PROJECT REPORT

Group Members: Clara Punzi (1894722), Meher Kavya Indrani Koppisetti (1921475), Fidan Aliyeva (1921618)

Project Name: Home Credit Default Risk

Project Score: 0.78230 (Public), 0.77501 (Private)

Project Description:

The aim of this project is to build a model based on machine learning and statistical methods to help Home Credit predict their clients' repayment abilities. We have built our model according to the following steps:

1. **Exploratory Data Analysis**
    1.1. Analysis of the distribution of the target column
    1.2. Aggregation of the *train* and *test* datasets, detection of anomalies
    1.3. Processing of the other available datasets, anomalies' detection and aggregation to a single dataset; definition of new features from aggregation phases
    1.4. Encoding categorical variables and removing missing values
2. **Feature Engineering**
    2.1. Domain knowledge features
    2.2. Dimensionality reduction
        2.2.1. Standardization of the features
        2.2.2. PCA
3. **Logistic Regression Model**
    3.1. LightGBM on reduced dataset
    3.2. LightGBM on full dataset