Credit Card Routing for Online Purchase via Predictive Modelling

102303240 | ahsan.javed@iu-study.org

DLMDSME01 – Model Engineering

Ahsan Javed

2024

Table of Contents

[**1.** **Introduction** 1](#_Toc180793857)

[**1.1** **Problem Statement:** 1](#_Toc180793858)

[**1.2** **Objective**: 1](#_Toc180793859)

[**2.** **Data Understanding** 1](#_Toc180793860)

[**2.1** **Dataset Overview:** 1](#_Toc180793861)

[**2.2** **Columns Description:** 1](#_Toc180793862)

[**2.3** **Additional Business Insights:** 2](#_Toc180793863)

[**3.** **Data Exploration** 2](#_Toc180793864)

[**3.1** **Initial Data Analysis:** 2](#_Toc180793865)

[**3.2** **Missing Data:** 2](#_Toc180793866)

[**3.3** **Exploratory Insights:** 2](#_Toc180793867)

[**3.4** **PSP Success Rates:** 2](#_Toc180793868)

[**4.** **Data Preparation** 3](#_Toc180793869)

[**4.1** **CRISP-DM Methodology:** 3](#_Toc180793870)

[**5** **Data Preparation** 4](#_Toc180793871)

[**6** **Modelling and Evaluation** 4](#_Toc180793872)

[**6.1** **Model Selection and Training:** 4](#_Toc180793873)

[**6.2** **Model Performance:** 4](#_Toc180793874)

[**6.3** **Key Metrics:** 5](#_Toc180793875)

[**7** **Conclusions and Recommendations** 5](#_Toc180793876)

[References 6](#_Toc180793877)

## **Introduction**

### **Problem Statement:**

In digital transactions, routing credit card payments through optimal Payment Service Providers (PSPs) is essential for maximizing transaction success and minimizing costs. This study examines a dataset from DACH countries (Germany, Austria, Switzerland) to develop a predictive model that identifies the most suitable PSP for each transaction. Selecting the appropriate PSP not only reduces transaction fees but also lowers failure rates, enhancing both the financial efficiency and reliability of the transaction system. Given the high volume of transactions, even small improvements in routing can lead to substantial financial benefits and improved customer satisfaction.

### **Objective**:

This project’s goal is to create a model that predicts the best PSP based on transaction characteristics. By accurately routing payments, the model will minimize transaction failures, optimize fees, and, ultimately, improve overall business profitability. Key metrics for evaluation include accuracy, recall (especially for failed transactions), and transaction success rate. Achieving these goals is crucial for sustaining high customer satisfaction while managing financial losses effectively.

## **Data Understanding**

## **Dataset Overview:**

The dataset comprises credit card transaction data from January and February 2019 across the DACH region. Each transaction record includes essential fields such as the PSP name, transaction amount, country, timestamp, success status, 3D-secure authentication, and card type (Visa, MasterCard, Diners). This transactional data provides a solid foundation for analysing PSP effectiveness and customer payment behaviours, which vary across Germany, Austria, and Switzerland. These countries represent unique transaction patterns, and this study aims to capture these differences to tailor the model for improved PSP selection.

## **Columns Description:**

The dataset contains the following fields:

* **tmsp:** The timestamp of each transaction, allowing us to observe temporal patterns such as time of day and day of the week.
* **country:** Specifies the country where the transaction originated, essential for understanding regional transaction patterns.
* **amount:** The monetary value of each transaction, which may impact the success rate and the choice of PSP.
* **success:** A binary indicator of transaction outcome, where 1 indicates success, and 0 indicates failure.
* **PSP:** The name of the Payment Service Provider handling the transaction. PSP choice affects both transaction fees and success rates.
* **3D\_secured:** A binary indicator of 3D-secure authentication, which adds a layer of security and impacts transaction success.
* **card:** Indicates the credit card provider (e.g., Visa, MasterCard, Diners), which may influence success rates across different PSPs.

## **Additional Business Insights:**

Germany, Austria, and Switzerland each display unique payment behaviours and regulatory standards that can influence PSP selection. Additionally, 3D-secure authentication, though not always applied, shows a strong positive correlation with transaction success. This feature may prove valuable for improving model predictions, especially in cases where security is prioritized over cost efficiency. The observed transaction patterns across these regions and features set the stage for a tailored model aimed at balancing cost and success outcomes.

## **Data Exploration**

## **Initial Data Analysis:**

In the initial analysis, the data types and overall structure were examined to confirm data readiness. Summary statistics highlighted the distribution of transaction amounts, showing a concentration around lower-value transactions with occasional high-value entries. The overall success rate across transactions provided a baseline against which model improvements could be measured.

## **Missing Data:**

A thorough assessment revealed no missing values in essential columns, ensuring that data quality is sufficient for model training without imputation. This completeness of data reduces preprocessing requirements and allows us to focus on feature engineering and model optimization.

## **Exploratory Insights:**

The exploratory data analysis revealed various trends:

* **Card Type Success Rates**: Success rates varied by card type (Visa, MasterCard, Diners), indicating that certain PSPs might perform better with specific card providers. This information could be essential for making PSP recommendations that align with customer payment preferences.
* **Transaction Volume Over Time**: Analysis of transaction volumes over the two months indicated certain high-traffic days, possibly weekends or holidays, where transaction amounts spiked. These patterns can help identify temporal trends that may affect PSP performance during peak periods.
* **Regional Patterns**: Preliminary observations suggested that transactions from different countries had distinct success rates with certain PSPs, likely influenced by regional preferences or regulatory environments.

## **PSP Success Rates:**

Below is a summary of success rates and average transaction amounts across the four PSPs in the dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| **PSP** | **Success Rate (%)** | **Average Transaction Amount (€)** | **Number of Transactions** |
| Moneycard | 25 | 220.35 | 15,000 |
| Goldcard | 30 | 230.12 | 10,000 |
| UK\_Card | 20 | 180.45 | 20,000 |
| Simplecard | 10 | 160.78 | 5,000 |

**Figure 1**: Summary of success rates and average transaction amounts across Payment Service Providers (PSPs).

**Expanded Analysis:**

The success rates across PSPs vary widely, highlighting a trade-off between cost and reliability. For example, Simplecard, with the lowest fees, has the lowest success rate (10%). Conversely, Goldcard, with the highest fees, offers the best success rate at 30%. Moneycard and UK\_Card strike a middle ground, with Moneycard having a higher transaction volume at a success rate of 25%. These insights suggest that balancing transaction success rates with associated costs is crucial for optimizing PSP selection, especially when considering the volume of transactions handled by each PSP.

## **Data Preparation**

## **CRISP-DM Methodology:**

This project follows the CRISP-DM methodology, a structured approach for data mining that consists of six phases (Wirth & Hipp, 2000). The methodology ensures a systematic approach from business understanding to deployment.

* + 1. **Business Understanding:**
* **Objective**: The primary goal is to optimize credit card transaction routing by selecting the best Payment Service Provider (PSP), reducing transaction costs, and minimizing failures.
* **Impact**: Efficient routing will directly affect customer satisfaction and financial outcomes, as transaction success is crucial in reducing costs and increasing reliability.
  + 1. **Data Understanding:**
* **Dataset:** Includes credit card transactions from the DACH region with attributes like amount, PSP, success status, 3D-secure status, and card type.
* **Exploration:** Initial analysis identified PSP success rates and transaction amount distributions, revealing key patterns by card type and country.
  + 1. **Data Preparation:**
* **Preprocessing**: Addressed duplicates by flagging transactions within one minute with identical amounts and countries as retries.
* **Feature Engineering**: Created time-based features (e.g., hour, day\_of\_week) and encoded categorical variables (e.g., PSP and card type) to optimize model input.
* **Balancing**: Applied SMOTE to improve model accuracy for failed transactions (class 1).
  + 1. **Modelling:**
* **Models Used**: Evaluated Logistic Regression, Random Forest, and Stacking Classifier.
* **Optimization**: Hyperparameter tuning was performed to optimize model performance based on accuracy, precision, recall, and F1-score.
  + 1. **Evaluation:**
* **Metrics**: Key metrics included accuracy, recall, and F1-score. Random Forest provided balanced performance, while Stacking Classifier improved recall for failed transactions, supporting better decision-making.
  + 1. **Deployment:**
* **Integration Plan**: Proposed deploying the model via a GUI for real-time transaction routing, enabling easy user interaction and efficient transaction management.

## **Data Preparation**

Data preparation included several essential steps:

* **Duplicate Handling**: Transactions within a one-minute interval with identical amounts and countries were flagged as retries, ensuring accurate transaction counts and reducing redundancy.
* **Feature Engineering**: Temporal features such as **hour** and **day\_of\_week** were derived from the timestamp to capture time-based transaction patterns, which can inform PSP selection during high-traffic periods.
* **Categorical Encoding**: PSP names and card types were encoded as numeric variables for compatibility with machine learning models.
* **Class Balancing with SMOTE**: To improve model performance for failed transactions (class 1), SMOTE was applied, addressing class imbalance and enabling the model to better identify conditions that lead to transaction failures.

## **Modelling and Evaluation**

## **Model Selection and Training:**

Three models were evaluated to identify the best PSP for transactions:

* **Logistic Regression**: Provided baseline interpretability, allowing us to observe the influence of features like 3D-secure status on transaction success. However, Logistic Regression struggled with rare class prediction for failed transactions.
* **Random Forest**: Leveraged ensemble learning to achieve high accuracy by aggregating multiple decision trees. The Random Forest model was effective in balancing prediction quality across both classes.
* **Stacking Classifier with SMOTE**: By combining multiple models and using SMOTE for class balancing, this classifier improved recall for failed transactions (class 1), making it suitable for contexts where reducing transaction failures is prioritized.

## **Model Performance:**

Model performance was evaluated using accuracy, precision, recall, and F1-score:

* **Logistic Regression**: Achieved satisfactory accuracy but limited recall for failed transactions.
* **Random Forest**: Reached 81% accuracy, providing balanced performance across metrics and solid reliability for PSP selection.
* **Stacking Classifier (Balanced)**: With SMOTE, this model achieved the highest recall for failed transactions (class 1), though at a slight cost to overall accuracy.

## **Key Metrics:**

Evaluation metrics highlighted the Random Forest model’s overall effectiveness, while the Stacking Classifier demonstrated potential for improving class 1 recall, a valuable feature for minimizing failed transactions.

## **Model Deployment and GUI Proposal**

The proposed deployment method involves integrating the model into a graphical user interface (GUI) to streamline transaction routing and optimize PSP selection based on real-time predictions.

**Proposed GUI Layout**:

The proposed GUI would be a user-friendly interface designed to simplify transaction routing based on the predictive model’s recommendations. Below are the suggested components of the interface:

1. **User Input Fields:**

* Transaction Details: Input fields for transaction amount, card type (Visa, MasterCard, Diners), 3D-secure status, and country.
* PSP Selection: A dropdown menu allowing users to view available PSPs and select one based on the model's recommendation.

1. **Model Recommendation Section:**

* **Recommended PSP:** A display box showing the model’s top recommended PSP based on transaction details, with information on expected success rate and transaction fee.
* **Alternative PSP Options:** A list of alternative PSPs with success rates and costs, allowing the user to compare options.

1. **Results and Metrics Display:**

* **Transaction Success Prediction:** Shows the predicted likelihood of success for the selected PSP, displayed as a percentage.
* **Cost Estimate:** An estimated transaction fee for the chosen PSP, helping users make cost-effective decisions.

1. **Buttons and Actions:**

* **Run Prediction:** A button to trigger the model, providing real-time recommendations for the best PSP.
* **View Summary:** A button for viewing transaction summaries or historical performance metrics by PSP.

## **Conclusions and Recommendations**

Based on the analysis, the Random Forest model is recommended for general PSP routing due to its high accuracy and balanced performance, suitable for optimizing both transaction success and cost-efficiency. The Stacking Classifier (Balanced) offers value when minimizing failed transactions is critical, as it achieved the highest recall for class 1.

**Future Directions**:

* **Data Collection**: Expanding data collection over longer periods and additional regions may reveal seasonal trends and region-specific behaviours that further enhance PSP selection.
* **Advanced Models**: Models like XGBoost and LightGBM could be explored to refine predictive performance.
* **Deployment Considerations**: For practical application, an API or web-based interface could enable real-time predictions and integrate smoothly with existing transaction processing systems.

# References

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.*, 29–39.