Applied Machine Learning Project

Home Credit Default Risk

Group-4

Team Members:

- Vishal Bhalla
- Ritu Sanjay
- David Drummond
- Ayush Srivastav

Project Abstract:

Home Credit is a non-banking financial institution that uses sources like transnational data and telecom data, to get insights into the applicant's finances and determine their repayment capabilities and accordingly help them to obtain loans. This dataset is shared by the company to the Kagglers' to come up with better machine learning solutions than the company currently uses and better serve its customers.

We are required to understand the dataset, extract relevant features, perform experiments using various machine learning algorithms and determine the best performing parameters.
We will do EDA, visualize the data for better presentation, find the correlation between the various data elements, create new features from existing data, roll up and aggregate wherever necessary. All this has been done using pipelines so that our solutions remain clean and reproducible.

We will extract new features from the dataset that are currently hidden, and our model will perform better than the existing model once we train the model on the existing and derived features.

The HCDR project was a binary classification problem with numerous datasets which were merged to form a master training/testing set. In phase-0 we explored the datasets and ran the base model that included only the application_train to fit the model. Phase-1 involved performing EDA and feature selection - PCA, chi-squared scores and correlation metrics were used for the same. We were also able to create new features for the training set. The next step i.e. phase-2 included creating transformer classes for each of the datasets and creating the respective pipelines. We tested models on the transformed application_train dataset in this phase (results above). The last phase included merging all transformed datasets and implementing XGBoost, LightGBM and Neural Network. We conclude that the best model was the LightGBM with an AUC (Kaggle) score of 76.67%.