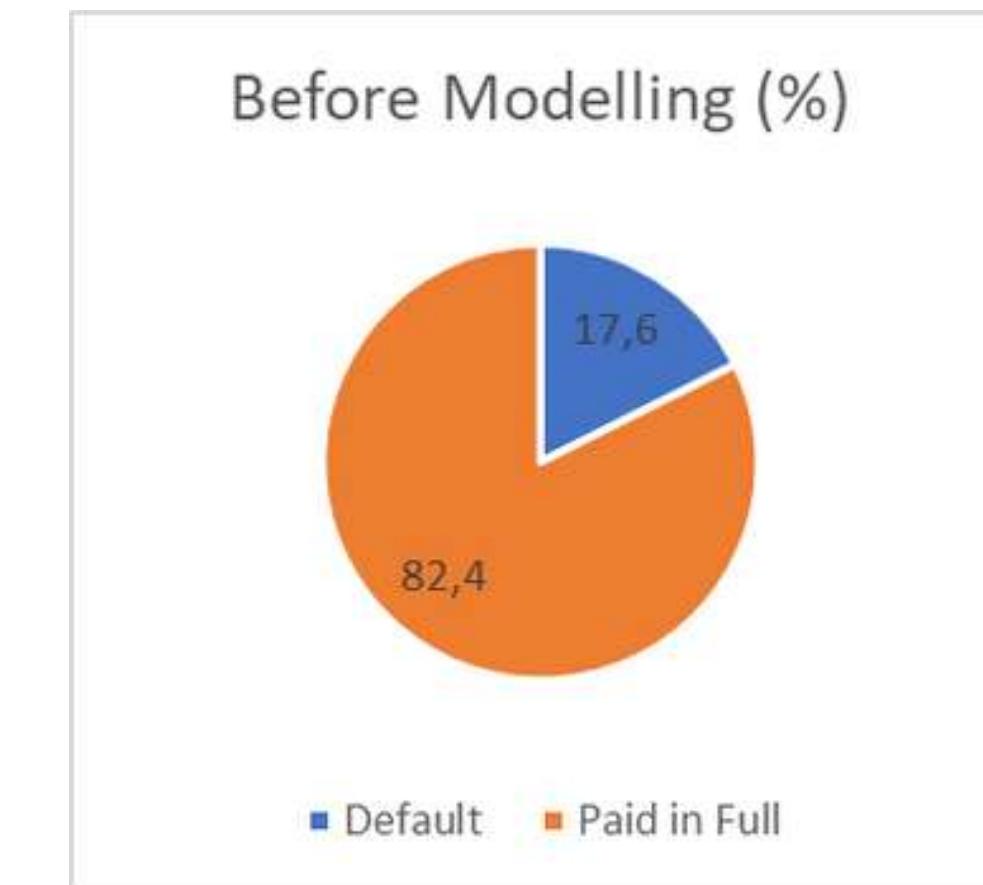
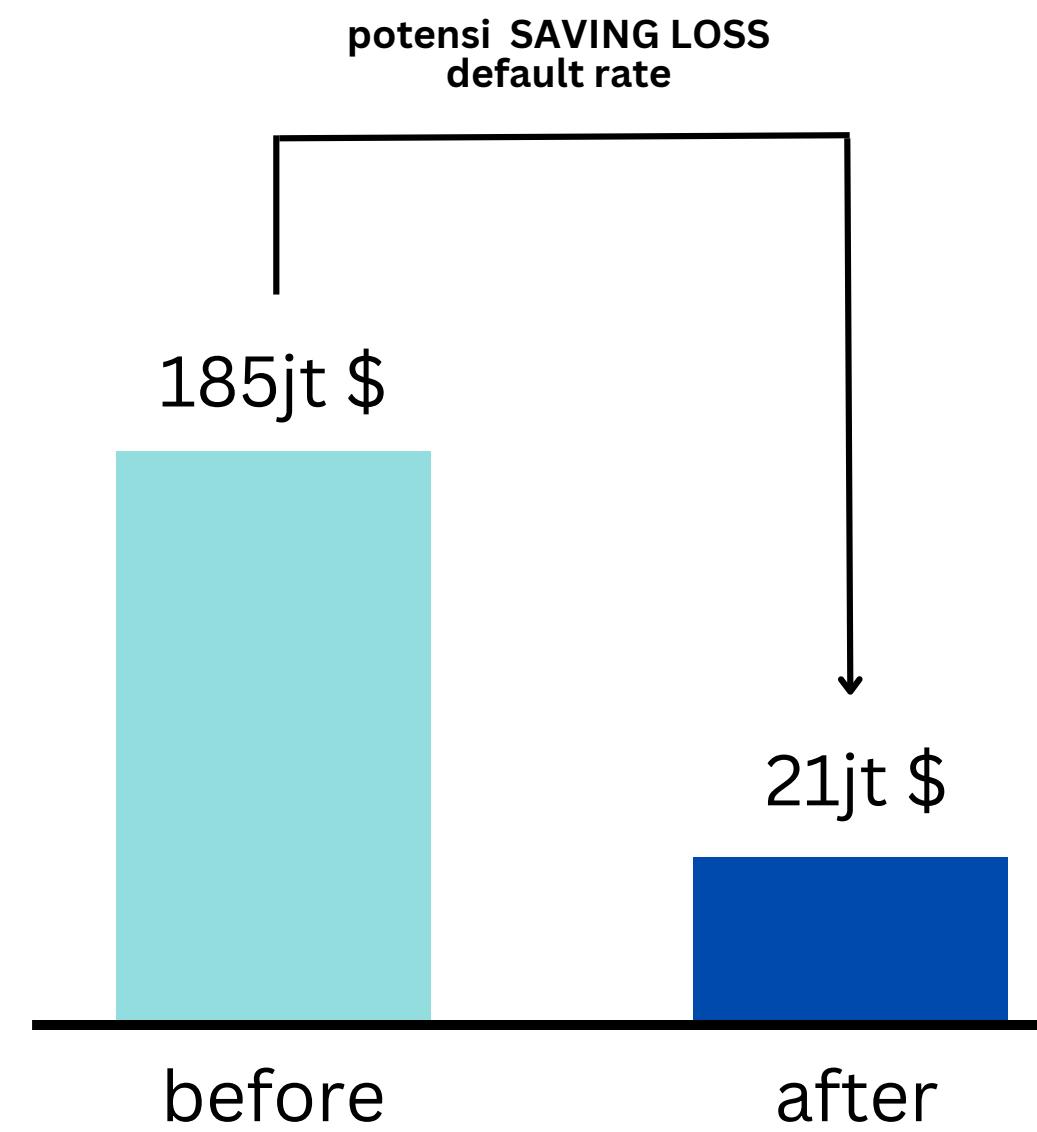
IMPACT ML MODELS

| | Default Borrowers | Disbursement Gross | Total Disbursement | Collateral | Loan covered by SBA | Loss | Interest Rate | Fees | Profit from potential loss |
|---------------|--------------------------|---------------------------|---------------------------|-------------------|----------------------------|-------------|----------------------|-------------|-----------------------------------|
| Before | 34.838 | 106.228 | 3.700.771.064 | 740.154.213 | 2.775.578.298 | 185.038.553 | 92.519.277 | 148.030.843 | 1,5% |
| After | 4.096 | 106.228 | 435.109.888 | 87.021.978 | 326.332.416 | 21.755.494 | 10.877.747 | 17.404.396 | 16,4% |

Peningkatan Collateral dari 20% menjadi 35% pada debitur akan meningkatkan profit dari potential loss sebesar **14,9%**

MODELING APPLICATION INSIGHT

SUMMARY



Penurunan default dari 17,6 % menjadi 2,2 %

▼ 88%

164jt USD *Years : 1988 - 2010

Potensi penurunan kerugian
dari tingkat gagal bayar



THANK YOU!



Jupyteron Group



<https://www.sba.gov/>



1-800-827-5722

U.S. Small Business Administration (SBA)
409 3rd St SW, Washington, DC 20416, Amerika Serikat