

Home Credit Default Risk competition

Project Description

- Our group consists of:

Davide Zingaro 1873954

Negin Amininodoushan 1915791

- The final leaderboard score is as following:

Private Score	Public Score
.78391	.78824

- Project features :

- Adding polynomial features:

We created another train and test data frames with polynomial combinations of 5 most correlated features with the target class. In modeling phase, we also tried training models with this polynomial features to see if it can improve the AUC score.

- Creating new features:

We created new columns that seemed to have a meaningful relationship with the target value.

- Adding data from extra datasets:

We added features from other 5 datasets based on the mean value for each SK_ID_CURR. We also created some new features like number of previous loans, number of previous applications, frequency of applying for loans, risk factor based on delays in paying loans, etc.

- Using 3 classification models:

We used logistic regression, random forest and light GBM models for training data. We applied these models on both datasets with and without polynomial features. The model with the best AUC score was light GBM on polynomial dataset.