

Loan Default Prediction Project | Feature Engineering (Milestone 4)

Executive Summary Report


Prepared by: PhD Aleksandar Osmanli


Project Overview

I'm currently developing a data analytics project aimed at decreasing overall loan default among borrowers by predicting which client attributes most contribute their loan default. For the purposes of this project, loan default means borrowers failed to pay their loans to the financial institution.


Feature engineering offer possibility to enhance the ML models and achieve better results by combining the most predictive basic features into new features, which can add predictive power to the modeling. **This report offers details and key insights from Milestone 4, which impact the future development of the overall project.**

Milestone 4 – Feature Engineering

 **Target Goal:** Create new interactive features and check their additional predictive power on user subscription churn.

 **Methods:**

- Assessment of loan default among the most predictive basic features
- Creation of interactive features
- Evaluation of their correlation with loan default

 **Impact:** By deep engineering, the new features can reveal important relationships and predict the target feature

Correlation of Features with Loan Default

Feature	Correlation
Age	-0,168
InterestRate	0,131
Income	-0,099
MonthsEmployed	-0,097
LoanAmount	0,087
HasCoSigner_Yes	-0,039
HasCoSigner_No	0,039
EmploymentType_Full-time	-0,039
EmploymentType_Unemployed	0,035
HasDependents_No	0,035
HasDependents_Yes	-0,035
CreditScore	-0,034
NumCreditLines	0,028
MaritalStatus_Married	-0,027
HasMortgage_Yes	-0,023
HasMortgage_No	0,023
Education_HighSchool	0,023
LoanPurpose_Home	-0,022
MaritalStatus_Divorced	0,020
DTIRatio	0,019
Education_PhD	-0,018
Education_Masters	-0,013
LoanPurpose_Business	0,011
Education_Bachelors	0,009
MaritalStatus_Single	0,007
EmploymentType_Part-time	0,006
LoanPurpose_Auto	0,004
LoanPurpose_Education	0,004
LoanPurpose_Other	0,003
EmploymentType_Self-employed	-0,003
LoanTerm	0,001

Next steps

→ Taking into account the benefits of tree-based ML models, evaluate different types of models to assess the best roc_auc score on the validation dataset.