CREDIT SCORE MODEL

FAJAR ADI PRASETIO

Fajar Adi Prasetio

FajarAdi25

fajaradiprasetio@gmail.com

Fajaradi25.github.io

HOME CREDIT

Kamu Bisa!

TABLE OF CONTENT

HOME CREDIT

Kamu Bisa!

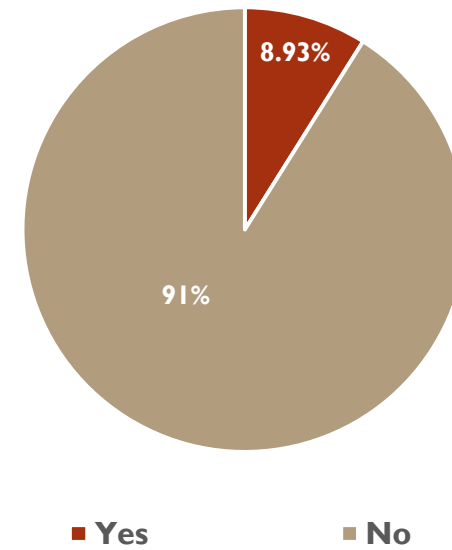# BUSINESS UNDERSTANDING

HOME CREDIT
Kamu Bisa!

Problem

8.93% of creditors have problems paying off credit applications

Goals

The Credit Score Model can help companies ensure that creditors who can make repayments are not rejected and creditors who are unable to make repayments will be rejected when applying for a loan
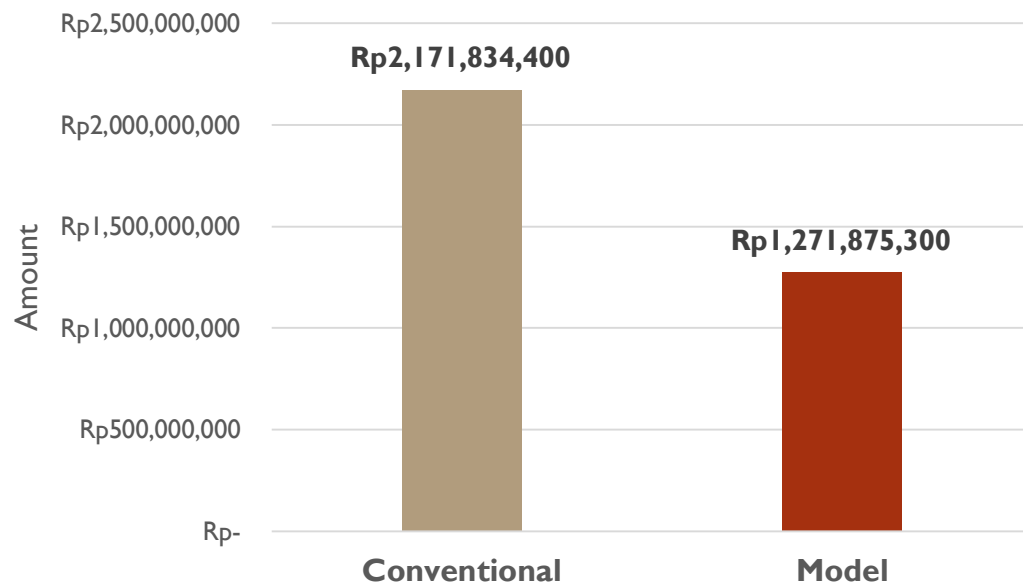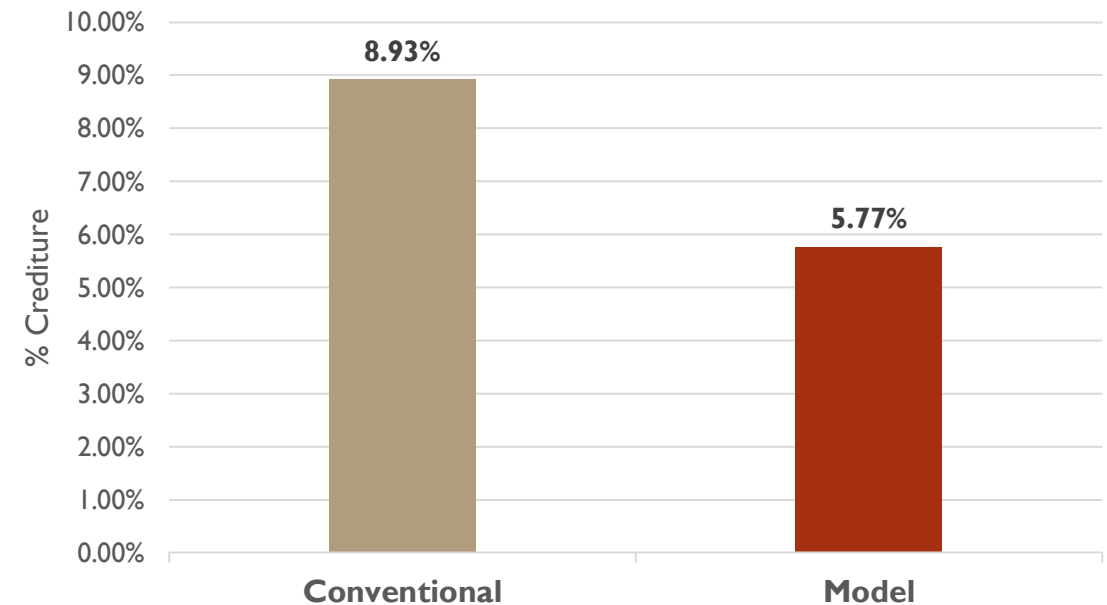
Creditors Issues

8.93%

91%

■ Yes   ■ No

BUSINESS UNDERSTANDING

Impact Model to Business Metric

Conventional: Rp2,171,834,400
Model: Rp1,271,875,300

Impact Model to Business Metric

Conventional: 8.93%
Model: 5.77%

Impact

The model can reduce number of defaults by 3.16% and amount of credit by Rp. 899,959,100

DATA PREPROCESSING

Dataset Info

- 33 Columns & 223008 Rows

- 7 Categorical Data & 25 Numerical Data

Cleaning Data

1. Missing Value
   - Replace blank values in categorical data with the most values.
   - Replace blank values in numeric data with 0
2. Duplicate Data
   - There are no duplicate data
3. Outliers
   - Using a boxplot to detect outliers
   - Using interquartile range to determine limit of the data range

Data Transformation

Weight of Evidence (WoE) methodology

Feature Selection

Using Multicollinearity Prevention with Pearson Correlation

EXPLORATORY DATA ANALYSIS (EDA)

	VAR_NAME	MIN_VALUE	MAX_VALUE	COUNT	EVENT	EVENT_RATE	NONEVENT	NON_EVENT_RATE	DIST_EVENT	DIST_NON_EVENT	WOE	IV
19	NAME_INCOME_TYPE	Maternity leave	Maternity leave	4	2.0	0.500000	2.0	0.500000	0.000098	0.000010	2.298385	0.024147
115	MICROLOAN	2.0	79.0	1037	254.0	0.244937	783.0	0.755063	0.012481	0.003864	1.172586	0.010179
105	CREDIT_DAY_OVERDUE	2.0	5134.0	2457	445.0	0.181115	2012.0	0.818885	0.021866	0.009928	0.789575	0.009571
50	OCCUPATION_TYPE	Low-skill Laborers	Low-skill Laborers	2039	358.0	0.175576	1681.0	0.824424	0.017591	0.008295	0.751774	0.058652
27	NAME_EDUCATION_TYPE	Lower secondary	Lower secondary	2129	305.0	0.143260	1824.0	0.856740	0.014987	0.009000	0.509909	0.062998

EXPLORATORY DATA ANALYSIS (EDA)

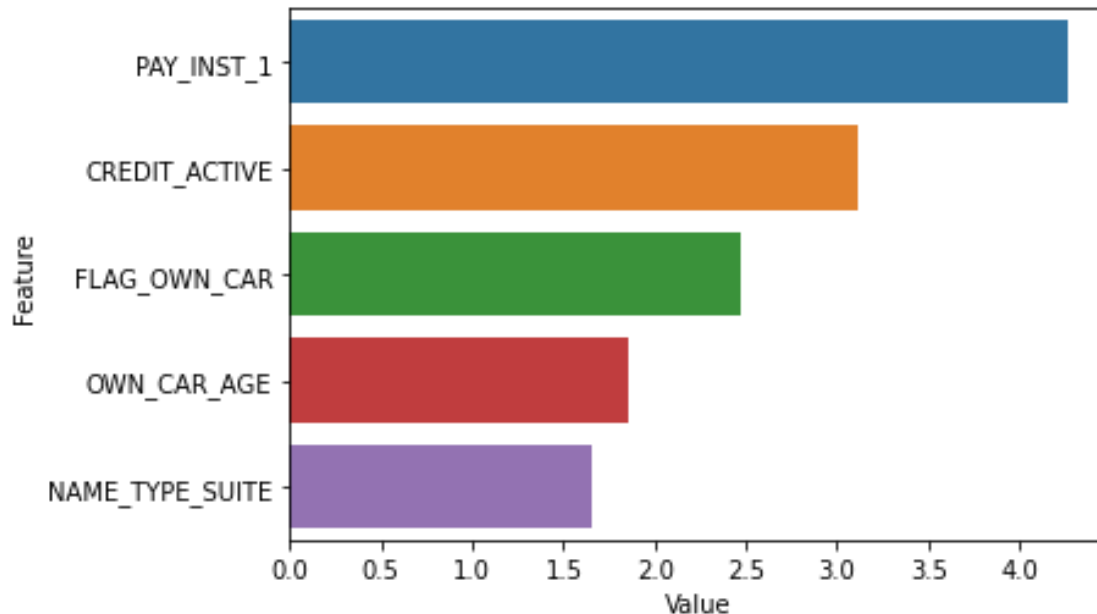OCCUPATION_TYPE vs target

NAME_EDUCATION_TYPE vs target
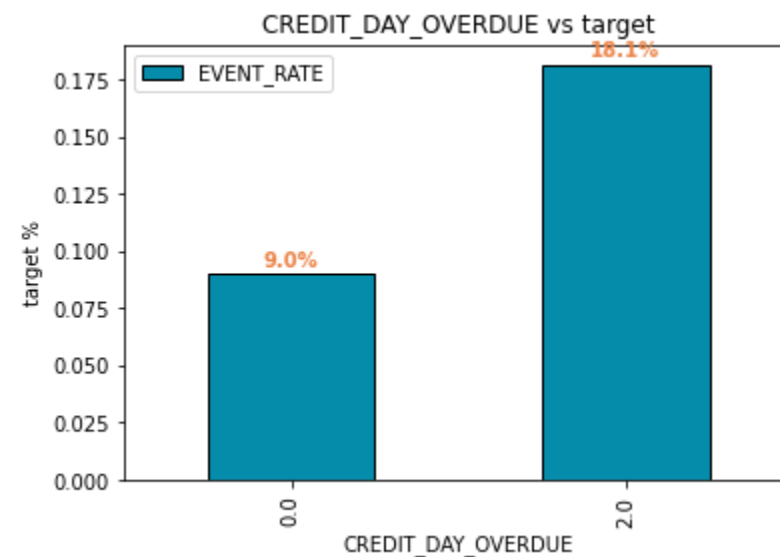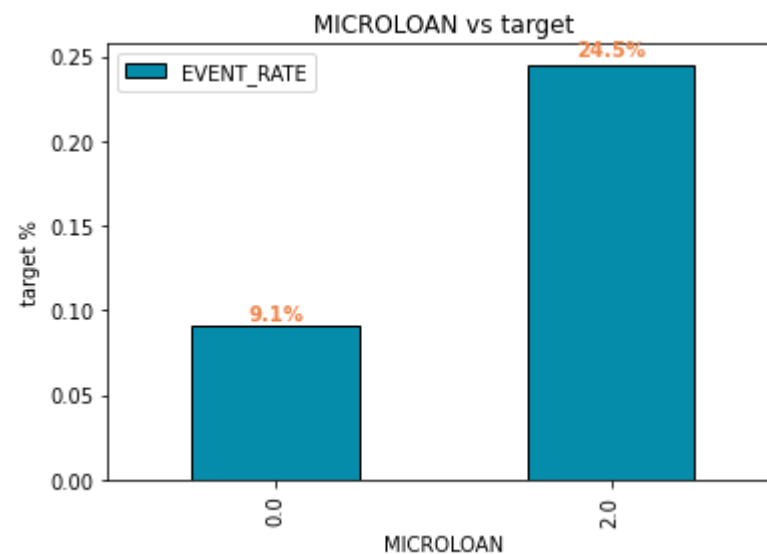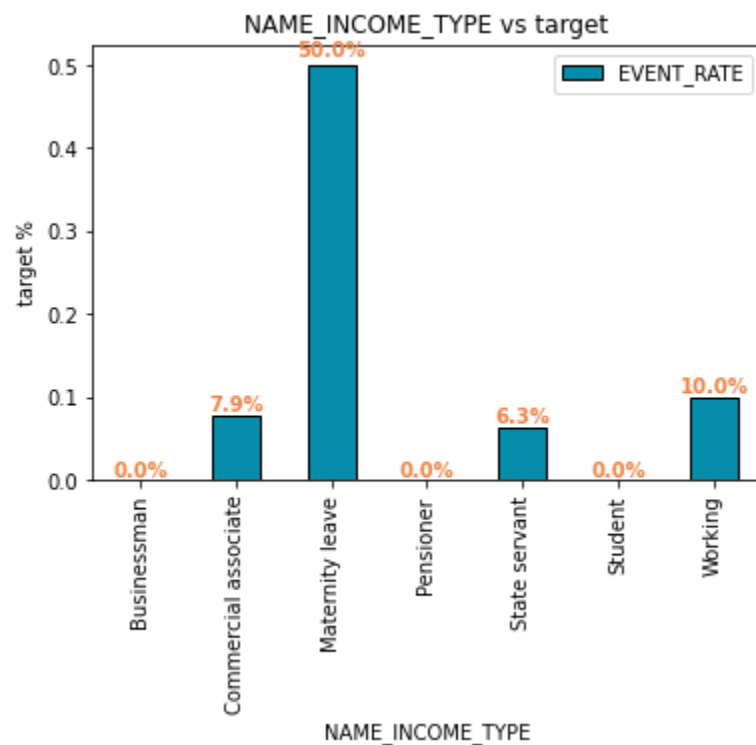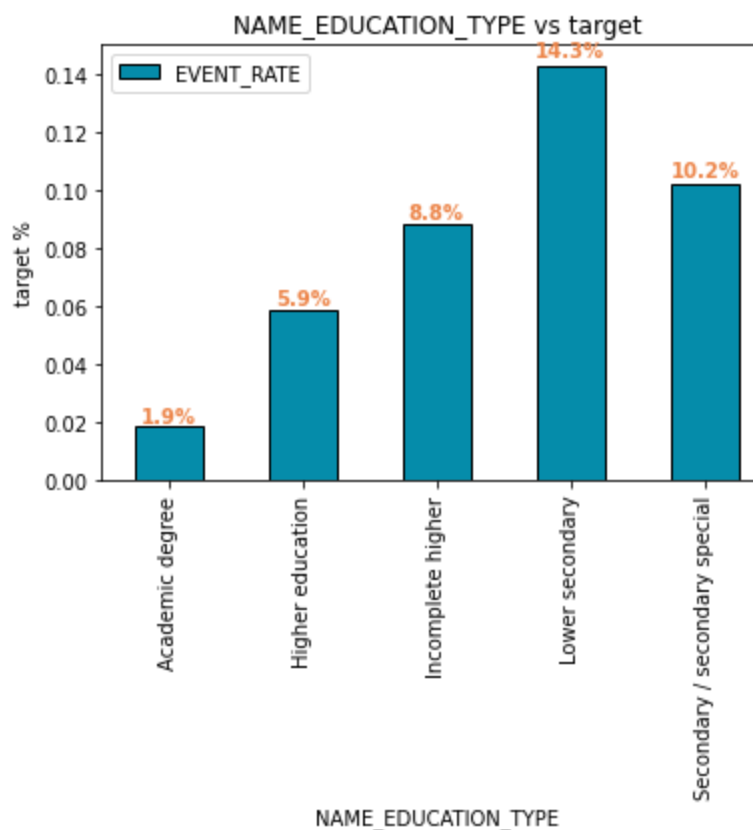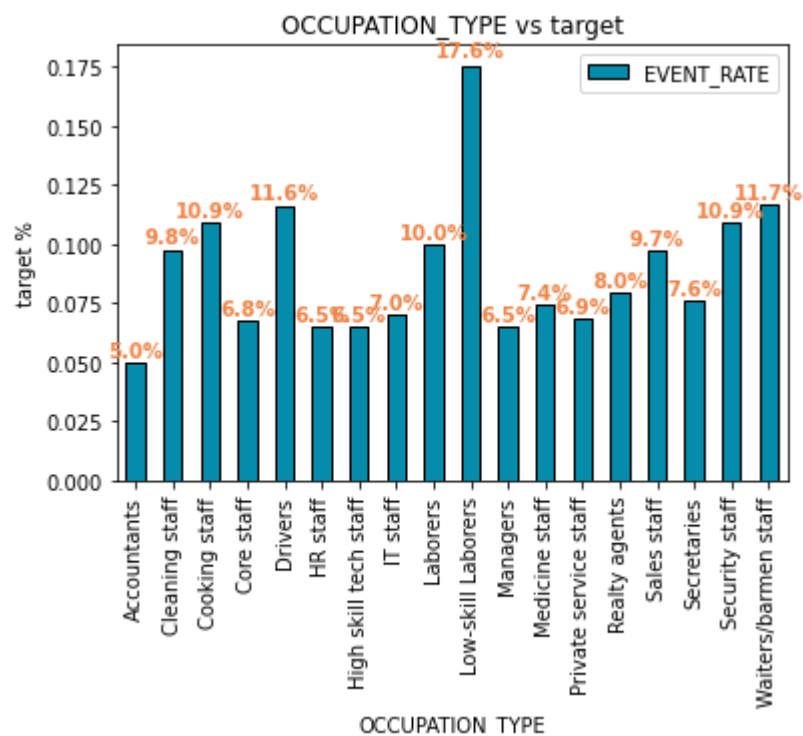
Insight

Based on the highest WoE, some of the features describe default creditors:

- NAME_INCOME_TYPE = Maternity Leave
- MICROLOAN = 20%
- CREDIT_DAY_OVERDUE = 2 Days
- OCCUPATION_TYPE = Low-Skill Laborers
- NAME_EDUCATION_TYPE = Lower Secondary

FEATURE IMPORTANT

Business Recommendation

- Credit analyst team must be aware of creditors who have a history of problems paying off credit applications
- Based on historical data, if creditors have more credit active, the probability of default is high

MODELING & EVALUATION

	Model	Accuracy Model	AUC Test	AUC Train	Precision Test	Precision Train
4	Light GBM	91.065423	65.380360	96.741107	0.000000	89.589479
0	Logistic Regression	60.221515	65.241815	67.052199	62.365051	64.320098
6	Linear SVC	59.979373	65.234904	67.053369	62.616119	64.570656
3	XGBoost	91.076633	64.997677	97.129315	0.351494	89.803626
8	Gradient Boosting	91.018340	62.001093	96.198048	0.225960	87.448623
2	Random Forest	90.796377	60.702073	99.998643	0.677881	99.877189
7	Naive Baiyes	36.469217	60.430297	66.832182	79.613357	86.229156
5	MLP Classifier	84.973768	60.399350	93.028016	14.963595	79.330157
1	Decision Tree	81.810233	52.066263	99.999826	15.767010	99.845098

DECILE	MIN	MAX
1	0.68723	0.95209
2	0.63183	0.68723
3	0.58848	0.63183
4	0.54811	0.58847
5	0.50796	0.54810
6	0.46644	0.50796
7	0.42112	0.46644
8	0.36818	0.42112
9	0.29705	0.36817
10	0.03655	0.29705

Very High Risk
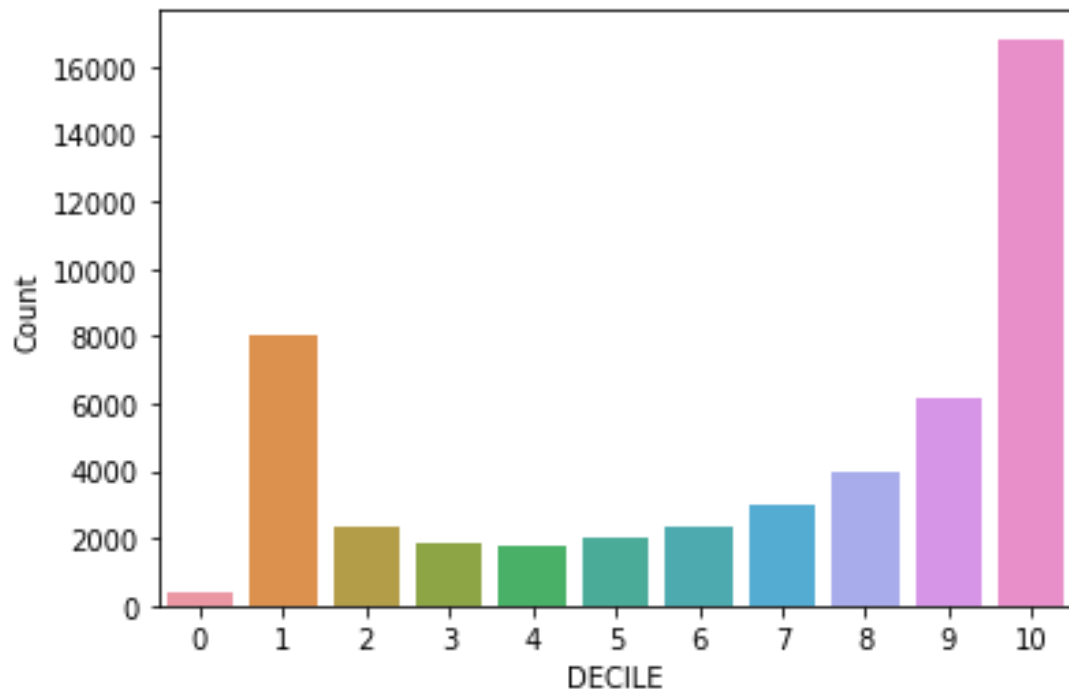High Risk
Medium Risk
Low Risk
Very Low Risk

- The selected model is Logistic Regression
- Using the Supervised Learning algorithm
- Have a high Precision Test score to minimize false-negatives
- Have a high AUC Test score to maximize credit scoring

Insight

- The higher the score, the higher the probability that creditors will default
- Creditors on **Decile 1 can be rejected because it has a high probability of default**

RESULT

Business Recommendation

From **48,788 creditors, it has 8,434 creditors were found to have a high risk of default.** Therefore, these creditors can be rejected on applying credits to minimize the number of credit defaults.