

PROJECT MILESTONE 4

PACE: Plan Stage

- What is the main purpose of this project milestone?

The main purpose of this project milestone is to conduct feature engineering between the predictive features and the target feature.

- What is the research question for this project?

"Which of the features have the most predictive power on client's loan default?"

- What is the importance of feature engineering?

Feature engineering is a critical step in machine learning that can significantly impact model performance.

Here's why feature engineering is important:

1. Improved Model Performance:

- Relevant Features: By creating relevant features, you can provide the model with more information about the data, leading to better predictions.
- Reduced Noise: Removing irrelevant or noisy features can improve model accuracy and reduce overfitting.
- Enhanced Generalization: Well-engineered features can help the model generalize better to unseen data.

2. Faster Training:

- Reduced Feature Space: By selecting the most important features, you can reduce the dimensionality of the data, leading to faster training times.

3. Better Interpretability:

- Meaningful Features: Well-engineered features can make the model more interpretable by humans.

4. Handling Complex Relationships:

- Feature Interactions: By creating interaction terms between features, you can capture complex relationships that might not be apparent in the raw data.

PACE: Analyze & Construct Stages

- In general, why is feature engineering useful?

In general, feature engineering is useful because it improves model performance, the training is faster, the engineered features are more interpretable, and also, they can capture complex relationships between the predictive and the target features.

- How did feature engineering help you analyze your data?

Without feature engineering, by using only the basic features the roc_auc score was significantly lower compared with the same score when using both basic and engineered features.

PACE: Execute Stage

- What key business or organizational insight(s) emerged from the feature engineering?

The key business insight is that interactions between the most predictive basic features: client's age, income, months employed, and loan interest rate and amount contribute most to the client's loan default.

- What recommendations do you propose based on your results?

After feature engineering, taking into account the benefits of tree-based ML models, I propose to evaluate different types of models to assess the best roc_auc score on the validation dataset.