1. Data Understanding & Initial Exploration:

- Load the dataset.

- Perform an initial examination to understand the structure, features, and distribution of data.

- Identify potential anomalies or missing values.

2. Data Cleaning & Pre-processing:

- Handle missing or erroneous data.

- Convert categorical variables to suitable numerical formats.

- Address any outliers or anomalies that might affect the model.

3. Feature Engineering & Data Transformation:

- Recognize and label transactions that are likely multiple payment attempts.

- Generate new features or transform existing ones based on insights to improve the model's predictive power.

4. Exploratory Data Analysis (EDA):

- Deep dive into data patterns, correlations, and trends.

- Visualize key insights using charts and graphs.

5. Model Development:

- Baseline Model: Start with a simple model to get an initial prediction accuracy.

- Advanced Predictive Modeling: Use more sophisticated algorithms to refine predictions, taking into account both the success rate and the associated transaction fees.

6. Model Evaluation & Fine-tuning:

- Measure the model's performance using appropriate metrics.

- Iterate and refine the model, testing different algorithms or parameters to enhance accuracy.

7. Feature Importance & Model Interpretation:

- Understand which features most influence the model's predictions.

- Provide insights into the reasons behind the model's decisions.

8. Deployment & Integration Proposal:

- Design a practical method or system (like a GUI) for the company to use the model in daily operations.

- This ensures that the model's benefits can be easily accessed by non-technical staff.

9. Feedback & Iteration:

- Based on feedback from the company or any new data, make necessary updates or refinements to the model and deployment method.

**Project Description:**

Task : Credit Card Routing for Online Purchase via Predictive Modelling

Case description:

This is your first day as a data scientist at one of the world’s largest retail companies. Already on your first day, you are invited to a meeting with important business stakeholders from the online payment department, who ask for your help: Over the past year they have encountered a high failure rate of online credit card payments. The company loses a lot of money due to failed transactions and customers become increasingly unsatisfied with the online shop.

Such online credit card payments are performed via so-called payment service providers, referred to as “PSPs” by the business stakeholders. Your company has contracts with four different PSPs and pays transaction fees for every single payment.

The current routing logic is manual and rule-based. Business decision-makers, however, hope that with predictive modelling and with your help, a smarter way of routing a PSP to a transaction is possible.

Project Aim:

Help the business to automate the credit card routing via a predictive model. Such a model should increase the payment success rate by finding the best possible PSP for each transaction and at the same time keep the transaction fees low.

Data Set:

The data set and all relevant information from the business side (name of PSPs, transaction fees) are given in a separate \*.zip folder, which is available in myCampus, under the section Case Study.

Task Description:

The task consists of both coding and conceptual steps. Here is a list of tasks, which should be included in your final document:

● Structure the project via the CRISP-DM or Team DS methodologies and give a recommendation of how a git repository for the project could look like. Note that you do not have to structure your final code according to your git-repository proposal.

● Assess the quality of the provided data set. Prepare and visualize your findings of the initial data analysis in order that business stakeholders can understand them in a clear and easy way.

● Provide a baseline model as well as an accurate predictive model, which fulfills business requirements, i.e. increase credit card success rate and keep fees low.

● In order that the business places confidence in your model, discuss the importance of the individual features and make the results of the model interpretable. Moreover, a sophisticated error analysis is very important for the business to understand the drawbacks of your approach.

● In the last step of the project, give a proposal of how your model could be used by the business in everyday work, for instance, via a graphical user interface (GUI).

● Finally, do not forget to attach the code to the final submission document.

info:1

name Fee on successful transactions Fee on failed transactions

Moneycard 5 Euro 2 Euro

Goldcard 10 Euro 5 Euro

UK\_Card 3 Euro 1 Euro

Simplecard 1 Euro 0,5 Euro

info2:

tmsp: timestamp of transaction

• country: country of transaction

• amount: transaction amount

• success: is 1 if payment is successful

• PSP: name of payments service provider

• 3D\_secured: is 1 if customer is 3D identified (i.e. more secure online credit card payments)

• card: credit card provider (Master, Visa, Diners)

info3:

Many transactions fail at the first try. Therefore, customers try several times to transfer the money. If two transactions are within one minute, with the same amount of money and from the same country, it is (for a decent number of tries) safe to assume that they are payment attempts of the same purchase. Consider this possibility of several payment attempts of the same purchase in your machine learning model