PROJECT MILESTONE 5

PACE: Plan Stage

- What am I trying to solve or accomplish?

I'm trying to build best prediction model in order to predict financial loan default.

- What are my initial observations when I explore the data?

The company's dataset has 364.782 rows of data divided into training set of 255.347 rows (70%) and test set of 109.435 rows (30%) and 18 columns (variables) in the training set, and 17 columns (variables) in the test set.

- What resources do you find yourself using as you complete this stage?

I was using the following tree-based ML models to evaluate the churn prediction: LightGBM, XGBoost, CatBoost and HistGBM.

I was using the GridSearch technique to find the optimal hypeparameter values on different number of cross-validation folds.

PACE: Analyze Stage

- What are some purposes of EDA before constructing a multiple tree-based ML models?

Some of the purposes of the EDA are analyzing and discovering data from the dataset and looking for correlations.

PACE: Construct Stage

- Can you improve the "champion" model? Is there anything you would change about the model?

The model could be improved by continuously fine tuning of the hyperparameters, but the improvements would be very small, especially when I stabilize the predictive power of the model.

It could also be helpful to scale the predictor variables, and/or to reconstruct the model with different combinations of predictor variables to reduce noise from unpredictive features.

PACE: Execute Stage

- What key insights emerged from your model(s)?

Besides the basic features, the newly engineered features bring combined predictive power with new insights in correlation with the client's loan default.

- What business recommendations do you propose based on the models built?

This model should be used to make significant business decisions because of its excellent roc_auc score on the test set – 75,74%

- What potential recommendations would you make?

Due to the model results, I recommend using the key insights from this project milestone to guide further exploration.

- What business/organizational recommendations would you propose based on the models built?

It would be helpful to have information about the client's spending habits and financial literacy (overuse of credit cards or payday loans, or lack of savings or emergency funds). Additionally, the industries prone to layoffs or low job security would be a valuable information in predicting loan default.