FINAL PROJECT

SBA Loan Approval

D-Alchemist

D-Alchemist

OUR DATA SCIENCE TEAM

Mentor

Kevin

Lead	Member	Member	Member	Member
				
Omega Delima Munthe	**Ilham Muhammad Shuhada**	**Rafael Nicholas Tanaja**	**Johnny Lim**	**Ageng Pamungkas**

TABLE OF CONTENT

D-Alchemist

Rakamin Academy

BUSINESS UNDERSTANDING

D-Alchemist

SBA?

SBA (Small Business Administration) adalah lembaga pemerintah Amerika Serikat yang memberikan dukungan finansial kepada small business (nasabah) untuk bisa berkembang, seperti memberikan jaminan pinjaman.

Problems

Tingginya gagal bayar pinjaman mencapai 32.3% dengan total nominal yang gagal dibayar mencapai $5.48 miliar.

D-Alchemist adalah Data Scientist di **SBA**

Objective

Membuat model yang akan memprediksi calon nasabah akan membayar lunas atau gagal membayar pinjaman.

Goals

- Menurunkan persentase gagal bayar
- Menurunkan total nominal yang dinyatakan gagal bayar

Business Metric

- **Default Percentage** (Persentase gagal bayar mengacu pada kolom MIS_Status).
- **Charged-off Total** (total nominal yang dinyatakan gagal bayar, mengacu pada kolom ChgOffPrinGr).

D-Alchemist

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis

Contohnya seperti UrbanRural, NewExist, SBA_Appv, etc.

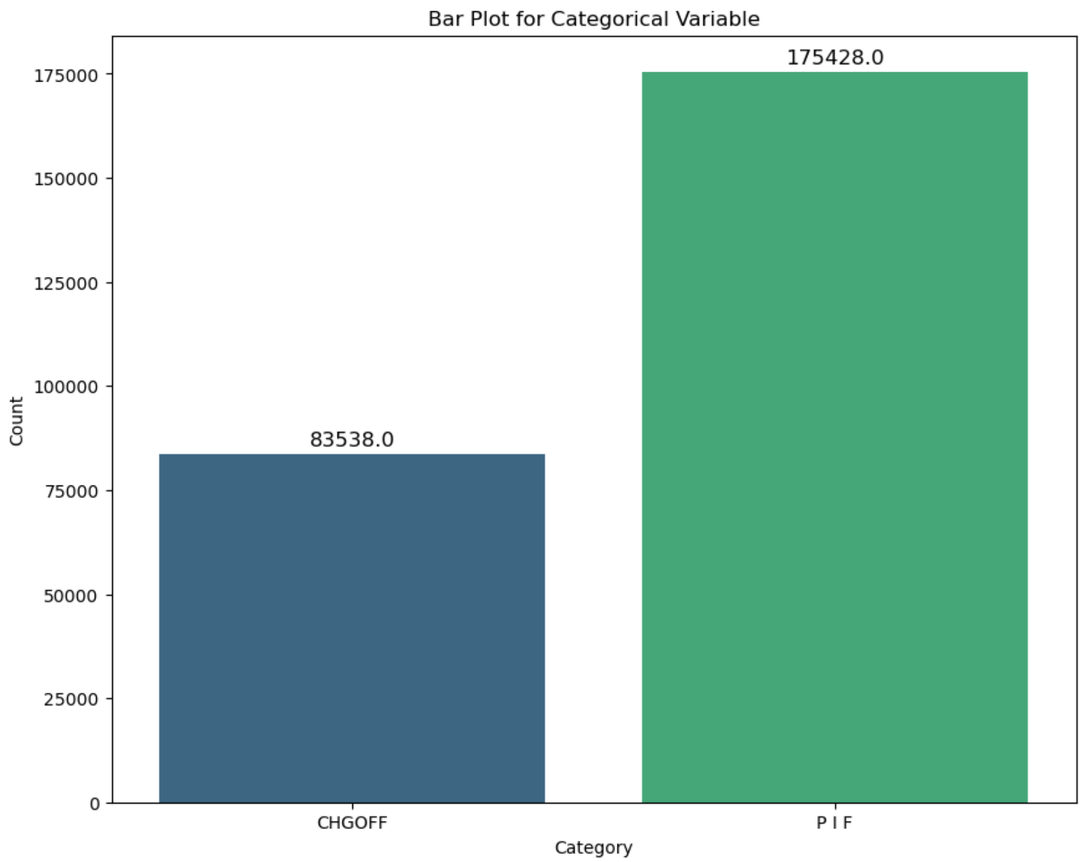
Contoh Name, City, State, Bank, BankState, RevLineCr, LowDoc, ChgOffDate, DisbursementDate.

Beberapa fitur memiliki positif skewed seperti ChgOffPrinGr, BalanceGross, CreateJob.

Terdapat banyak outliers pada kolom numeric seperti ChgoffPrinGr, CreateJob, GrAppv, etc.

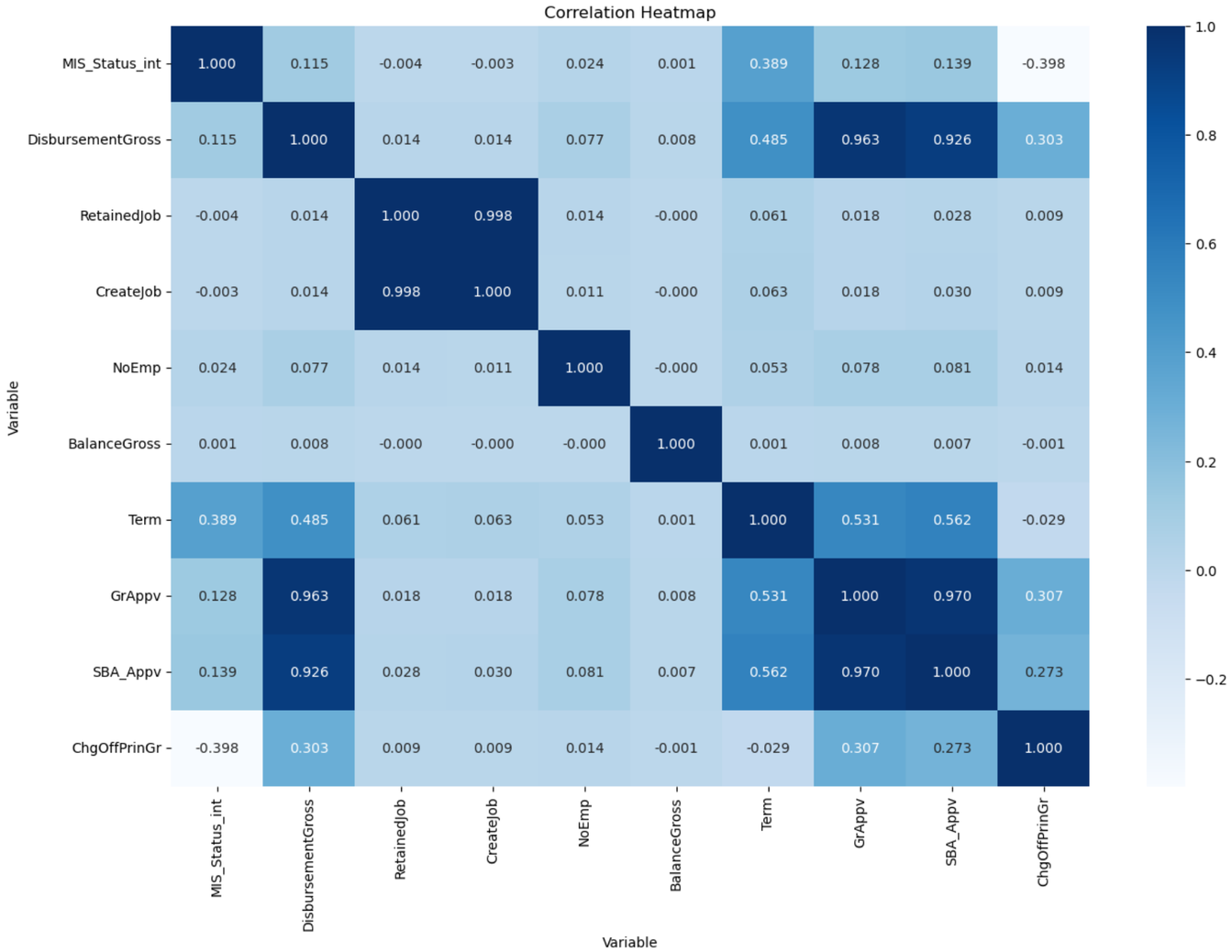
Kolom "Zip", “RevLineCr”, dan “LowDoc” memiliki beberapa data dengan karakter yang tidak bisa diinterpretasikan dengan jelas.

Univariate Analysis



Kolom target yaitu **MIS_Status memiliki 2 kategori yaitu CHGOFF (gagal bayar) dan PIF (lunas) dengan proporsi yang tidak seimbang.**

Multivariate Analysis



- Fitur **Retained Job** (jumlah pekerjaan yang ada) **memiliki korelasi yang tinggi dengan Create Job** (jumlah lapangan kerja baru).

- Fitur **GrAppv** (jumlah pinjaman kotor dsetujui) juga **memiliki korelasi yang tinggi dengan Disbursement Gross** (jumlah pinjaman cair) **dan SBA_Appv** (jumlah pinjaman disetujui SBA)

D-Alchemist

Rakamin Academy

DATA PRE-PROCESSING

Data Cleaning

Handling Missing Value

Mengisi missing values (***0.04 %*** *dari total data*) dengan modus

Handling Outlier

Menggunakan Z-Score untuk mendeteksi dan menghapus outliers.
Data yang terhapus: **2,76%**

Handling Duplicated Data

Tidak ada baris yang duplicate.

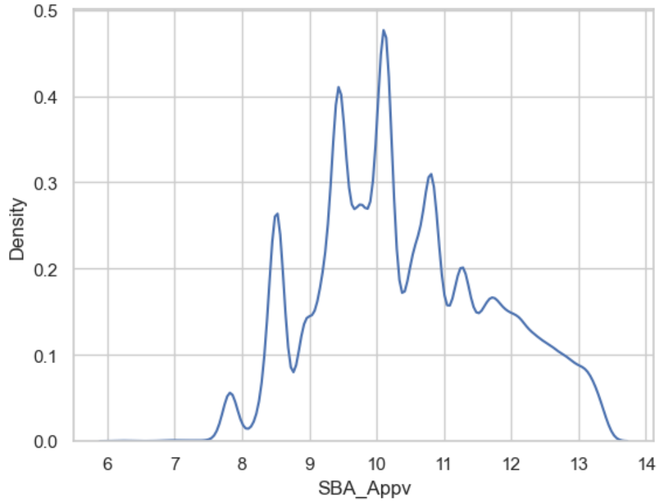
Handling Invalid Data

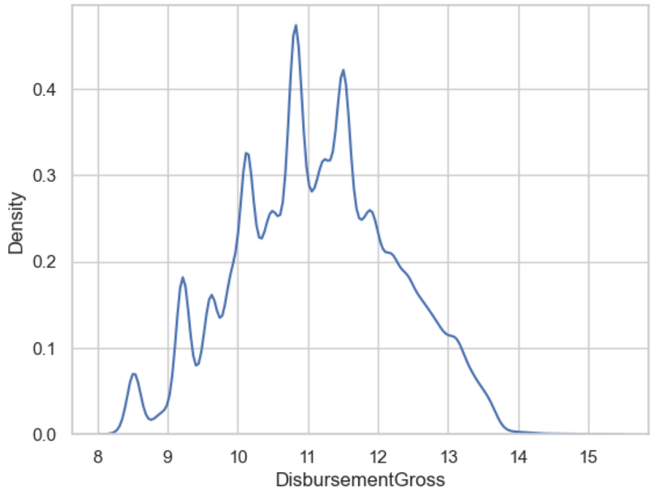
Mengubah 0 dan T menjadi No, 1 menjadi Yes, dan menghapus invalid data lainnya yang tidak dapat diinterpretasi

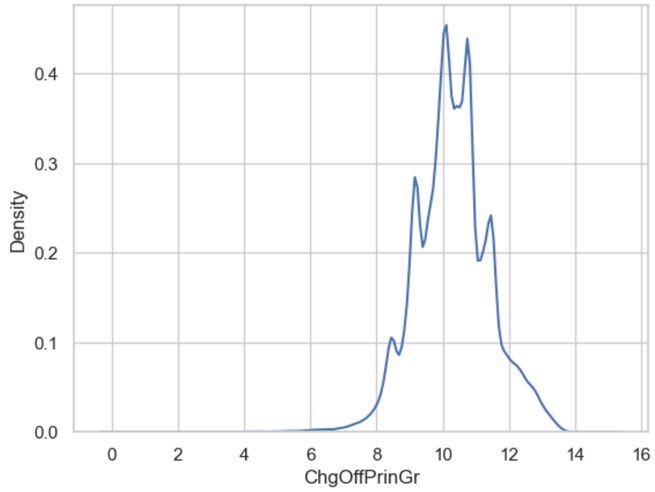
Feature Transformation

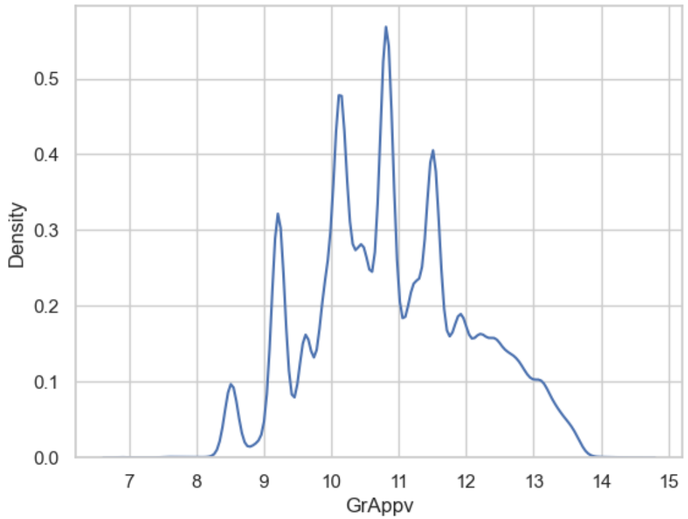
Semua fitur yang memiliki tipe data numeric melakukan Feature Transformation **menggunakan log** hasilnya **hanya beberapa fitur** yang mendekati **distribusi normal** setelah dilakukan Feature Transformation :

- SBA_Appv
- DisbursementGross
- ChgOffPrinGr
- GrAppv









Feature Engineering
(Extraction)

Recession

Kategori pinjaman dilakukan pada tahun resesi atau tidak berdasarkan Feature DisbursementDate

term_category

Pengelompokkan jangka waktu pinjaman (term)

Job Stability

Kategori 0 dan 1 untuk mengindikasikan sejauh mana pekerjaan dipertahankan (RetainedJob) > pekerjaan yang dibuat (CreateJob)

CompanySize

Pengelompokkan ukuran perusahaan berdasarkan jumlah karyawan

Industry

Mendefinisikan industry peminjam berdasarkan feature NAICS

Feature Encoding

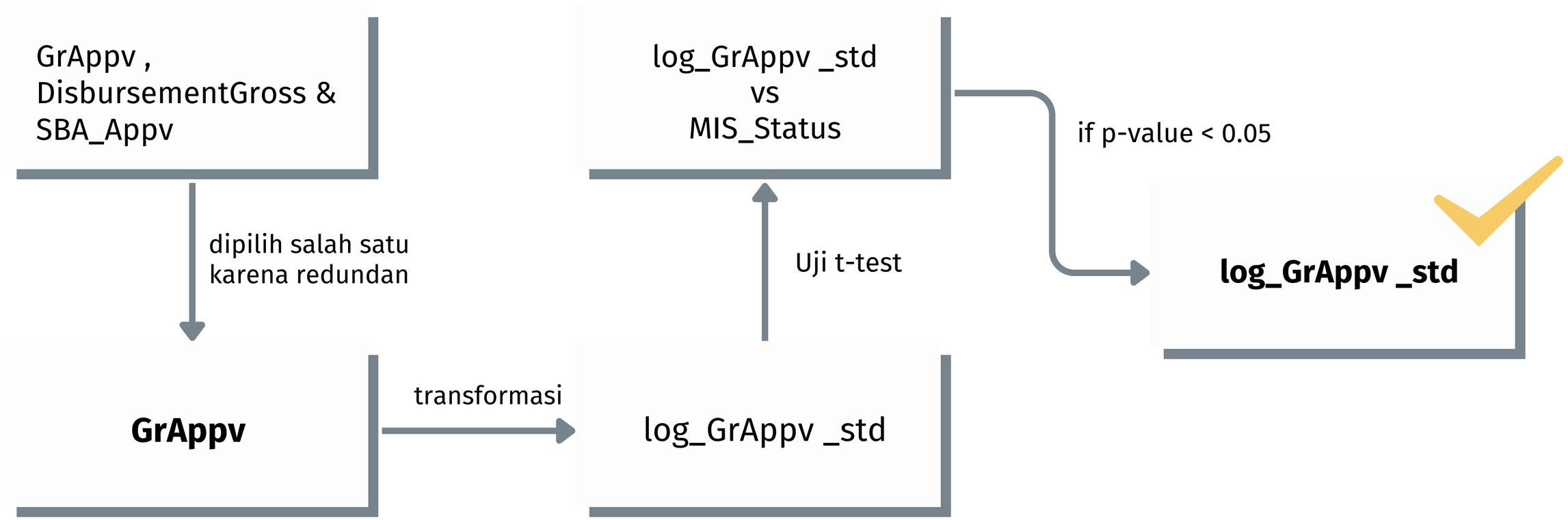
Menggunakan **Label Encoding** untuk mengubah Feature Categorical seperti FranchiseCode, term_category, RevLineCr, LowDoc, Industry, NewExist menjadi numeric agar dapat menjadi input untuk model machine learning.

Feature Engineering (Scaling)

Menerapkan standardisasi terhadap kolom numerik yang telah dilakukan log transformation.

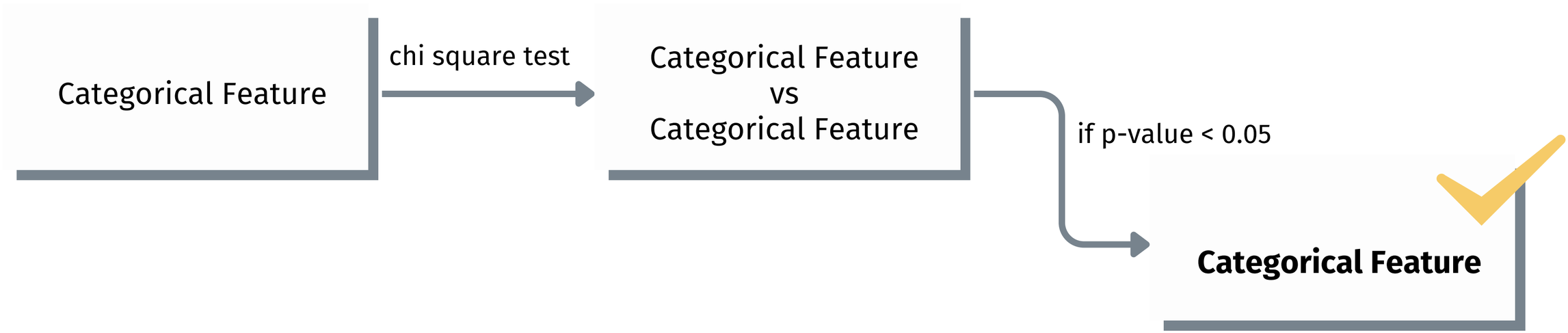
Feature Engineering

(Numerical Feature Selection)



Feature Engineering

(Categorical Feature Selection)



Feature Engineering
(Selection)

Fitur yang diambil :

- **NewExist_encoded**: Pengelompokkan perusahaan menjadi bisnis baru atau bisnis yang sudah lama didirikan
- **RevLineCr**: Revolving Line of Credit
- **LowDoc**: Low Documentation
- **Industry**: Kategori sektor bisnis perusahaan
- **Job Stability**: Indikator sejauh mana pekerjaan yang dipertahankan lebih besar daripada pekerjaan yang dibuat
- **CompanySize**: Kategori ukuran perusahaan berdasarkan jumlah karyawan
- **Franchise**: Kategori perusahaan termasuk ke bisnis franchise atau tidak
- **Recession**: Pinjaman diberikan saat resesi atau tidak
- **Term Category**: Kategori jangka waktu pinjaman
- **GrAppv**: Nominal pinjaman ke bank
- **MIS Status**: Kolom target yang menyatakan lunas atau gagal bayar

32 Fitur → **11 Fitur**

Class Imbalance

Menggunakan **Oversampling SMOTE** untuk **handle class imbalance** menjadi perbandingan 50:50

D-Alchemist

Rakamin Academy

MODELING & EVALUATION

Model Evaluation Metric

Precision as Primary Metric

Memilih **Precision** sebagai primary metric evaluation sesuai konteks bisnis dari dataset : lebih baik fokus untuk mereduksi False Positive (nasabah yang diprediksi akan membayar pinjaman secara lunas, namun kenyataannya gagal bayar).

Accuracy as Secondary Metric

Accuracy sebagai metrik sekunder dapat **memberikan pemahaman keseluruhan** tentang seberapa baik model bekerja pada seluruh dataset. **Accuracy** memberikan gambaran umum tentang sejauh mana model benar-benar memprediksi dengan benar, baik True Positive maupun True Negative.

Modeling Result
(Before Hypertuning)

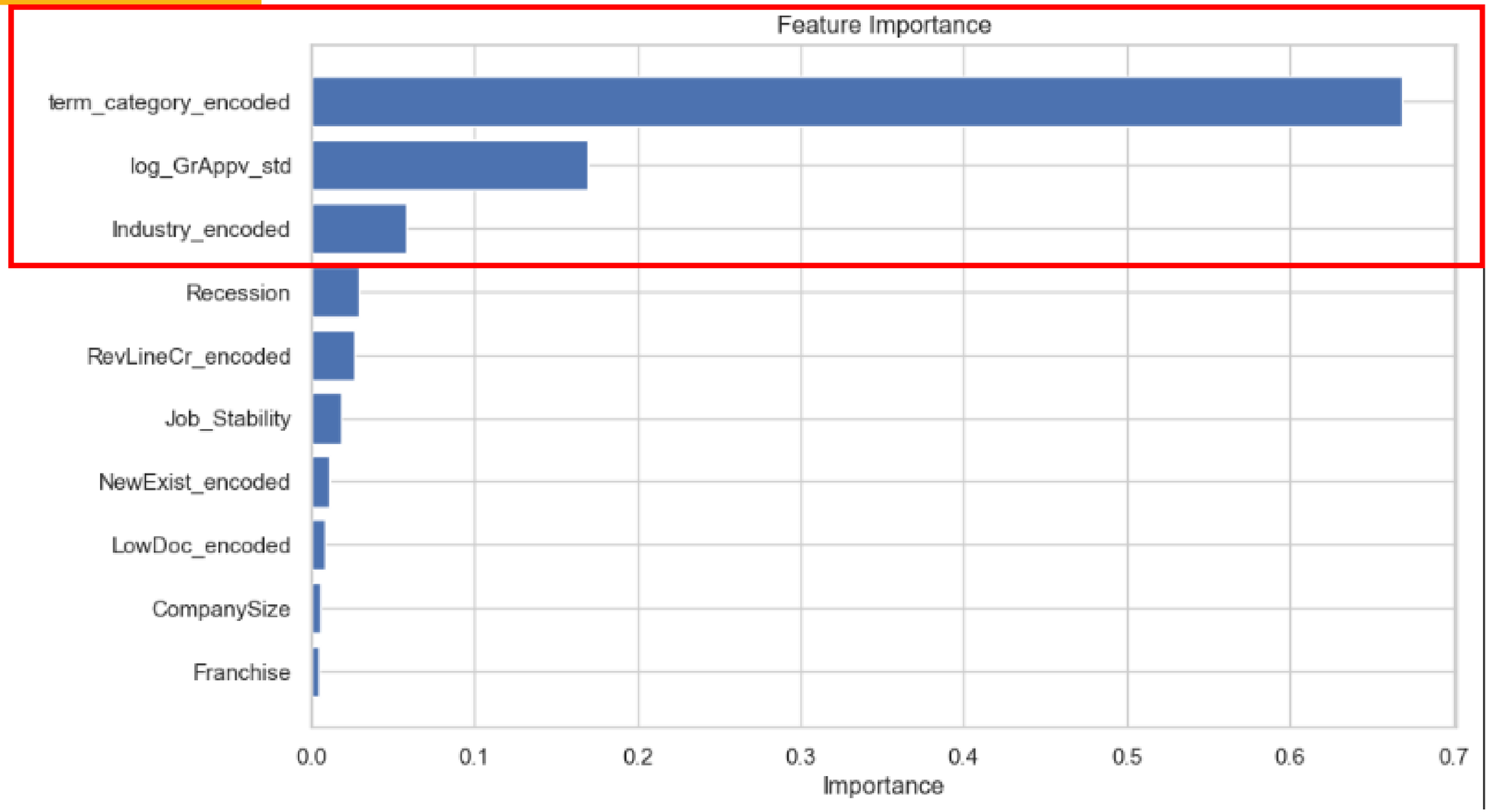
Model	Train Precision	Test Precision	Train Accuracy	Test Accuracy
Logistic Regression	0.79	0.88	0.80	0.80
K-Nearest Neighbor	0.87	0.90	0.88	0.85
Decision Tree	0.94	0.91	0.93	0.83

Modeling Result
(After Hypertuning)

Model	Train Precision	Test Precision	Train Accuracy	Test Accuracy
Logistic Regression	0.79	0.88	0.80	0.80
K-Nearest Neighbor	0.87	0.91	0.88	0.86
Decision Tree	**0.93**	**0.91**	**0.91**	**0.83**

Model yang kami pilih adalah **Decision Tree** yang sudah dilakukan tuning hyperparameter karena model tersebut memiliki score Data Train dan Data Test (baik itu Precision maupun Accuracy) yang lebih tinggi dibandingkan model lainnya, dengan gap antara data score Data Train dan Data Test yang rendah

Feature Importance

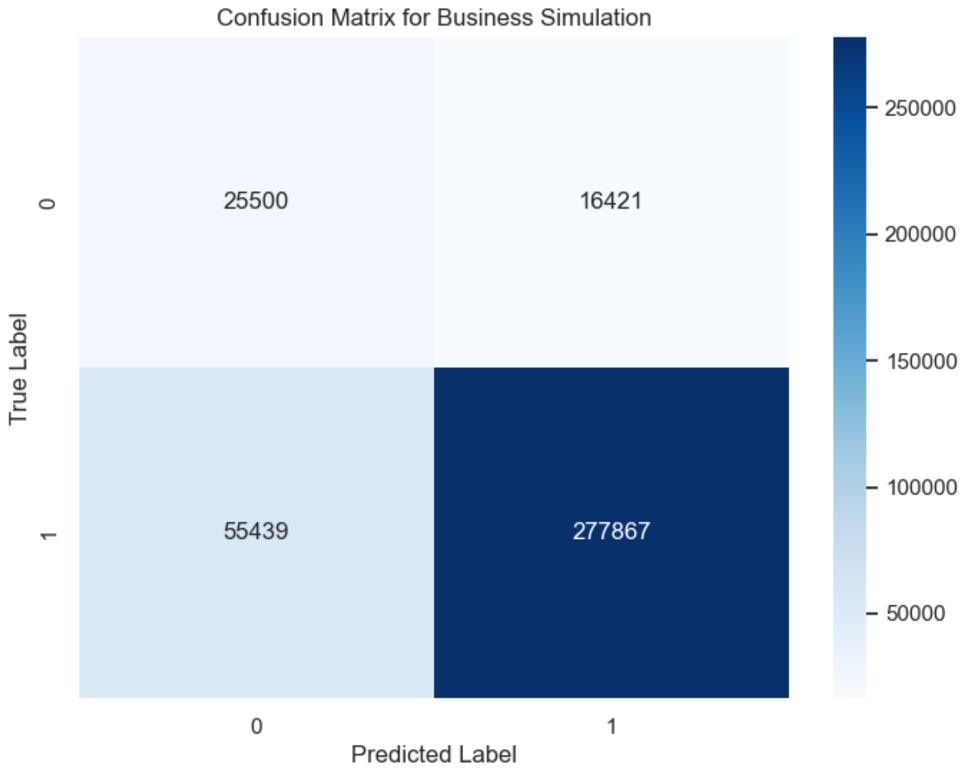


D-Alchemist

BUSINESS SIMULATION

Rakamin Academy

Confusion Matrix



Precision

94%

Accuracy

81%

277867 **True Positive** : diprediksi berhasil bayar, dan itu benar

25500 **True Negative** : diprediksi gagal bayar, dan itu benar

16421 **False Positive** : diprediksi berhasil bayar, dan itu salah

55439 **False Negative** : diprediksi gagal bayar, dan itu salah

Default Percentage

11,17%

Before Model

Decrease

5,59 %

FP / (FP + TP)

5,58%

After Model

Charged-Off Total

$ 3.38 BILLION

Before Model

Decrease

$ 1.58 BILLION

$ 1.8 BILLION

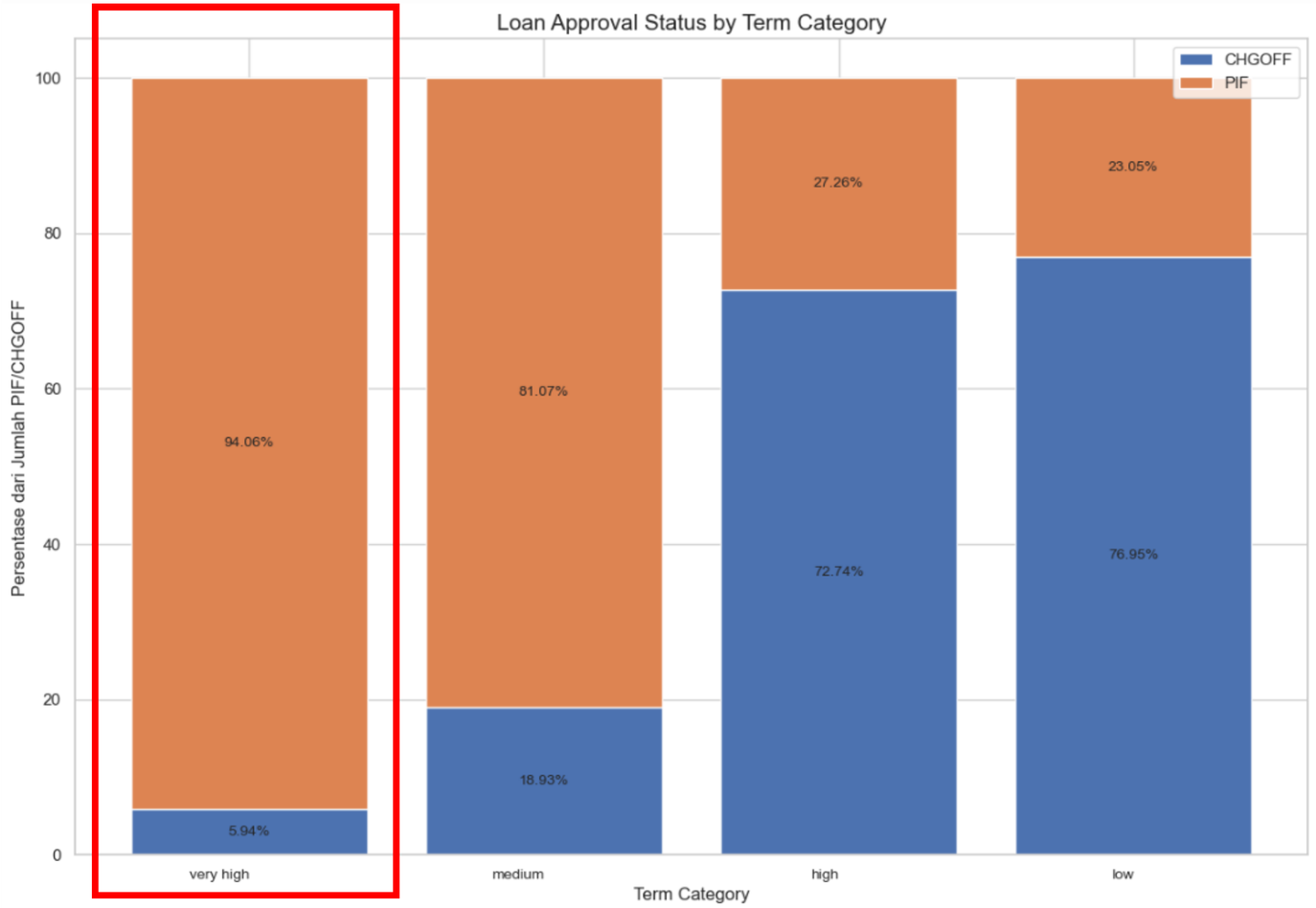
After Model

D-Alchemist

Rakamin Academy

BUSINESS RECOMMENDATION

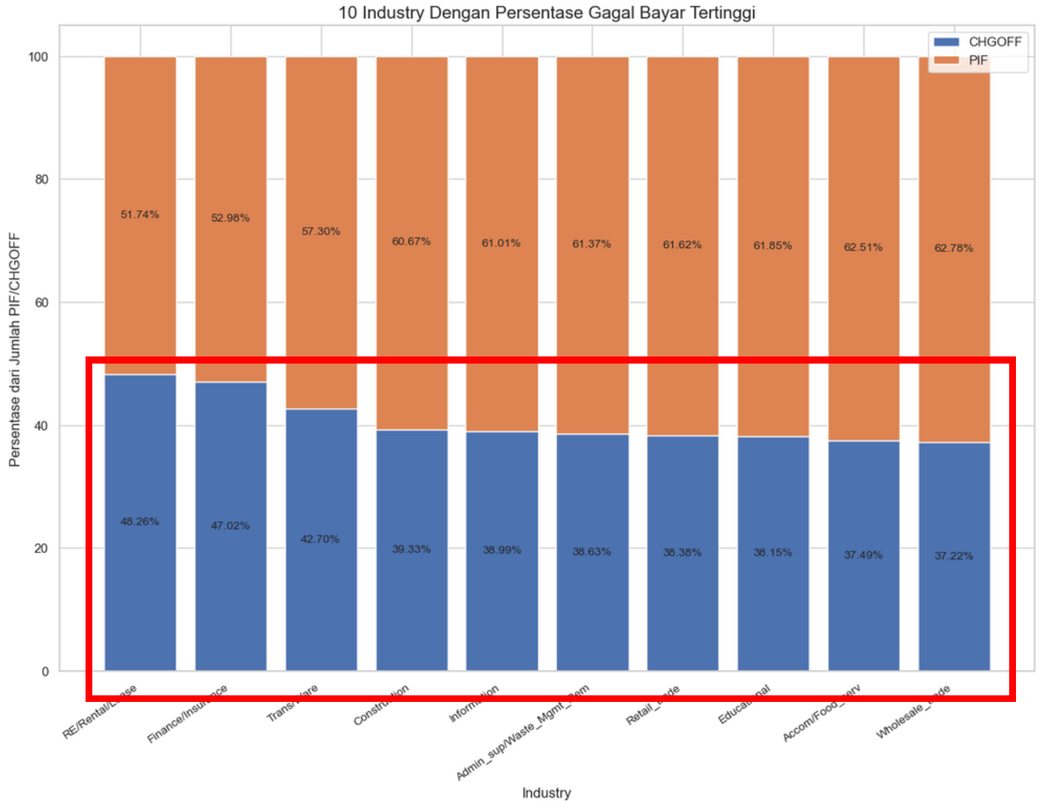
Business Recommendation - Term



Dapat dilihat bahwa term kategori **very high** (lebih dari 82 bulan) memiliki tingkat gagal bayar sangat rendah.

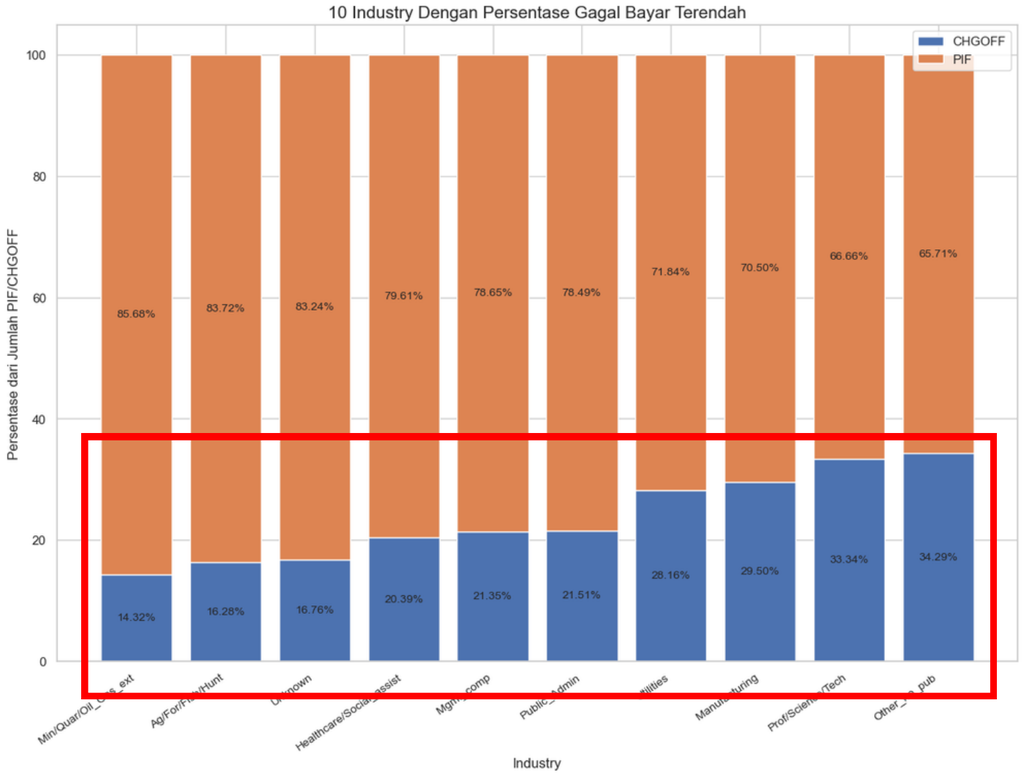
Rekomendasi yang dapat dilakukan adalah **Mengembangkan Layanan Pinjaman Jangka Panjang.** Dengan cara, Bank dapat menawarkan produk pinjaman dengan durasi yang lebih lama (lebih dari 82 bulan) dengan suku bunga yang bersaing.

Business Recommendation - Industry



Tiga Industri dengan **tingkat gagal bayar tertinggi** yaitu:

- **RE/Rental/Lease (48,26%)**
- Finance/Insurance (47,02%)
- Trans/Ware (42,7%).



Tiga Industri dengan **tingkat gagal bayar terendah** yaitu:

- **Min/Quar/Oil_Gas_ext (14,32%)**
- Ag/For/Fish/Hunt (16,28%)
- Unknown (16,76%)

Business Recommendation - Industry

Fokus **pemberian jaminan pinjaman pada perusahaan manajemen, sektor Min/Quar/Oil_Gas_ext, dan Ag/For/Fish/Hunt** karena persentase gagal bayar rendah. Dapat diberikan jalur khusus pada sektor tersebut untuk mendorong jumlah pengajuan jaminan pinjaman.

Tingkatkan manajemen risiko pada sektor dengan tingkat gagal bayar tinggi seperti **Real Estate/Rental/Lease, Finance/Insurance, Transportation/Warehousing.**

Rakamin Academy

D-Alchemist

THANK YOU