Welcome to our ML Challenge

We are very excited to share this take home challenge as part of the recruitment process for Bankaya's Data Science team.

Please review these instructions carefully.

You will have 7 natural days to complete the challenge. Please make sure you agree on a starting date that works best for you with your recruiter.

Risk Model Challenge

Bankaya needs to differentiate between good customers and risky customers in order to make better decisions in terms of credit approvals. They ask you as a new Data Scientist to develop a model that would help them reach this goal and make better decisions on new customers applying to get their first loan to buy a smartphone.

The challenge is to create a model that would help them approve or reject clients, you are free to develop any kind of model you think might be able to resolve the task. Each step of the process should be justified and documented in the notebook or presentation that you send us. Be as thorough as you can be.

This part of the challenge is the main one and you can only submit your work solving this part.

The second part of the challenge is for extra points. As a data scientist you can give recommendations and proposals on how to use your model and how you would assign an interest rate to each client.

Datasets

1. **Main dataset:**

   This primary dataset contains our internal credit portafolio, providing data of the credits we have extended. Each row corresponds to an individual credit. All loans in these dataset are first-time loans. Additionally, the dataset includes the target variable.

Additional Insights:

This dataset also provides insights into the customer's history within our organization:

- **Previous Internal Applications:** Information about previous internal applications for different products (BNPL - Buy Now Pay Later, SF - Smartphone Finance).
- **Credit Bureau Inquiries:** Details about inquiries made by our company to the credit bureau.

main_dataset.parquet

2. **External credit bureau data:**

This dataset captures the credit history of the customers, with each row corresponding to a specific credit record. It's important to clarify that these credits are external to our company and represent credit activities of the customers with various financial entities. This report might have multiple credits per customer.

credit_reports.parquet.

Note: The relationship between datasets is one internal loan (main_dataset) to many external credit bureau loans (credit_reports), using customer_id column as the primary key.

Deliverable

We expect you to use a **Github repo** to upload your code and any relevant datasets or images to support your work.

Make sure any graphs can be visualized in your notebook without the need to re-run any cells. If requiered, you can attach graph images in a subfolder.

Additional Resources

Datasets description:

- customers with loan for the first time

    - 'LOAN_ID': id of loan
    - 'CUSTOMER_ID': internal id for customer
    - 'ACC_CREATION_DATETIME': date in which the account was created
    - 'APPLICATION_DATETIME': date in which the customer applied to a loan,
    - 'LOAN_ORIGINATION_DATETIME': date in which the loan was disbursed.
    - 'FINANCED_AMOUNT': amount financed.
    - 'LOAN_TERM_MONTHLY': time to pay tha loan in months.
    - 'LOAN_TYPE': type of credit.
    - 'max_days_late': maximum days late in the first 77 days since loan origination.
    - 'target': 1 if 34 days late or more in 77 days of contract, 0 otherwise.
    - 'account_to_application_days': days from account creation to application.
    - 'n_sf_apps': number of previous applications to smartphone financing product
    - 'first_app_date': first application date
    - 'last_app_date': last application date
    - 'n_bnpl_apps': number of previous applications to bnpl product
    - 'n_bnpl_approved_apps': number of previous bnpl applications approved
    - 'first_bnpl_app_date': first bnpl application date
    - 'last_bnpl_app_date': last bnpl application date
    - 'n_inquiries_l3m': inquiries to credit reports from external entities in the last 3 months.
    - 'n_inquiries_l6m': inquiries to credit reports from external entities in the last 6 months.

- credit reports

    - 'CUSTOMER_ID': internal id for customer
    - 'INQUIRY_TIME': date of report consultation.
    - 'LOAN_ID': id of loan
    - 'CDC_INQUIRY_ID': unique id of inquiry.
    - 'PREVENTION_KEY': prevention code.
    - 'CURRENCY': currency.
    - 'MAX_CREDIT': Maximum amount of credit used by the client.

- 'CREDIT_LIMIT': credit limit.
- 'PAYMENT_AMOUNT': amount of the next payment.
- 'UPDATE_DATE': date of the last time the info was updated.
- 'LOAN_OPENING_DATE': opening date of credit.
- 'LOAN_CLOSING_DATE': closing date of credit.
- 'WORST_DELAY_DATE': date of worst delay.
- 'REPORT_DATE': date of report.
- 'LAST_PURCHASE_DATE': date of last purchase.
- 'LAST_PAYMENT_DATE': date of last payment.
- 'PAYMENT_FREQUENCY': how frequent the client has to pay the credit.
- 'BUSINESS_TYPE': type of financial entity.
- 'CREDIT_TYPE': credit type
- 'ACCOUNT_TYPE': account type
- 'RESPONSABILITY_TYPE': type of responsability of the client.
- 'TOTAL_PAYMENTS': total number of payment through the entire credit.
- 'DELAYED_PAYMENTS': delayed payments.
- 'CURRENT_PAYMENT': current payment.
- 'WORST_DELAY': maximum number of due payments.
- 'TOTAL_REPORTED_PAYMENTS': total number of payments.
- 'CURRENT_BALANCE': current balance
- 'BALANCE_DUE': amount due.
- 'BALANCE_DUE_WORST_DELAY': worst amount due