|  |  |
| --- | --- |
|  | FleetCor Project Report |
|  |  |
|  | Group 1  Database Management for Business Analytics  12/12/24 |

# **Cross-Sell Model Development**

## Group 1: FleetCor Project 1

### Department of Economics and Business Analytics, University of New Haven

### BANL6430-01: Database Management for Business Analytics

#### Dr. Pindaro Demertzoglou

#### December 12th, 2024

## **Executive Summary**

FleetCor Cross-Sell Strategy Enhancement Project aimed at enhancing FleetCor’s strategy for cross selling their Universal Card. The fundamental focus was on filtering the existing eligibility model to better identify and target high-value customers for inclusion (“swap-in”) while filtering out underperforming ones (“swap-out”). The goal was to attain a more profitable and risk-mitigated customer base, thereby driving overall satisfaction and profitability through data-informed decisions.

The final implementation of a Random Forest model significantly improved predictive accuracy of the overall model by achieving a remarkable 99.33%. It highlighted key factors such as credit limit and Vantage Score as the most essential variables in determining customer eligibility. The final strategy distinguishing eligible from ineligible customers more effectively, this model provided accurate insights, leading to optimized cross-sell decisions and better financial outcomes for FleetCor.

The Random Forest model was chosen for its strong predictive capabilities and ability to manage complex, non-linear relationships among variables which were very critical in nature. When the random forest model compared with logistic regression, it proved to be more effective, as evidenced by an Area Under the Curve (AUC) score of 1. This robust model offers FleetCor a dependable solution for accurately predicting customer eligibility and elevating the effectiveness of their cross-sell strategy.

The project reflected some of the key factors critical to determining customer eligibility for the cross-sell strategy. Firstly, **Credit Limit** emerged as the most influential factor, reflecting a customer’s financial capacity to manage additional credit responsibly. Besides, **Vantage Score** widely accepted measure of creditworthiness, provided insight into a customer’s reliability and risk level. Lastly, **Fuel and Non-Fuel Spending** patterns highlighted customer engagement and their potential for generating value through increased spending. These factors were selected after a thorough analysis of their strong predictive power, as demonstrated by the Random Forest model’s feature importance metrics. The model consistently indicated that customers with higher credit limits, strong Vantage Scores, and balanced spending patterns were the best candidates for the Universal Card, aligning directly with the company’s goals of increasing profitability and minimizing risk.

## **Dashboard**

The project also involved the development of an interactive dashboard using Tableau. This dashboard is designed to monitor key performance metrics related to the cross-sell strategy, such as delinquency rates, credit utilization, and customer segmentation. It integrates seamlessly with FleetCor’s CRM system, allowing for real-time updates and actionable insights. The visuals, including bar charts and heatmaps, were chosen for their ability to convey critical information clearly and support better decision-making processes.

## **Data Integration Work Performed**

* **Data Sourcing and Preparation:**

FleetCor provided 10 Excel files in a zip format, containing data on customer transactions, payment history, and credit-related metrics. Using **R**, the data was extracted, cleaned, and prepared for analysis, where key tasks included identifying missing values, standardizing inconsistent formats, and removing duplicate entries. These steps ensured the dataset was accurate and consistent for subsequent modeling efforts.

* **Data Consolidation:**

The cleaned files were merged into a single dataset to facilitate efficient analysis. This involved harmonizing column names, ensuring data type consistency, and resolving discrepancies across the individual files.

* **Data Modeling:**

A high-level dimensional model was conceptualized to aid analysis and reporting.

The star schema design included:

* **Fact Table:**Containing key transactional data, such as credit usage, payment history, and risk indicators.
* **Dimension Tables:** Capturing descriptive attributes, such as customer profiles, product categories, and time periods. This model allowed for efficient querying, slicing, and dicing of data to extract actionable insights.

## **Data Integration Steps**

1. **Export and Save the Model:**
   * After we train the Random Forest model in R, we save it as a serialized file (e.g., .rds) to ensure easy reuse. This step ensures the trained model can be loaded directly into the production environment without retraining.
2. **Model Deployment in CRM:**
   * Incorporate the saved model into FleetCor's CRM system using APIs or a backend integration layer.
   * For example:
     + Use **R APIs (e.g., plumber)** to expose the model as a web service.
     + Alternatively, convert the model to a format compatible with the CRM’s analytics engine (if supported).
3. **Real-Time Scoring:**
   * Set up a mechanism where new customer data is fed into the model in real time for scoring. This ensures immediate predictions for customer eligibility (swap-in or swap-out decisions).
   * Example: Customer data updates in the CRM trigger the model to calculate predictions.
4. **Data Flow Setup:**
   * Establish an ETL pipeline to extract relevant data fields (e.g., Credit Limit, Vantage Score, Spending Metrics) from the CRM, preprocess them as required by the model (normalization, encoding), and pass them as inputs for scoring.
5. **Integration with Dashboards:**
   * Use tools like Tableau for visualizing model outputs directly from the CRM. This can help stakeholders monitor eligibility results, risk scores, and segmentation in real time.
6. **User-Friendly Interface:**
   * Design an intuitive interface within the CRM for non-technical users to view predictions, explanations of eligibility, and feature importance (e.g., how much Credit Limit or Vantage Score influenced the decision).
7. **Scheduled Retraining:**
   * Automate periodic retraining of the model with updated data from the CRM. This ensures the model adapts to evolving customer behaviors and maintains accuracy over time.

## **Models and Tools Used**

## **Random Forest Model:**

## The Random Forest algorithm was selected for its superior performance in handling non-linear relationships and complex interactions between variables. We have evaluated other models like Logistic Regression, the Random Forest was chosen based on its exceptional predictive accuracy (AUC: 1) and ability to provide insights into feature importance. The key predictors identified included Credit Limit, Vantage Score, and Spending Patterns.

**Rationale for Selection:**

* **Accuracy:** Random Forest achieved 99.33% accuracy, minimizing false positives and negatives.
* **Feature Importance Analysis:** It highlighted influential factors such as Credit Limit, enabling targeted decision-making.
* **Scalability:** Its ability to process large datasets made it ideal for FleetCor's needs.

**Tools Utilized:**

* **R:** Used for data cleaning, preparation, and model development. Libraries like caret and random forest facilitated seamless implementation.
* **Tableau:** Interactive dashboards were created to visualize model outputs, such as customer segmentation and eligibility scores.
* **Excel:** Used for initial data exploration and sensitivity analysis, allowing quick checks and validations.

## **Group Names**

* **Group Members:**
  + Ms. Khyati Chauhan
  + Ms. Ponvarshini Arumugam
  + Mr. Varshith Bandi
* **Total Members:** 3

**Appendix:**

**R Code Outputs During Model Comparison of Logistics Regression and Random Forest:**

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

A graph with numbers and a red line

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph with dots on it

Description automatically generated

A green and orange pie chart

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

**Dashboard on Tableau**

A screenshot of a graph

Description automatically generated

**SQL Code for Data Modelling**

**a. Customer Dimension Table**

sql

Copy code

CREATE TABLE DimCustomer (

CustomerID INT PRIMARY KEY,

CustomerName VARCHAR(100),

Segment VARCHAR(50),

VantageScore INT,

City VARCHAR(100),

State VARCHAR(100),

ZipCode VARCHAR(10)

);

**b. Product Dimension Table**

sql

Copy code

CREATE TABLE DimProduct (

ProductID INT PRIMARY KEY,

ProductCategory VARCHAR(50),

ProductName VARCHAR(100)

);

**c. Time Dimension Table**

sql

Copy code

CREATE TABLE DimTime (

TimeID INT PRIMARY KEY,

Year INT,

Month INT,

Day INT,

Quarter INT,

YearMonth VARCHAR(7)

);

**2. Create the Fact Table**

sql

Copy code

CREATE TABLE FactTransactions (

TransactionID INT PRIMARY KEY,

CustomerID INT,

ProductID INT,

TimeID INT,

CreditUsage DECIMAL(10, 2),

PaymentHistory DECIMAL(10, 2),

RiskIndicator VARCHAR(50),

Revenue DECIMAL(10, 2),

FOREIGN KEY (CustomerID) REFERENCES DimCustomer(CustomerID),

FOREIGN KEY (ProductID) REFERENCES DimProduct(ProductID),

FOREIGN KEY (TimeID) REFERENCES DimTime(TimeID)

);

**Insert Sample Data**

You can populate these tables with sample data to test their functionality.

**1. Insert Data into Dimension Tables**

**a. Insert Customers**

INSERT INTO DimCustomer (CustomerID, CustomerName, Segment, VantageScore, City, State, ZipCode)

VALUES

(1, 'John Doe', 'High Risk', 400, 'New York', 'NY', '10001'),

(2, 'Jane Smith', 'Medium Risk', 650, 'San Francisco', 'CA', '94105'),

(3, 'Alice Johnson', 'Low Risk', 750, 'Austin', 'TX', '73301');

**b. Insert Products**

INSERT INTO DimProduct (ProductID, ProductCategory, ProductName)

VALUES

(101, 'Fuel', 'Diesel'),

(102, 'Fuel', 'Petrol'),

(103, 'Non-Fuel', 'Lubricant');

**c. Insert Time Periods**

INSERT INTO DimTime (TimeID, Year, Month, Day, Quarter, YearMonth)

VALUES

(1, 2023, 1, 1, 1, '2023-01'),

(2, 2023, 2, 1, 1, '2023-02'),

(3, 2023, 3, 1, 1, '2023-03');

**2. Insert Data into the Fact Table**

INSERT INTO FactTransactions (TransactionID, CustomerID, ProductID, TimeID, CreditUsage, PaymentHistory, RiskIndicator, Revenue)

VALUES

(1001, 1, 101, 1, 500.00, 250.00, 'High', 1000.00),

(1002, 2, 102, 2, 300.00, 150.00, 'Medium', 750.00),

(1003, 3, 103, 3, 100.00, 50.00, 'Low', 500.00);

**Explanation**

1. **Fact Table (FactTransactions)**:
   * Contains transactional data like credit usage, payment history, risk indicators, and revenue.
   * Includes foreign keys (CustomerID, ProductID, TimeID) linking to the dimension tables.
2. **Dimension Tables**:
   * DimCustomer: Stores customer-related attributes (e.g., segment, vantage score, location).
   * DimProduct: Stores product-related attributes (e.g., category, name).
   * DimTime: Stores time-related attributes (e.g., year, month, quarter).