HOME CREDIT

SCORECARD MODEL

Shinta Amalia Paradita

PROBLEM RESEARCH

01. Improving the accuracy of credit score predictions using Machine Learning models so that lending decisions are more accurate and the risk of bad debt can be minimized.

02. Identify customers who are truly capable of repaying loans on time so that loan schemes can be adjusted to support successful repayment.

PROBLEM RESEARCH

01.

GOALS

Manage credit risk by reducing potential financial losses by selecting applicants who are capable of repaying loans on time.

Optimizing credit risk prediction models to improve accuracy in assessing loan applicant eligibility.

02.

OBJECTIVE

Developing machine learning models that can predict credit risk with high precision and identify applicants at risk of default.

Identifying factors that contribute to credit risk to provide a clearer picture of the factors that increase the likelihood of default.

03.

METRICS

Accuracy

F1-Score

ROC-AUC

Precision

Recall

MCC

DATASET

- **application_train**

Contains complete customer data with targets (e.g., whether loans are repaid on time or not). This dataset is used to train Machine Learning models, including preprocessing, feature engineering, and model evaluation.

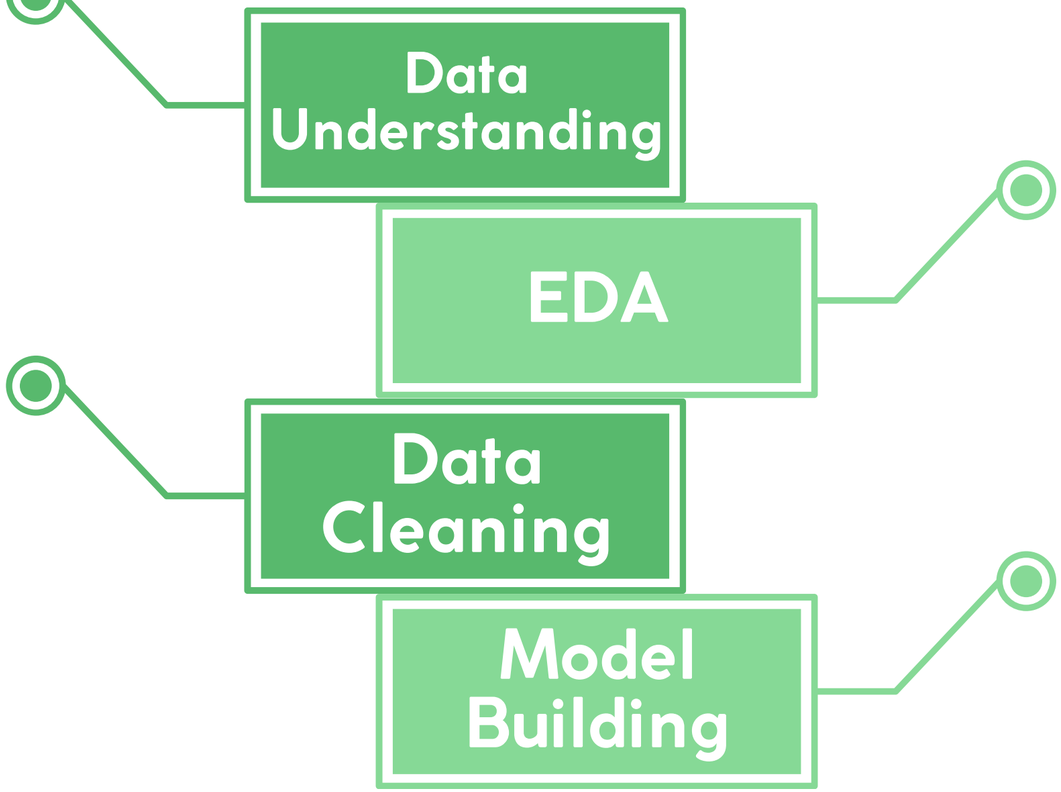
- **application_test**

Contains customer data without targets. This dataset is used to generate credit score predictions for new customers or to test model performance on previously unseen data.

WORKFLOW

Loading the dataset, checking the structure and data types, and analyzing the initial statistical summary to understand the characteristics of the data.

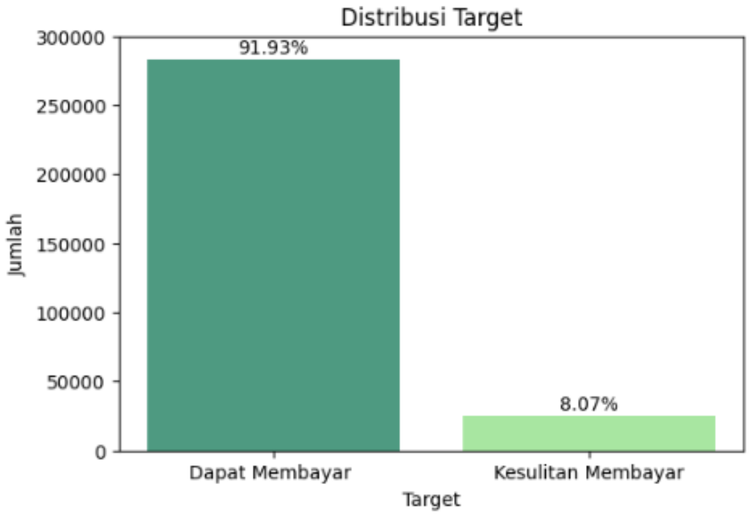
Handle missing values and duplicates, and perform data transformations such as normalization or encoding to prepare it for use in the modeling stage.



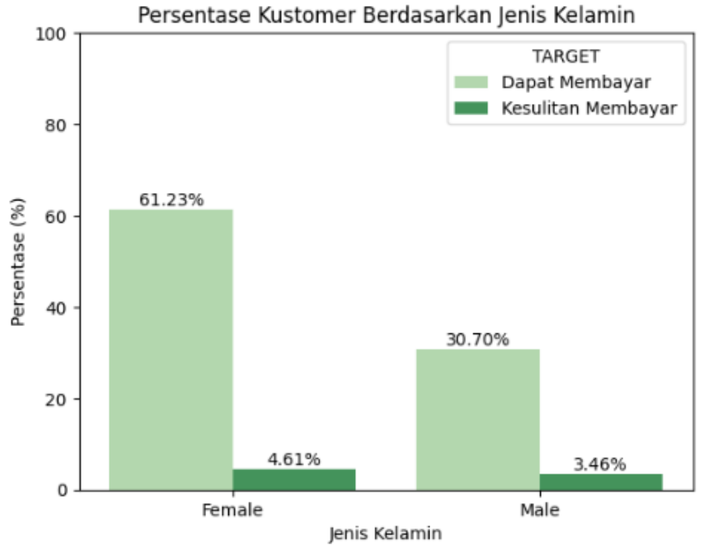
Visualize data, see distributions and relationships between variables, and explore relevant business patterns or insights.

Selecting the most influential features, training the machine learning model with the cleaned data, and evaluating the model's performance to obtain the best prediction results.

BUSINESS INSIGHT

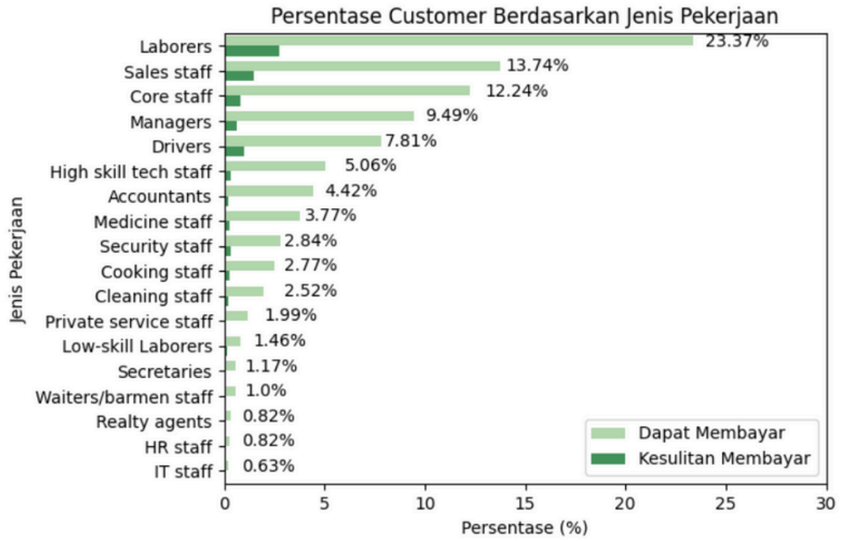


A total of 8.07% of clients had difficulty repaying their loans, while 91.93% of customers were able to repay their loans smoothly. Although the proportion of defaults is relatively small, this is still important to note as it can affect business risk.

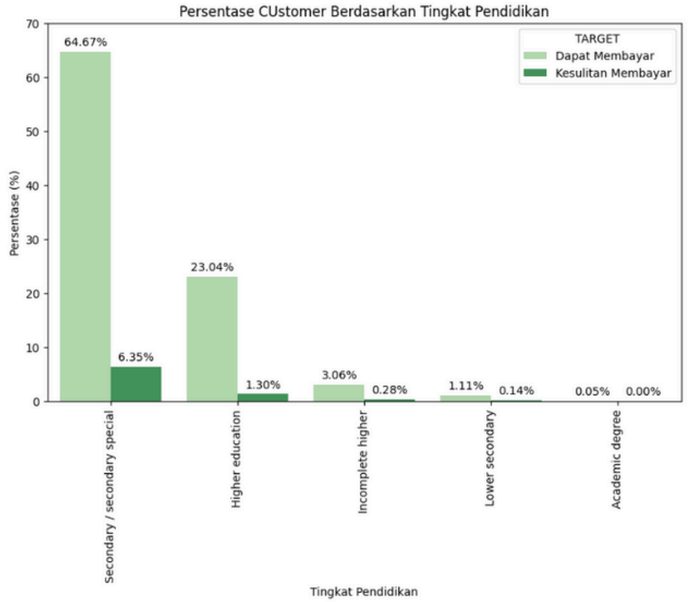


The client base is dominated by women (61.23%), but they also contribute to a higher proportion of defaults (4.61%) compared to men (3.46%). This shows that although women dominate the number of customers, the risk of default in this group is slightly higher.

BUSINESS INSIGHT



The majority of clients come from the Laborers group (23.37%), followed by Sales staff (13.74%) and Core staff (12.24%). Although Laborers dominate the number of customers who are able to repay their loans, this group also has a large proportion who have difficulty repaying. Meanwhile, Managers (9.49%) and Drivers (7.81%) also contribute a significant proportion with relatively more stable payment distributions.



Most clients come from secondary/secondary special education (64.67%), which also recorded the highest proportion of repayment difficulties (6.35%). Meanwhile, the higher education group ranked second (23.04%) with a lower level of payment difficulties (1.30%).

MODEL BUILDING

ML MODEL	ACCURACY	F1-SCORE	PRECISION	RECALL	ROC-AUC	MCC
LOGISTIC REGRESSION	0.608	0.700	0.880	0.610	0.645	0.118
K-NEIGHBORS CLASSIFIER	0.701	0.770	0.860	0.70	0.551	0.040
RANDOM FOREST	0.909	0.880	0.860	0.910	0.605	0.034
DECISION TREE	0.814	0.830	0.860	0.810	0.516	0.026

Based on the evaluation results, the Random Forest model showed the best performance compared to other models. This proves that Random Forest is most effective in identifying customers who are able to make repayments and predicting potential credit problems, thereby supporting improved accuracy in assessing borrower eligibility.

Bussiness Recommandation

01. **Improving Credit Assessment Quality**

Use the model's predictions to refine the credit scoring process, enabling more targeted lending decisions. The model can also help distinguish between high- and low-risk customers more accurately, ultimately improving the quality of the company's credit portfolio and supporting sustainable business growth.

02. **Regular Monitoring and Validation**

Monitor the model's performance regularly to ensure the consistency of prediction results. If there are changes in customer behavior or market trends, perform retraining to keep the model relevant. This approach ensures that the company's risk management strategy remains adaptive and effective in supporting long-term business objectives.

03. **Strategies Based on Occupation and Education**

Customer groups such as laborers and sales staff, as well as customers with a secondary education, have a higher proportion of defaults. Companies can adjust limits and interest rates and provide financial education programs for these at-risk groups, while customers with more stable occupations or education receive more flexible credit facilities.

Terima Kasih