MSDS PT2025 -Data Mining and Wrangling 1

Learning Team 4

Co, Julie | Quiddaoen, Gianni | Seumal, Anthony | Tan, Allan

For code instructions, you may refer to our Github repo:
https://github.com/MSDSPT2025B-Term-2-LT4/HOME_CREDIT_KAGGLE

Home Credit – Credit Scoring & Client Default Predictor Model

I. Introduction

The Home Credit - Credit Risk Model Stability competition on Kaggle challenges participants to build machine learning models to predict the repayment capabilities of potential clients. The competition addresses the significant issue of financial inclusion, focusing on individuals with little or no credit history. Accurate risk prediction models can enhance the accessibility of loans, potentially improving the lives of many who are traditionally excluded from financial services. Our main objective was to develop a credit default model for the dataset provided by the Home Credit team.

Business Objective

Founded in 1997, Home Credit is an international consumer finance provider focused on responsible lending, primarily to people with little or no credit history. By broadening financial inclusion, Home Credit aims to create a positive and safe borrowing experience for the unbanked population. Home Credit has a significant presence in several countries, offering a range of financial products including personal loans, credit cards, and point-of-sale financing. Their mission is to provide accessible financial services that improve the lives of their customers.

Home Credit's business model focuses on serving the unbanked population which traditional and well-established credit institutions barely tap due to rigid requirements and conservative risk assessment. By developing this credit default model, we aim to provide a decision support tool for Home Credit's loans approval process.

Key Elements of the Business Model

Easier Application Process:
- Simplified procedures for applying for credit, reducing the barriers to entry for potential borrowers.
- Less paperwork and faster approval times make it more convenient for applicants.

Less Rigid Metrics for Approval:
- Flexible criteria for assessing creditworthiness, allowing individuals with little or no credit history to qualify for loans.
- Utilizes alternative data sources and innovative scoring models to evaluate applicants.

Interest Rates Include Risk Premiums:
- Interest rates are adjusted to include risk premiums, which account for the higher risk associated with lending to individuals with less traditional credit profiles.
- This ensures that the financial institution remains profitable while offering inclusive financial services.

Taken into consideration the emphasis on inclusivity, the model should be flexible enough to to score default risk of applications even under these certain conditions:

1. Lack of Credit History
2. Lack of Other Supporting Documents/Information

II. Feature Engineering & Selection

Data Overview

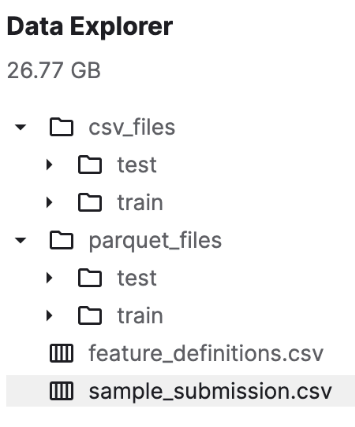
The dataset provided by Home Credit included a comprehensive set of features related to clients' personal information, previous credit history, loan details, and more. Key aspects of the dataset included:

- **Client Demographics**: Age, employment status, income, family structure, etc.
- **Credit History**: Previous loans, repayment records, defaults, etc.
- **Loan Details**: Amount, duration, purpose, etc.
- **Tax Registry:** Tax deductions, payment dates, tax queries, etc

See feature_definitions.csv for Data Dictionary

Data Directory

Dataset can be downloaded in the [Home Credit Risk Model Stability Competition](#).




```
|-- train
    |-- train_person_2.parquet
    |-- train_other_1.parquet
    |-- train_deposit_1.parquet
    |-- train_debitcard_1.parquet
    |-- train_credit_bureau_b_2.parquet
    |-- train_base.parquet
    |-- train_credit_bureau_a_2_10.parquet
    |-- train_credit_bureau_a_2_1.parquet
    |-- train_credit_bureau_a_2_0.parquet
    |-- train_credit_bureau_b_1.parquet
    |-- train_tax_registry_b_1.parquet
    |-- train_credit_bureau_a_2_7.parquet
    |-- train_static_cb_0.parquet
    |-- train_tax_registry_a_1.parquet
    |-- train_static_0_0.parquet
    |-- train_applprev_1_0.parquet
    |-- train_person_1.parquet
    |-- train_tax_registry_c_1.parquet
    |-- train_credit_bureau_a_2_9.parquet
    |-- train_static_0_1.parquet
    |-- train_credit_bureau_a_2_8.parquet
    |-- train_credit_bureau_a_2_6.parquet
    |-- train_credit_bureau_a_1_3.parquet
    |-- train_credit_bureau_a_2_3.parquet
    |-- train_credit_bureau_a_2_5.parquet
    |-- train_credit_bureau_a_2_4.parquet
    |-- train_credit_bureau_a_1_2.parquet
    |-- train_credit_bureau_a_1_1.parquet
    |-- train_credit_bureau_a_2_2.parquet
    |-- train_applprev_2.parquet
    |-- train_applprev_1_1.parquet
    |-- train_credit_bureau_a_1_0.parquet
```

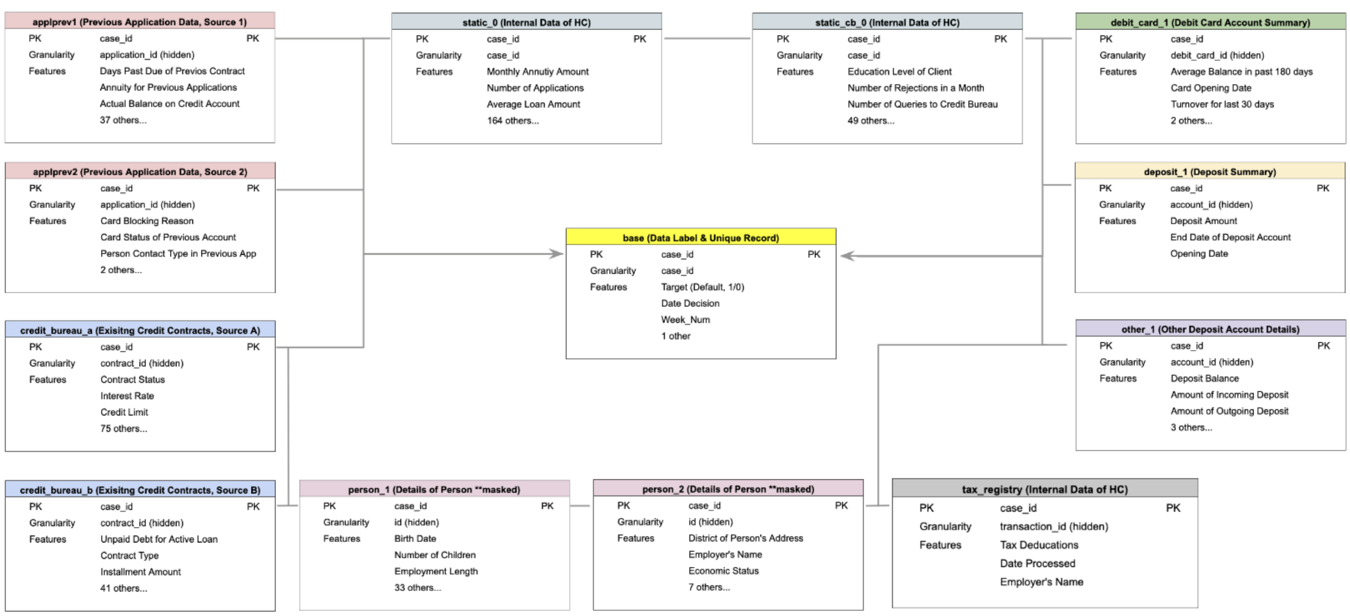
```
|-- test
    |-- test_tax_registry_a_1.csv
    |-- test_static_cb_0.csv
    |-- test_tax_registry_b_1.csv
    |-- test_static_0_2.csv
    |-- test_static_0_0.csv
    |-- test_deposit_1.csv
    |-- test_other_1.csv
    |-- test_static_0_1.csv
    |-- test_person_2.csv
    |-- test_credit_bureau_a_2_7.csv
    |-- test_debitcard_1.csv
    |-- test_person_1.csv
    |-- test_credit_bureau_b_2.csv
    |-- test_credit_bureau_a_2_6.csv
    |-- test_credit_bureau_a_2_5.csv
    |-- test_credit_bureau_a_2_2.csv
    |-- test_credit_bureau_a_2_10.csv
    |-- test_credit_bureau_a_2_4.csv
    |-- test_credit_bureau_a_2_11.csv
    |-- test_credit_bureau_b_1.csv
    |-- test_credit_bureau_a_1_2.csv
    |-- test_credit_bureau_a_1_4.csv
    |-- test_credit_bureau_a_1_3.csv
    |-- test_credit_bureau_a_2_0.csv
    |-- test_credit_bureau_a_2_8.csv
    |-- test_credit_bureau_a_1_1.csv
    |-- test_credit_bureau_a_2_9.csv
    |-- test_credit_bureau_a_2_1.csv
    |-- test_base.csv
    |-- test_applprev_1_2.csv
    |-- test_credit_bureau_a_1_0.csv
    |-- test_credit_bureau_a_2_3.csv
    |-- test_applprev_1_0.csv
    |-- test_applprev_1_1.csv
    |-- test_tax_registry_c_1.csv
    |-- test_applprev_2.csv
```

- A total of 138 files with a size of ~27 GB was provided by Kaggle. Each file has a csv and parquet
- Folders ‘csv_files/train’ and ‘parquet_files/train’ contain the total training dataset
- Folders ‘csv_files/test’ and ‘parquet_files/test’ contain the same files under the ‘train/’ folders. They have fewer content as they are only for submission testing purposes.
- Folders ‘test/’ also do not have a target variable in ‘test_base’ unlike in folders ‘train’

- Data dictionary is provided in 'feature_definitions.csv'

Data Schema

Due to the large size, some tables were chunked into smaller dataframes. Each table contains a 'case_id' column which serves as the joining key for all tables. Other tables are on a higher granularity (e.g. application_id of each case_id for past applications). Hence, there is a need to aggregate the tables.



Feature Extraction

Given the differing granularities of the tables, there is a need to further aggregate features on a 'case_id' level in order to build Model Base table – data to be plugged into the model where each row corresponds to a unique case_id or observation. General functions for aggregating numeric, categorical, and date columns were used to generate the base table.

1. *class Pipeline*

Before processing the data,

Home Credit provided information on the data types and transformations done on each of the original columns. They have attached an identifier letter on the end of the column names to provide information on this:

- P - Transform DPD (Days past due)
- M - Masking categories
- A - Transform amount
- D - Transform date
- T - Unspecified Transform
- L - Unspecified Transform

So for example, column *amount_1115A* is a numeric column calculated from a transformed amount column. It represents the credit amount of the active contract provided by the credit bureau.

2. *class Aggregator*

Due to the complexity and high-dimensionality of the dataset, aggregation was simplified and generalized for specific column data types.

- **Numerical Columns (int, float)**: Extracted statistical features such as mean, max, median, sum, last, first, and standard deviation.
- **Categorical Columns (string/objects)**: Extracted features like minimum, maximum, last, and first values.
- **Dates**: Converted date differences from the decision date to numerical values.

The result of this feature engineering process was a dataset with 1.53 million IDs and 774 columns.

Feature Selection

Feature selection is the process of selecting a subset of relevant features (variables, predictors) for use in model construction. The main goal is to improve the model's performance by removing redundant or irrelevant data, reducing overfitting, and decreasing the computational cost.

The feature selection process used in the Home Credit Risk Model Stability project. It explains how columns with low usability, high complexity, and high correlation with other features were eliminated from the dataset. The steps are as follows:

Missing Values:

- Features with more than 95% missing (null) values were excluded from the dataset.

No Variance:

- Features with zero variance were excluded. This includes numeric features with a standard deviation of zero and categorical features with only one unique value.

Complex Variables:

- Categorical features with more than 50 unique values were excluded to avoid complexity.

Correlated Variables:

- Highly correlated features (correlation > 0.9) were reduced based on their importance against the target variable. Statistical tests were used such as Chi-square tests, and ANOVA F-values to determine which features from a set of highly correlated variables are best to keep.

The result of this feature selection process was a dataset consisting of 1.5 million unique cases and 455 features, ensuring that the most relevant and manageable data was used for model training.

III. Model Development

Study Base

The dataset was divided into two subsets: an 80% training set and a 20% test set to build and evaluate the predictive models. This split was necessary to ensure that the models were trained on a large portion of the data while leaving enough data for unbiased evaluation.

The training set consists of 1,221,327 records, including 1,182,886 non-defaulters and 38,441 defaulters, resulting in a default rate of 3.15%. The test set has 305,332 records, with 295,779 non-defaulters and 9,553 defaulters, giving a default rate of 3.13%. Overall, the entire dataset contains 1,526,689 records with a default rate of 3.14%.

The data exhibits a significant class imbalance, with a very low percentage of defaulters (approximately 3.14%). This imbalance poses a challenge for model training, as it can lead to biased predictions towards the majority class (non-defaulters). Addressing this imbalance is crucial for developing an effective credit risk prediction model. Techniques such as resampling, using different performance metrics, or applying specialized algorithms can help mitigate the effects of class imbalance and improve model performance.

Baseline Models

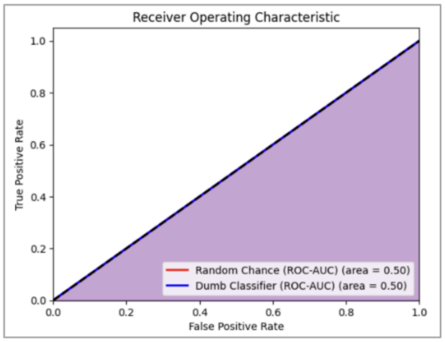
To evaluate the performance of our advanced models, we first established baseline models to serve as benchmarks. These models include:

Dumb Classifier:

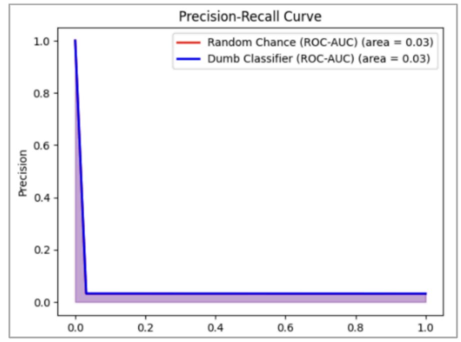
- This classifier predicts the majority class, which in this case is non-defaulters. It is used to benchmark the worst performance to beat for any imbalance dataset. By always predicting the majority class, the dumb classifier sets a lower bound for performance, highlighting the challenge posed by the highly imbalanced dataset.

Random Chance Classifier:

- This classifier makes random predictions based on the event rate (default rate) in the dataset. It provides a baseline by randomly guessing the target variable according to the observed distribution of defaulters and non-defaulters.



Metrics	Dumb Classif	Random Chance
Accuracy	96.85%	93.96%
Precision	0.00%	3.19%
Recall	0.00%	3.13%
F1 Score	0.00%	3.16%
ROC-AUC	0.5000	0.5002
PR-AUC	0.0315	0.0315



These two baseline models expectedly only performed at minimum with an ROC-AUC of 0.5 and a PR-AUC equal to the event rate at ~3%. These baseline models are crucial for understanding the effectiveness of our advanced models. By comparing their performance to these simple classifiers, we can gauge the improvements achieved by more sophisticated models.

Model Training & Tuning

As mentioned earlier in the business model, Home Credit's loan approval process needs to be robust against vastly different levels of missing information. Some applications may lack credit history, tax information, or other data necessary to predict probability of default. On top of this, credit information also contains multiple categorical features with high dimensionality. Hence, we need to pick a Machine Learning model that is robust against null values and categorical columns.

The chosen algorithms for this project are Gradient Boosting Trees specifically:

- LightGBM
- CatBoost

We defined a hyperparameter space for these 2 models to find the best-performing model. Provided below are the objective function and search method used to optimize the machine learning models:

Objective Function:

- Our goal was to maximize the average PR-AUC (Precision-Recall Area Under the Curve) using 5-fold cross-validation. This metric is particularly useful for evaluating models on imbalanced datasets.

Search Algorithm:

- We employed the Hyperopt library with the Tree of Parzen estimator to conduct the hyperparameter search. This Bayesian optimization method explores the hyperparameter space efficiently. We limited the search to a maximum of 250 trials per algorithm to ensure thorough exploration without excessive computational cost.

Due to usage limitations in Kaggle of 12 hours and the dependency of Hyperopt-Tree of Parzen estimators with results of previous Hyperopt trials, we were only able to run 53 trials in a 24-hour compute session (2 concurrent Kaggle session).

LightGBM performed the best out of all the trials with an ROC-AUC of 0.84925 and PR-AUC of 0.18499 on the average 5-Fold CV performance on the training set.

Average of 5-Fold Cross Validation (Training Dataset - 80%)			
model_name	**ROC AUC**	**PR AUC**	**Gini**
LGB_Tune34	0.84925	0.18499	0.69850
LGB_Tune28	0.84923	0.18492	0.69846
LGB_Tune3	0.84956	0.18449	0.69911
LGB_Tune2	0.84909	0.18426	0.69819
LGB_Tune33	0.84851	0.18343	0.69702
LGB_Tune23	0.84866	0.18339	0.69732
LGB_Tune26	0.84862	0.18276	0.69725
LGB_Tune25	0.84778	0.18271	0.69556
LGB_Tune22	0.84774	0.18259	0.69548
LGB_Tune27	0.84726	0.18207	0.69453
LGB_Tune5	0.84733	0.18148	0.69467
LGB_Tune20	0.84778	0.18144	0.69556
LGB_Tune21	0.84819	0.18111	0.69638
LGB_Tune24	0.84786	0.18049	0.69572
LGB_Tune6	0.84447	0.17763	0.68894
LGB_Tune14	0.84529	0.17756	0.69057
CatBoost_Tune12	0.84359	0.17745	0.68718
CatBoost_Tune0	0.84425	0.17736	0.68850
LGB_Tune12	0.84364	0.17728	0.68728
CatBoost_Tune6	0.84302	0.17720	0.68604
CatBoost_Tune14	0.84400	0.17672	0.68800
LGB_Tune30	0.84382	0.17652	0.68764
CatBoost_Tune15	0.84173	0.17499	0.68347
LGB_Tune10	0.84192	0.17495	0.68384
CatBoost_Tune10	0.84176	0.17164	0.68352

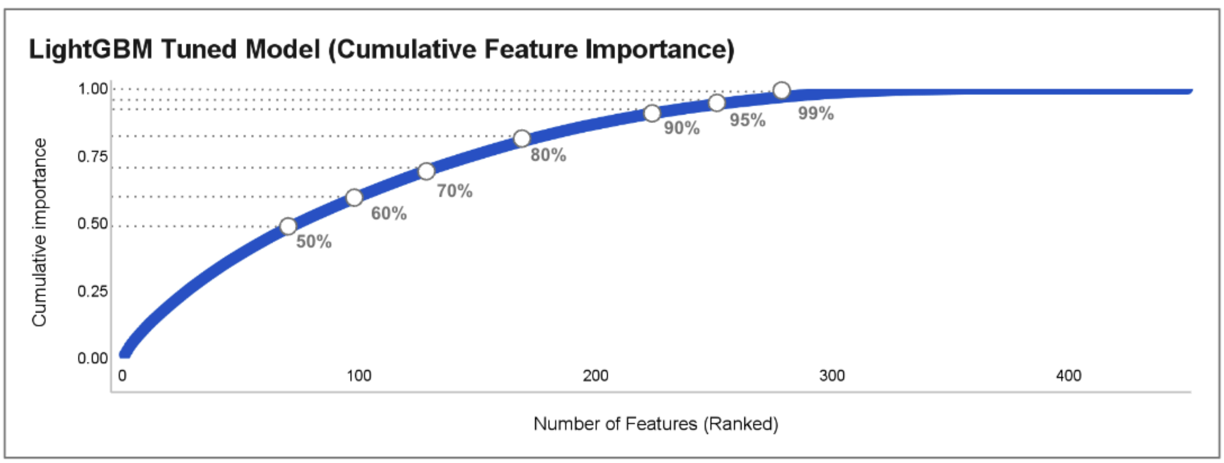
Optimizing Features

The LightGBM Tuned Model, with 466 features, was selected as the best performing model based on its evaluation metrics. The ROC curve AUC of 0.86 and PR curve AUC of 0.19 demonstrate its strong predictive capabilities and effectiveness in handling the imbalanced dataset. By optimizing thresholds and cutoffs, this model can be fine-tuned further for deployment in real-world credit risk assessment scenarios.

To enhance the performance of our LightGBM Tuned Model, we employed Recursive Feature Elimination (RFE) using the importance scores of the features from the optimized model. The aim was to identify the most relevant features and eliminate those that contributed the least to the model's predictive power.

Process of Recursive Feature Elimination:

- We started with all the features and recursively removed the least important ones, evaluating the model's performance at each step.
- The cumulative importance of the features was calculated to determine the contribution of the top features.



%Cum Imp	50%	60%	70%	80%	90%	95%	99%	100%
n_features	73	99	129	166	218	258	312	450
ROC-AUC	0.8508	0.8534	0.8551	0.8556	0.8565	0.8561	0.8566	0.8561
PR-AUC	0.1830	0.1867	0.1887	0.1897	0.1907	0.1919	0.1903	0.1896

Findings:

- The chart shows the cumulative importance of the features ranked by their importance scores.
- Performance metrics (ROC-AUC and PR-AUC) were evaluated at different levels of cumulative importance (50%, 60%, 70%, 80%, 90%, 95%, 99%, and 100%).
- We observed that slightly better performance was achieved when the top 95% of features were used (removing 192 variables).
- A continued decrease in performance was noted when using less than 90% of the features, indicating that some important features were being excluded.

Performance Metrics:

- At 90% cumulative importance (218 features), the model achieved an ROC-AUC of 0.8565 and a PR-AUC of 0.1907, which were among the highest values observed.
- The final model's performance metrics at various feature levels are shown in the table, highlighting the impact of feature selection on the model's accuracy and precision.

By optimizing the features, we were able to enhance the model's performance and reduce complexity. The RFE process helped in identifying the most significant features that contributed to the predictive accuracy of the LightGBM Tuned Model, ensuring a robust and efficient credit risk assessment tool.

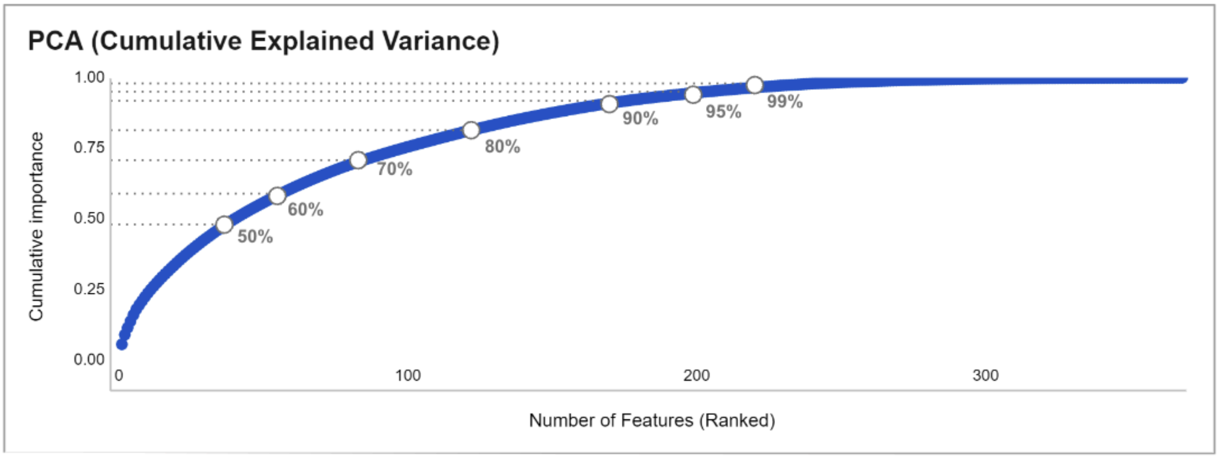
Experimenting with PCA

To explore the effect of dimensionality reduction on model performance, we applied Principal Component Analysis (PCA) to the dataset. PCA helps in transforming the original features into a smaller set of uncorrelated components, capturing the maximum variance in the data. The cumulative explained variance was used to evaluate the impact of varying the number of principal components.

Process of PCA Dimensionality Reduction:

- We experimented with different levels of cumulative explained variance (50%, 60%, 70%, 80%, 90%, 95%, 99%, and 100%) to analyze the effect on model performance.

- The number of principal components corresponding to each level of cumulative explained variance was determined.



%Cum Imp	50%	60%	70%	80%	90%	95%	99%	100%
n_features	122	140	164	198	245	285	351	450
ROC-AUC	0.8198	0.8223	0.8232	0.8247	0.8259	0.8274	0.8289	0.8335
PR-AUC	0.1467	0.1496	0.1511	0.1522	0.1529	0.1534	0.1550	0.1580

Findings:

- The chart shows the cumulative explained variance for the ranked principal components.
- Performance metrics (ROC-AUC and PR-AUC) were evaluated at different levels of cumulative explained variance.
- The overall performance was observed to be worse than the original dataset with all features.
- Reducing the number of principal components continuously decreased the performance, indicating that important information was being lost.

Performance Metrics:

- At 100% cumulative explained variance (450 features), the model achieved an ROC-AUC of 0.8335 and a PR-AUC of 0.1580.
- The final model's performance metrics at various levels of cumulative explained variance are shown in the table, highlighting the impact of PCA on the model's accuracy and precision.

By experimenting with PCA, we found that while it is a useful technique for dimensionality reduction, the performance of our credit risk model deteriorated as the number of principal components was reduced. This suggests that retaining all features or using other feature selection methods might be more effective for this particular dataset.

Using PCA for dimensionality reduction, we observed a decrease in model performance with fewer principal components. The cumulative explained variance analysis revealed that reducing the number of features led to a loss of important information, adversely affecting the model's predictive power. This experiment highlights the importance of carefully selecting and retaining relevant features for building a robust credit risk model.

Result and Impact

Loan Approval

Highlighting the top 20 features that provide a balance between credit data and non-credit data. It categorizes the features into those relevant for individuals with little to no credit history and those with established credit histories. The SHAP (SHapley Additive exPlanations) values plot further elucidates the impact of these features on the likelihood of default.

Key Features

Little to None Credit History:

- **Income**: Total income of the applicant.
- **Income Type**: Type of income (e.g., salary, business).
- **Age**: Age of the applicant.
- **Education**: Education level of the applicant.
- **Tax Payments**: History and consistency of tax payments.

With Credit History:

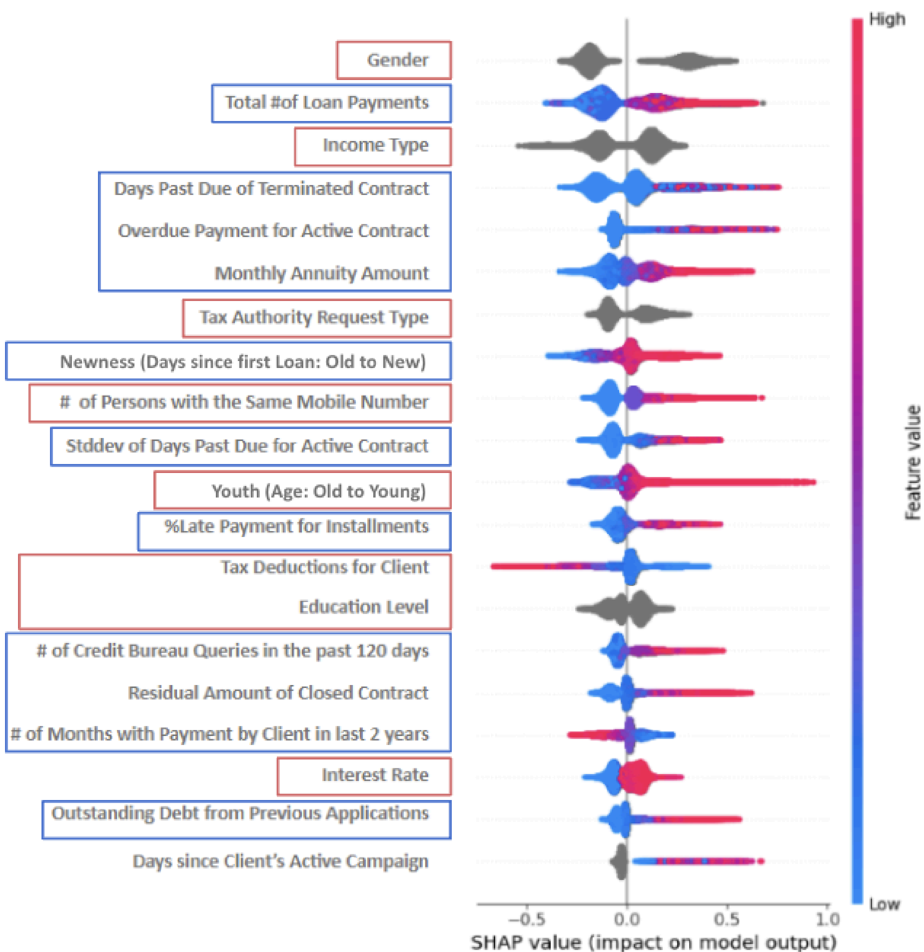
- **Promptness of Payments**: Timeliness in making payments.
- **Consistency of Payments**: Regularity and reliability of payments.
- **Delinquency History**: History of late or missed payments.
- **Annuity Amount**: Monthly annuity payment amount.
- **Loan Tenure**: Duration of previous loans.

These features are essential on top of those required for individuals with little to no credit history.

SHAP Values Plot

The SHAP values plot provides a detailed visualization of the impact of each feature on the model's output. The features are ranked by their importance, and the SHAP values indicate how much each feature contributes to the likelihood of default. Key insights include:

- **Days Past Due of Terminated Contract**: High values increase the likelihood of default.
- **Gender and Income Type**: These demographic features also significantly impact the model.
- **Monthly Annuity Amount**: Higher annuity payments are associated with increased risk.
- **Education Level and Tax Deductions**: Higher education levels and tax deductions tend to decrease the likelihood of default.
- **Interest Rate and Outstanding Debt**: Higher interest rates and outstanding debts from previous applications increase the risk of default.



Understanding the critical features necessary for loan approval and their impact on the likelihood of default is crucial for building robust credit risk models. The combination of credit and non-credit data provides a comprehensive view of an applicant's financial behavior, enabling better risk assessment and more informed decision-making. The SHAP values plot offers transparency into the model's decision-making process, highlighting the significance of each feature in predicting loan defaults.

Defaulter Profile

The types of applicants with a high likelihood of defaulting on loans, based on their credit history. It categorizes applicants into two groups: those with little to no credit history and those with established credit histories. It also presents a comparative analysis of features between defaulters and good loaners.

Types of Applicants Likely to Default

Little to None Credit History:

- **Private Employees**: Often have unstable or insufficient income.
- **Duplicate/Redundant Applications**: May indicate desperation or multiple failed attempts to secure credit.
- **Younger Loaners**: Typically have less financial stability and credit experience.
- **Lower Tax Bracket**: Indicates lower income and potential financial constraints.

With Credit History:

- **Late Payers**: History of making late payments.
- **Delinquent Loaners**: Previous delinquencies on loans.
- **Irregular Payers**: Inconsistent payment patterns.
- **Larger Annuity Amounts**: Higher monthly financial obligations.
- **Newer Loaners**: Less established credit history with the current lender.

Comparative Analysis of Features

The table provides a comparative analysis of key features between defaulters and good loaners:

Features	Defaulters	Good Loaners
Gender	Male	Female
Payment Frequency	16x	12x
Income Type	Private Sector Employee	Other, Retired Pensioner
Days Past Due of Terminated Contract	7.37	0.28
Monthly Annuity Amount	3,288.40	3,142.20
Tax Authority Request Type	DEDUCTION_6	PENSION_6
Years since First Credit	1.5 Years	2 Years
Persons with the same Mobile Number	2	1
Std Dev of Days Past Due	0.45	0
Age	38 yrs	43 yrs
%Late Payments on Installation	29%	8%
Taxes	5,800.00	8,520.60
Education Level	MASKED CATEGORIES	
# of Credit Bureau Queries	2	1
Residual Amount of Closed Contract	23,339.93	10,449.90
Months with Payment by Client	8	10
Interst Rate	0.30	0.28
Oustanding Debt from Previous Application	11,695.25	7,837.31

- **Gender**: Males have a higher likelihood of defaulting compared to females.
- **Payment Frequency**: Defaulters tend to have a higher payment frequency.
- **Income Type**: Private sector employees are more likely to default compared to retirees or pensioners.
- **Days Past Due of Terminated Contract**: Defaulters have significantly more days past due.
- **Monthly Annuity Amount**: Defaulters have higher monthly annuity amounts.
- **Tax Authority Request Type**: Different tax request types correlate with default likelihood.
- **Years Since First Credit**: Defaulters generally have a shorter credit history.
- **Persons with the Same Mobile Number**: Defaulters tend to share mobile numbers with others more frequently.

- **Standard Deviation of Days Past Due**: Higher variability in days past due among defaulters.
- **Age**: Defaulters tend to be younger on average.
- **% Late Payments on Installments**: Higher percentage of late payments among defaulters.
- **Taxes**: Lower tax payments among defaulters.
- **Education Level**: Higher education levels correlate with lower default rates.
- **Number of Credit Bureau Queries**: More queries indicate higher default risk.
- **Residual Amount of Closed Contract**: Higher amounts are associated with defaulters.
- **Months with Payment by Client**: Fewer months with consistent payments among defaulters.
- **Interest Rate**: Higher interest rates are associated with defaulters.
- **Outstanding Debt from Previous Applications**: Higher outstanding debts correlate with default risk.

The analysis identifies specific characteristics and behaviors that increase the likelihood of loan default. By understanding these patterns, lenders can improve their risk assessment processes, tailoring their strategies to mitigate default risks. This comprehensive evaluation of applicants' features enables more accurate and fair credit decisions, contributing to better financial outcomes for both lenders and borrowers.

Impact and Strategy

The goal of this analysis was to simulate the financial impact of different rejection rates using the predictions from our credit risk model. By adjusting the rejection rate, we aimed to find the optimal balance between accepting more applicants and maintaining profitability.

Traditional Rejection Rate:

Traditionally, credit companies reject about 20% of loan applications. This conservative approach minimizes the risk of defaults but may also limit potential profits.

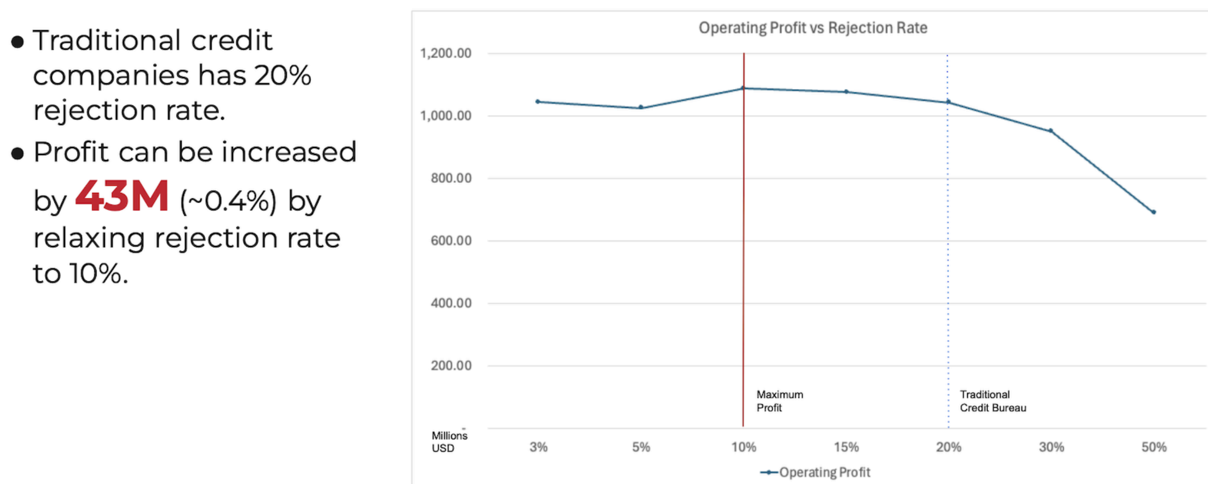
Profit Maximization:

Our analysis indicates that by lowering the rejection rate to 10%, the operating profit can be increased by $43 million (~0.4%). This is based on the model's ability to accurately assess and predict the credit risk of applicants, allowing the company to accept more loans without significantly increasing default rates.

Rejection Rate vs. Profit:

The chart illustrates the relationship between the rejection rate and operating profit. As the rejection rate decreases from 20% to 10%, the profit increases, reaching a peak at the 10% rejection rate.

Beyond this point, further decreasing the rejection rate may lead to an increase in defaults, thereby reducing profitability.



Recommendation

Model Enhancements

- **Experiment with Handling Imbalanced Datasets**:
  Explore different techniques to address class imbalance, such as oversampling, under sampling, and synthetic data generation (e.g., SMOTE).
- **Separate Models for Those with Credit History and Those without**:

Develop distinct models for applicants with and without credit history to tailor the risk assessment process to the specific characteristics of each group.

- **Explore Other Algorithms**:
  Test various machine learning algorithms (e.g., Neural Networks, Support Vector Machines, Random Forest) to identify the most effective models for predicting credit risk.

Consider Market Nuances

- **Demographics Might Be Skewed for Each Location**:
  Take into account demographic variations across different locations when developing and deploying models.
- **Different Markets (i.e., Jurisdictions) Have Different Financial Legal Frameworks**:
  Adapt models to comply with local financial regulations and legal requirements in different jurisdictions.
- **Identify Locations with Good Credit Scores for Expansion**:
  Use data-driven insights to identify regions with favorable credit profiles for business expansion.

Dynamic Product Tiers

- **Adjust Credit Limit, Fees, and Max Loan Terms According to Default Risk and Other Significant Features**:
  Implement a tiered product structure that adjusts credit terms based on the applicant's risk profile. This can include varying credit limits, interest rates, fees, and loan terms to better align with the default risk.

To optimize the credit risk modeling process and enhance financial performance, we recommend the following steps:

1. **Enhance Models**: Address imbalanced datasets, develop separate models for different applicant groups, and explore a variety of algorithms.
2. **Consider Market Nuances**: Adapt to demographic variations, comply with local regulations, and identify promising markets for expansion.
3. **Implement Dynamic Product Tiers**: Tailor credit products to the risk profiles of applicants, adjusting terms to mitigate risk and maximize profitability.

By following these recommendations, financial institutions can improve their risk assessment capabilities, offer more personalized financial products, and achieve better overall financial outcomes.

Future Work

Future improvements could focus on:

- **Advanced Modeling Techniques**: Exploring deep learning approaches and automated machine learning (AutoML) for further performance gains.
- **Real-time Monitoring**: Implementing real-time monitoring systems to promptly identify and address model performance issues.

APPENDIX

Variable	Description
actualdpd_943P	Days Past Due (DPD) of previous contract (actual).
actualdpdtolerance_344P	DPD of client with tolerance.
addres_district_368M	District of the person's address.
addres_role_871L	Role of person's address.
addres_zip_823M	Zip code of the address.
amount_1115A	Credit amount of the active contract provided by the credit bureau.
amount_416A	Deposit amount.
amount_4527230A	Tax deductions amount tracked by the government registry.

amount_4917619A	Tax deductions amount tracked by the government registry.
amtdebitincoming_4809443A	Incoming debit card transactions amount.
amtdebitoutgoing_4809440A	Outgoing debit card transactions amount.
amtdepositbalance_4809441A	Deposit balance of client.
amtdepositincoming_4809444A	Amount of incoming deposits to client's account.
amtdepositoutgoing_4809442A	Amount of outgoing deposits from client's account.
amtinstpaidbefduel24m_4187115A	Number of instalments paid before due date in the last 24 months.
annualeffectiverate_199L	Interest rate of the closed contracts.
annualeffectiverate_63L	Interest rate for the active contracts.
annuity_780A	Monthly annuity amount.
annuity_853A	Monthly annuity for previous applications.
annuitynextmonth_57A	Next month's amount of annuity.
applicationcnt_361L	Number of applications associated with the same email address as the client.
applications30d_658L	Number of applications made by the client in the last 30 days.
applicationscnt_1086L	Number of applications associated with the same phone number.
applicationscnt_464L	Number of applications made in the last 30 days by other clients with the same employer as the applicant.
applicationscnt_629L	Number of applications with the same employer in the last 7 days.

applicationscnt_867L	Number of applications associated with the same mobile phone.
approvaldate_319D	Approval Date of Previous Application
assignmentdate_238D	Tax authority data - date of assignment.
assignmentdate_4527235D	Tax authority data - Date of assignment.
assignmentdate_4955616D	Tax authority assignment date.
avgdbddpdlast24m_3658932P	Average days past or before due of payment during the last 24 months.
avgdbddpdlast3m_4187120P	Average days past or before due of payment during the last 3 months.
avgdbdtollast24m_4525197P	Average days of payment before due date within the last 24 months (with tolerance).
avgdpdtolclosure24_3658938P	Average DPD (days past due) with tolerance within the past 24 months from the maximum closure date, assuming that the contract is finished. If the contract is ongoing, the calculation is based on the current date.
avginstallast24m_3658937A	Average instalments paid by the client over the past 24 months.
avglnamtstart24m_4525187A	Average loan amount in the last 24 months.
avgmaxdpdlast9m_3716943P	Average Days Past Due (DPD) of the client in last 9 months.
avgoutstandbalancel6m_4187114A	Average outstanding balance of applicant for the last 6 months.
avgpmtlast12m_4525200A	Average of payments made by the client in the last 12 months.
bankacctype_710L	Type of applicant's bank account.
birth_259D	Date of birth of the person.
birthdate_574D	Client's date of birth (credit bureau data).
birthdate_87D	Birth date of the person.

byoccupationinc_3656910L	Applicant's income from previous applications.
cacccardblochreas_147M	Card blocking reason.
cancelreason_3545846M	Application cancellation reason.
cardtype_51L	Type of credit card.
childnum_185L	Number of children of the applicant.
childnum_21L	Number of children in the previous application.
classificationofcontr_1114M	Classificiation of the active contract.
classificationofcontr_13M	Classificiation of the active contract.
classificationofcontr_400M	Classificiation of the closed contract.
clientscnt_100L	Number of applications with matching employer's phone and client's.
clientscnt_1022L	Number of clients sharing the same mobile phone.
clientscnt_1071L	Number of applications where the alternative phone number matches that of the client.
clientscnt_1130L	Number of applications where client's phone number matches the alternative phone contact.
clientscnt_136L	Number of applications associated with same email address as client's email.
clientscnt_157L	Number of clients whose employer has the same phone number as the client.
clientscnt_257L	Number of clients that share an alternative phone number with the applicant.
clientscnt_304L	Number of clients with the same phone number.
clientscnt_360L	Number of clients that have the same alternative phone number and employer's phone number.

clientscnt_493L	Number of clients with matching phone numbers for both the employer and the client.
clientscnt_533L	Number of clients with same client's and alternative's phone number
clientscnt_887L	Number of clients sharing the same employer's phone number.
clientscnt_946L	Number of clients with matching mobile and employer's number.
clientscnt12m_3712952L	Number of clients that have used the same mobile phone as the applicant in the past 12 months.
clientscnt3m_3712950L	Number of clients who have the same mobile phone number in the last 3 months.
clientscnt6m_3712949L	Total number of clients who have used the same mobile number in the last 6 months.
cntincpaycont9m_3716944L	Number of incoming payments in the past 9 months.
cntpmts24_3658933L	Number of months with any incoming payment in last 24 months.
collater_typofvalofguarant_298M	Collateral valuation type (active contract).
collater_typofvalofguarant_407M	Collateral valuation type (closed contract).
collater_valueofguarantee_1124L	Value of collateral for active contract.
collater_valueofguarantee_876L	Value of collateral for closed contract.
collaterals_typeofguarante_359M	Type of collateral that was used as a guarantee for a closed contract.
collaterals_typeofguarante_669M	Collateral type for the active contract.
commnoinclast6m_3546845L	Number of communications indicating low income in the last six months.
contaddr_district_15M	Zip code of a contact person's address.

contaddr_matchlist_1032L	Indicates whether the contact address is found in a code list.
contaddr_smempladdr_334L	Indicates whether the contact address is the same as the employment address.
contaddr_zipcode_807M	Zip code of contact address.
contractdate_551D	Contract date of the active contract
contractenddate_991D	End date of deposit contract.
contractmaturitydate_151D	End date of active contract.
contractssum_5085716L	Total sum of values of contracts retrieved from external credit bureau.
contractst_516M	Contract status.
contractst_545M	Contract status.
contractst_964M	Contract status of terminated credit contract.
contractsum_5085717L	Sum of other contract values.
contracttype_653M	Contract Type
conts_role_79M	Type of contact role of a person.
conts_type_509L	Person contact type in previous application.
creationdate_885D	Date when previous application was created.
credacc_actualbalance_314A	Actual balance on credit account.
credacc_cards_status_52L	Card status of the previous credit account.
credacc_credlmt_575A	Credit card credit limit provided for previous applications.
credacc_maxhisbal_375A	Maximal historical balance of previous credit account
credacc_minhisbal_90A	Minimum historical balance of previous credit accounts.
credacc_status_367L	Account status of previous credit applications.

credacc_transactions_402L	Number of transactions made with the previous credit account of the applicant.
credamount_590A	Loan amount or card limit of previous applications.
credamount_770A	Loan amount or credit card limit.
credlmt_1052A	Credit limit of an active loan.
credlmt_228A	Credit limit for closed loans.
credlmt_230A	Credit limit of the closed credit contracts from credit bureau.
credlmt_3940954A	Credit limit for active loan.
credlmt_935A	Credit limit for active loan.
credor_3940957M	Creditor's name
credquantity_1099L	Number of credits in credit bureau
credquantity_984L	Number of closed credits in credit bureau.
credtype_322L	Type of credit.
credtype_587L	Credit type of previous application.
currdebt_22A	Current debt amount of the client.
currdebt_94A	Previous application's current debt.
currdebtcredtyperange_828A	Current amount of debt of the applicant.
dateactivated_425D	Contract activation date of the applicant's previous application.
datefirstoffer_1144D	Date of first customer relationship management (CRM) offer.
datelastinstal40dpd_247D	Date of last instalment that was more than 40 days past due (DPD).
datelastunpaid_3546854D	Date of the last unpaid instalment.
dateofbirth_337D	Client's date of birth.
dateofbirth_342D	Client's date of birth.
dateofcredend_289D	End date of an active credit contract.

dateofcredend_353D	End date of a closed credit contract.
dateofcredstart_181D	Date when the credit contract was closed.
dateofcredstart_739D	Start date of a closed credit contract.
dateofrealrepmt_138D	Date of credit's closure (contract termination date).
days120_123L	Number of credit bureau queries for the last 120 days.
days180_256L	Number of credit bureau queries for last 180 days.
days30_165L	Number of credit bureau queries for the last 30 days.
days360_512L	Number of Credit Bureau queries for last 360 days.
days90_310L	Number of credit bureau queries for the last 90 days.
daysoverduetolerancedd_3976961L	Number of days that past after the due date (with tolerance).
debtoutstand_525A	Outstanding amount of existing contract.
debtoverdue_47A	Amount that is currently past due on a client's existing credit contract.
debtpastduevalue_732A	Amount of unpaid debt for existing contracts.
debtvalue_227A	Outstanding amount for existing debt contracts.
deductiondate_4917603D	Tax deduction date.
deferredmnthsnum_166L	Number of deferred months.
description_351M	Categorization of clients by credit bureau.
description_5085714M	Categorization of clients by credit bureau.
disbursedcredamount_1113A	Disbursed credit amount after consolidation.
disbursementtype_67L	Type of disbursement.
district_544M	District of the address used in the previous loan application.
downpmt_116A	Amount of downpayment.
downpmt_134A	Previous application downpayment amount.

dpd_550P	The number of days past due for active loans where a guarantee has been provided.
dpd_733P	Days past due (DPD) for guaranteed loans that were terminated according to credit bureau data.
dpdmax_139P	Maximal days past due for active contract.
dpdmax_757P	Maximum days past due for a closed contract.
dpdmax_851P	Maximal past due days for active contracts in the credit bureau.
dpdmaxdatemonth_442T	Max DPD occurrence month for terminated contracts from credit bureau data.
dpdmaxdatemonth_804T	Month when the maximum Day Past Due (DPD) occurred for active contracts on credit bureau's records.
dpdmaxdatemonth_89T	Month when maximum days past due occurred on the active contract with the credit bureau.
dpdmaxdateyear_596T	Year when maximum Days Past Due (DPD) occurred for the active contract.
dpdmaxdateyear_742T	Year of the maximum Days Past Due (DPD) on an active credit contract in the credit bureau.
dpdmaxdateyear_896T	Year of maximum Days Past Due of closed contract obtained from Credit Bureau.
dtlastpmt_581D	Date of last payment made by the applicant.
dtlastpmtallstes_3545839D	Date of the applicant's last payment.
dtlastpmtallstes_4499206D	Date of last payment made by the applicant.
education_1103M	Level of education of the client provided by external source.
education_1138M	Applicant's education level from their previous application.
education_88M	Education level of the client.
education_927M	Education level of the person.
eir_270L	Interest rate.

empl_employedfrom_271D	Start date of employment.
empl_employedtotal_800L	Employment length of a person.
empl_industry_691L	Employment Industry of the person.
empladdr_district_926M	District where the employer's address is located.
empladdr_zipcode_114M	Zipcode of employer's address.
employedfrom_700D	Employment start date from the previous application.
employername_160M	Employer's name.
empls_economicalst_849M	The economical status of the person (num_group1 - person, num_group2 - employment).
empls_employedfrom_796D	Start of employment (num_group1 - person, num_group2 - employment).
empls_employer_name_740M	Employer's name (num_group1 - person, num_group2 - employment).
equalitydataagreement_891L	Flag indicating sudden changes in client's social-demographic data (e.g. education, family status, housing type).
equalityempfrom_62L	Flag indicating a sudden change in the client's length of employment.
familystate_447L	Family state of the person.
familystate_726L	Family State in previous application of applicant.
financialinstitution_382M	Name of financial institution that is linked to a closed contract.
financialinstitution_591M	Financial institution name of the active contract.
firstclxcampaign_1125D	Date of the client's first campaign.
firstdatedue_489D	Date of the first due date.

firstnonzeroinstldate_307D	Date of first instalment in the previous application.
firstquarter_103L	Number of results obtained from credit bureau in the first quarter.
for3years_128L	Number of rejected applications in the past 3 years.
for3years_504L	Client's credit history data over the last three years.
for3years_584L	Number of cancellations in the last 3 years.
formonth_118L	Number of rejections in a month.
formonth_206L	Number of cancelations in the previous month.
formonth_535L	Credit history for the last month.
forquarter_1017L	Number of cancellations recorded in the credit bureau in the last quarter.
forquarter_462L	Number of credit applications that were rejected in the last quarter.
forquarter_634L	Credit history for the last quarter.
fortoday_1092L	Client's credit history for today.
forweek_1077L	Number of cancelations in the last week.
forweek_528L	Credit history for the last week.
forweek_601L	Number of rejected applications in the last week.
foryear_618L	Number of application rejections in the previous year.
foryear_818L	Number of cancelations that occurred in last year.
foryear_850L	Credit history for the last year.
fourthquarter_440L	Number of results in fourth quarter.
gender_992L	Gender of a person.
homephncnt_628L	Number of distinct home phones on client's application.
housetype_905L	House type of the person.
housingtype_772L	Type of housing of the person.
incometype_1044T	Type of income of the person
inittransactionamount_650A	Initial transaction amount of the credit application.

inittransactioncode_186L	Transaction type of the initial credit transaction.
inittransactioncode_279L	Type of the initial transaction made in the previous application of the client.
installmentamount_644A	Instalment amount of a closed and secured credit contract.
installmentamount_833A	Instalment amount for a secured and active contract in credit bureau.
instlamount_768A	Instalment amount for the active contract in credit bureau.
instlamount_852A	Instalment amount for closed contract.
instlamount_892A	Instalment amount for active credit contract.
interesteffectiverate_369L	Interest rate on active contract.
interestrate_311L	The interest rate of the active credit contract.
interestrate_508L	Interest rate for a closed contract in the credit bureau.
interestrategrace_34L	Interest rate during the grace period.
interestrateyearly_538L	Annual interest rate for existing contract obtained from credit bureau.
isbidproduct_1095L	Flag indicating if the product is a cross-sell.
isbidproduct_390L	Flag for determining if the product is a cross-sell in previous applications.
isbidproductrequest_292L	Flag indicating if the product is a cross-sell.
isdebitcard_527L	Previous application flag indicating if product being applied for is a debit card.
isdebitcard_729L	Flag indicating if the product is a debit card.
isreference_387L	Flag indicating whether the person is a reference contact.
language1_981M	The primary language of the person.
last180dayaveragebalance_704A	Average balance on debit card in the last 180 days.

last180dayturnover_1134A	Debit card's turnover within the last 180 days.
last30dayturnover_651A	Debit card turnover for the last 30 days.
lastactivateddate_801D	Contract activation date for previous applications.
lastapplicationdate_877D	Date of previous customer's application.
lastapprcommoditycat_1041M	Commodity category of the last loan applications made by the applicant.
lastapprcommoditytypec_5251766M	Commodity type of the last application.
lastapprcredamount_781A	Credit amount from the client's last application.
lastapprdate_640D	Date of approval on client's most recent previous application.
lastcancelreason_561M	Cancellation reason of the last application.
lastdelinqdate_224D	Date of the last delinquency occurrence.
lastdependentsnum_448L	Number of dependents in the client's last loan application.
lastotherinc_902A	Amount of other income reported by the client in their last application.
lastotherlnsexpense_631A	Monthly expenses on other loans from the last application.
lastrejectcommoditycat_161M	Category of commodity in the applicant's last rejected application.
lastrejectcommodtypec_5251769M	Commodity type of the last rejected application.
lastrejectcredamount_222A	Credit amount on last rejected application.
lastrejectdate_50D	Date of most recent rejected application by the applicant.

lastrejectreason_759M	Reason for rejection on the most recent rejected application.
lastrejectreasonclient_4145040M	Reason for the client's last loan rejection.
lastrepayingdate_696D	Date of the last payment made by the applicant.
lastst_736L	Status of the client's previous credit application.
lastupdate_1112D	Date of last update for an active contract from credit bureau.
lastupdate_260D	Last update date for the active contracts.
lastupdate_388D	Date of last update for a closed contract in the credit bureau.
maininc_215A	Client's primary income amount.
mainoccupationinc_384A	Amount of the main income of the client.
mainoccupationinc_437A	Client's main income amount in their previous application.
maritalst_385M	Marital status of the client.
maritalst_703L	Marital status of the client.
maritalst_893M	Marital status of the client
mastercontrelectronic_519L	Flag indicating the existence of the master contract for the client.
mastercontrexist_109L	Flag indicating whether or not the applicant has an existing master contract.
maxannuity_159A	Maximum annuity previously obtained by client.
maxannuity_4075009A	Maximal annuity offered to the client in the current application.
maxdbddpdlast1m_3658939P	Maximum number of days past due in the last month. A negative value indicates the number of days before the due date.
maxdbddpdtollast12m_3658940P	Maximum number of days past due in last 12 months. A negative value implies days before due date.

maxdbddpdtollast6m_4187119P	Maximum number of days past due in last 6 months. This predictor takes the value as a negative number when it represents days before due date.
maxdebt4_972A	Maximal principal debt of the client in the history older than 4 months.
maxdebtpduevalodued_3940955A	Days past due at the time of the maximum debt.
maxdpdfrom6mto36m_3546853P	Maximum Days Past Due (DPD) in the period ranging from 6 to 36 months.
maxdpdinstldate_3546855D	Date of instalment on which client was most days past due.
maxdpdinstlnum_3546846P	Instalment number of which client was most days past due.
maxdpdlast12m_727P	Maximum days past due in the past 12 months.
maxdpdlast24m_143P	Maximal days past due in the last 24 months.
maxdpdlast3m_392P	Maximum number of days past due in last 3 months.
maxdpdlast6m_474P	Maximum days past due in the last 6 months.
maxdpdlast9m_1059P	Maximum days past due in last 9 months.
maxdpdtolerance_374P	Maximum number of days past due (with tolerance).
maxdpdtolerance_577P	Maximum DPD with tolerance (on previous application/s).
maxinstallast24m_3658928A	Maximum instalment in the last 24 months
maxlnamtstart6m_4525199A	Maximum loan amount started in the last 6 months.
maxoutstandbalancel12m_4187113A	Maximum outstanding balance in the last 12 months.
maxpmtlast3m_4525190A	Maximum payment made by the client in the last 3 months.
mindbddpdlast24m_3658935P	Minimum days past due (or days before due) in last 24 months.

mindbdtollast24m_4525191P	Minimum days before due in last 24 months.
mobilephncnt_593L	Number of persons with the same mobile phone number.
monthlyinstlamount_332A	Monthly instalment amount for active contract.
monthlyinstlamount_674A	Monthly amount of instalment payment on a closed contract.
monthsannuity_845L	Monthly annuity amount for the applicant.
name_4527232M	Name of employer.
name_4917606M	Name of employer.
nominalrate_281L	Interest rate of the active contract.
nominalrate_498L	Interest rate for closed contract.
numactivecreds_622L	Number of active credits.
numactivecredschannel_414L	Number of active credits.
numactiverelcontr_750L	Number of active revolving credits.
numberofcontrsvalue_258L	Number of active contracts in credit bureau.
numberofcontrsvalue_358L	Number of closed credit contracts.
numberofinstls_229L	Number of instalments on closed contract.
numberofinstls_320L	Number of instalments of the active contract.
numberofinstls_810L	Number of instalments for the active contract.
numberofoutstandinstls_520L	Number of outstanding instalment for closed contract.
numberofoutstandinstls_59L	Number of outstanding instalments for the active contracts.
numberofoverdueinstlmax_1039L	Number of outstanding instalments for active contracts.

numberofoverdueinstlmax_1151L	Maximum number of past due installments for a closed contract.
numberofoverdueinstlmaxdat_148D	Date of maximum number of past due instalments for the closed contract.
numberofoverdueinstlmaxdat_641D	Date of maximum number of past due instalments for the active contract.
numberofoverdueinstls_725L	Maximum number of past due instalments for an active contract.
numberofoverdueinstls_834L	Number of past due instalments for a closed contract.
numberofqueries_373L	Number of queries to credit bureau.
numcontrs3months_479L	Number of contracts in last 3 months.
numincomingpmts_3546848L	Number of incoming payments.
numinstlallpaidearly3d_817L	Number of instalments paid at least 3 days prior to their due date.
numinstls_657L	Number of instalments.
numinstlsallpaid_934L	Number of paid instalments.
numinstlswithdpd10_728L	Number of instalments that were overdue for 10 or more days.
numinstlswithdpd5_4187116L	Number of instalments that were overdue by at least 5 days.
numinstlswithoutdpd_562L	Number of instalments that were not past due date.
numinstmatpaidtearly2d_4499204L	Number of instalments that have been paid more than 2 days before their due date.
numinstpaid_4499208L	Number of paid instalments.
numinstpaidearly_338L	Number of installments paid prior to the due date.

numinstpaidearly3d_3546850L	Number of instalments paid more than three days before the due date.
numinstpaidearly3dest_4493216L	Number of instalments that have been paid more than 3 days in advance of the due date.
numinstpaidearly5d_1087L	Number of instalments paid more than 5 days prior to the due date.
numinstpaidearly5dest_4493211L	Number of instalments that were paid more than 5 days before the due date.
numinstpaidearly5dobd_4499205L	Number of installments paid more than 5 days prior to the due date.
numinstpaidearlyest_4493214L	Number of instalments paid before the due date.
numinstpaidlastcontr_4325080L	Number of paid installments from the client's last contract.
numinstpaidlate1d_3546852L	Number of instalments paid more than 1 day past their due date.
numinstregularpaid_973L	Number of fully paid regular installments in the client's previous contracts.
numinstregularpaidest_4493210L	Number of fully paid regular installments on clients' previous contracts.
numinsttopaygr_769L	Number of unpaid instalments.
numinsttopaygrest_4493213L	Number of unpaid instalments.
numinstunpaidmax_3546851L	Maximum number of unpaid instalments.
numinstunpaidmaxest_4493212L	Maximum number of unpaid instalments.
numnotactivated_1143L	Number of non-activated credits.
numpmtchanneldd_318L	Number of previous loan contracts for the applicant that had direct debit as payment channel.
numrejects9m_859L	Number of credit applications that were rejected in the last 9 months.

opencred_647L	Number of active loans from the previous application.
openingdate_313D	Deposit account opening date.
openingdate_857D	Debit card opening date.
outstandingamount_354A	Outstanding amount for closed credit contract in credit bureau.
outstandingamount_362A	Active contract's outstanding amount.
outstandingdebt_522A	Amount of outstanding debt on the client's previous application.
overdueamount_31A	Past due amount for a closed contract.
overdueamount_659A	Past due amount for active contract.
overdueamountmax_155A	Maximal past due amount for active contract.
overdueamountmax_35A	Maximal past due amount for a closed contract.
overdueamountmax_950A	Maximal past due amount for active contract.
overdueamountmax2_14A	Maximal past due amount for an active contract.
overdueamountmax2_398A	Maximal overdue amount for a closed contract.
overdueamountmax2date_1002D	Date of maximal past due amount for a closed contract
overdueamountmax2date_1142D	Date of maximal past due amount for an active contract.
overdueamountmaxdatemonth_284T	Month when the maximum past due amount occurred for a closed contract.
overdueamountmaxdatemonth_365T	Month when maximum past due amount occurred for an active contract.
overdueamountmaxdatemonth_494T	Month when the maximum past due amount was recorded for an active contract with the credit bureau.

overdueamountmaxdateyear_2T	Year when the maximum past due amount occurred for active contracts.
overdueamountmaxdateyear_432T	Year when max past due amount occurred for active contract.
overdueamountmaxdateyear_994T	Year when maximum past due amount occurred for closed contract.
paytype_783L	Type of payment.
paytype1st_925L	Type of first payment of the client.
payvacationpostpone_4187118D	Date of last payment holiday instalment.
pctinstlsallpaidearl3d_427L	Percentage of installments paid at least 3 days prior to the due date.
pctinstlsallpaidlat10d_839L	Percentage of installments that were paid 10 or more days after the due date.
pctinstlsallpaidlate1d_3546856L	Percentage of installments that are paid 1 or more days after the due date.
pctinstlsallpaidlate4d_3546849L	Percentage of installments that were paid 4 or more days past their due date.
pctinstlsallpaidlate6d_3546844L	Percentage of installments that were paid 6 or more days past their due date.
periodicityofpmts_1102L	Frequency of instalments for a closed contract.
periodicityofpmts_837L	Frequency of instalments for an active contract.
periodicityofpmts_997L	Frequency of instalments for active credit contracts.
periodicityofpmts_997M	Frequency of instalments for active credit contracts.
personindex_1023L	Order of the person specified on the application form.
persontype_1072L	Person type.
persontype_792L	Person type.
pmtamount_36A	Tax deductions amount for credit bureau payments.

pmtaverage_3A	Average of tax deductions.
pmtaverage_4527227A	Average of tax deductions.
pmtaverage_4955615A	Average of tax deductions.
pmtcount_4527229L	Number of tax deductions.
pmtcount_4955617L	Number of tax deductions.
pmtcount_693L	Number of tax deductions.
pmtdaysoverdue_1135P	Number of days past due for existing contracts in the credit bureau.
pmtmethod_731M	Instalment payment method for existing contract in credit bureau.
pmtnum_254L	Total number of loan payments made by the client.
pmtnum_8L	Number of payments made for the previous application.
pmtnumpending_403L	Number of pending payments for active contract.
pmts_date_1107D	Payment date for an active contract according to credit bureau (num_group1 - contract, num_group2 - payment).
pmts_dpd_1073P	Days past due of the payment for the active contract (num_group1 - existing contract, num_group2 - payment).
pmts_dpd_303P	Days past due of the payment for terminated contract according to credit bureau (num_group1 - terminated contract, num_group2 - payment).
pmts_dpdvalue_108P	Value of past due payment for active contract (num_group1 - existing contract, num_group2 - payment).
pmts_month_158T	Month of payment for a closed contract (num_group1 - existing contract, num_group2 - payment).
pmts_month_706T	Month of payment for active contract (num_group1 - terminated contract, num_group2 - payment).
pmts_overdue_1140A	Overdue payment for an active contract (num_group1 - existing contract, num_group2 - payment).
pmts_overdue_1152A	Overdue payment for a closed contract (num_group1 - terminated contract, num_group2 - payment).
pmts_pmtsoverdue_635A	Active contract that has overdue payments (num_group1 - existing contract, num_group2 - payment).

pmts_year_1139T	Year of payment for an active contract (num_group1 - existing contract, num_group2 - payment).
pmts_year_507T	Payment year for a closed credit contract (num_group1 - terminated contract, num_group2 - payment).
pmtscount_423L	Number of tax deduction payments.
pmtssum_45A	Sum of tax deductions for the client.
posfpd10lastmonth_333P	Average FPD10 (Share of contracts with first installment past due more than 10 days) from point of sales that processed contract in the previous month.
posfpd30lastmonth_3976960P	Average FPD30 (Share of contracts with first installment past due more than 30 days) from point of sales that processed contract in the previous month.
posfstqpd30lastmonth_3976962P	Average FSTPD30 (share of contracts with first, second, or third installment past due more than 30 days) from point of sale that processed contract in the last month.
postype_4733339M	Type of point of sale.
previouscontdistrict_112M	Contact district of the client's previous approved application.
price_1097A	Credit price.
processingdate_168D	Date when the tax deduction is processed.
profession_152M	Profession of the client during their previous loan application.
prolongationcount_1120L	Count of prolongations on terminated contract according to credit bureau.
prolongationcount_599L	Count of active contract prolongations.
purposeofcred_426M	Purpose of credit for active contract.
purposeofcred_722M	Purpose of credit for active contracts.
purposeofcred_874M	Purpose of credit on a closed contract.
recorddate_4527225D	Date of tax deduction record.
refreshdate_3813885D	Date when the credit bureau's public sources have been last updated.

registaddr_district_1083M	District of person's registered address.
registaddr_zipcode_184M	Registered address's zip code of a person.
rejectreason_755M	Reason for previous application rejection.
rejectreasonclient_4145042M	Reason for rejection of the client's previous application.
relatedpersons_role_762T	Relationship type of a client's related person (num_group1 - person, num_group2 - related person).
relationshiptoclient_415T	Relationship to the client.
relationshiptoclient_642T	Relationship to the client.
remitter_829L	Flag indicating whether the client is a remitter.
requesttype_4525192L	Tax authority request type.
residualamount_1093A	Residual amount of closed guarantee contract.
residualamount_127A	Residual amount of active guarantee contract.
residualamount_3940956A	Residual amount for the active contract.
residualamount_488A	Residual amount of a closed contract.
residualamount_856A	Residual amount for the active contract.
responsedate_1012D	Tax authority's response date.
responsedate_4527233D	Tax authority's response date.
responsedate_4917613D	Tax authority's response date.
revolvingaccount_394A	Revolving account that was present in the applicant's previous application.
riskassesment_302T	Estimated probability that the client will default on their credit obligation within the next year.
riskassesment_940T	Estimate of client's creditworthiness.

role_1084L	Type of contact role.
role_993L	Person's role.
safeguarantyflag_411L	Flag indicating if client is using a flexible product with additional safeguard garanty.
score_940	Estimate of client's creditworthiness.
secondquarter_766L	Number of results in second quarter.
sellerplacecnt_915L	Number of sellerplaces where the same client's document was used.
sellerplacescnt_216L	Number of sellerplaces where the same client's mobile phone was used.
sex_738L	Gender of the client.
status_219L	Previous application status.
subjectrole_182M	Subject role in active credit contract.
subjectrole_326M	Subject role in active credit contract.
subjectrole_43M	Subject role in closed credit contract.
subjectrole_93M	Subject role in closed credit contract.
subjectroles_name_541M	Name of subject role in closed credit contract (num_group1 - terminated contract, num_group2 - subject roles).
subjectroles_name_838M	Name of subject role in active credit contract (num_group1 - existing contract, num_group2 - subject roles).
sumoutstandtotal_3546847A	Sum of total outstanding amount.
sumoutstandtotalest_4493215A	Sum of total outstanding amount.
tenor_203L	Number of instalments in the previous application.
thirdquarter_1082L	Number of results in third quarter.
totalamount_503A	Total amount of active secured credit for a client.
totalamount_6A	Total amount of closed contracts.
totalamount_881A	Total amount of secured credit from closed contracts.
totalamount_996A	Total amount of active contracts in the credit bureau.

totaldebt_9A	Total amount of debt.
totaldebtoverduevalue_178A	Total amount of past due debt on active contracts.
totaldebtoverduevalue_718A	Total overdue debt amount for closed credit contracts.
totaloutstanddebtvalue_39A	Total outstanding debt for active contracts in the credit bureau.
totaloutstanddebtvalue_668A	Total outstanding debt for the closed contracts in the credit bureau.
totalsettled_863A	Sum of all payments made by the client.
totinstallast1m_4525188A	Total amount of monthly instalments paid in the previous month.
twobodfilling_608L	Type of application process.
type_25L	Contact type of a person.
typesuite_864L	Persons accompanying the client during the loan application process.
validfrom_1069D	Date since the client has an active campaign.