# **Customer Transaction Prediction**

**Introduction:**

In today’s data-driven business environment, predicting customer behavior is vital for targeted marketing, customer engagement, and resource optimization. This project aims to build a machine learning model that can predict whether a customer will make a transaction in the future, using anonymized numerical features. The classification model is intended to help businesses identify high-potential customers and proactively engage with them.

**Abstract:**

The objective of this project is to predict customer transactions using a large, structured dataset provided for modeling. The dataset consists of **200,000 observations** with **202 columns**, including:

* ID\_code: Unique identifier for each record
* target: Binary label (1 for transaction, 0 for no transaction)
* var\_0 to var\_199: 200 anonymized numeric features

We employed various preprocessing and modeling techniques, including SMOTE **for class balancing**, and used models such as Logistic Regression and XGBoost to build a robust binary classifier. Evaluation metrics like accuracy, precision, recall, F1-score were used to assess performance.

**Tools & Libraries Used**

* **Python** – Programming language
* **Jupyter Notebook** – Development environment
* **Pandas** – Data handling
* **NumPy** – Numerical operations
* **Seaborn & Matplotlib** – Data visualization
* **Skimpy** – Exploratory data summary
* **Scikit-learn** – Machine learning models and metrics
* **XGBoost** – Gradient boosting classifier
* **SMOTE (imblearn)** – Synthetic oversampling for class imbalance

Steps Involved:

1. **Data Loading**
   * Imported the dataset with 200,000 rows and 202 columns.
2. **Exploratory Data Analysis (EDA)**
   * Checked data structure, column types, and summary statistics using .info(), .describe(), and skim().
3. **Class Imbalance Handling**
   * Used SMOTE to balance the dataset as the target variable was imbalanced.
4. **Preprocessing**
   * Scaled the features using StandardScaler and split the data into training and testing sets.
5. **Model Building**
   * Trained models including Logistic Regression and XGBoost.
6. **Model Evaluation**
   * Evaluated models using Accuracy, Precision, Recall, F1-score, and ROC-AUC.
7. **Visualization**
   * Visualized class distribution and performance metrics using Matplotlib and Seaborn.

**Conclusion:**

The project successfully demonstrated how machine learning can be leveraged to predict customer transactions using anonymized behavioral data. The XGBoost Classifier yielded strong predictive power with higher recall proving suitable for imbalanced classification tasks.