New Jersey Institute of Technology

AI-Driven Financial Advisory and Credit Risk Prediction: A Machine Learning Approach to Personal Finance Management

Jyothsna Karuparthi (jk773)
Jasleen Kaur (jl2448)
Afsheen Khan (ak2788)
Bhanu Sai Nikhil Vinnakota (bv258)

Submitted to: Dr. Arashdeep Kaur

Contents

Introduction

Managing personal finances effectively has become increasingly challenging in today's fast-paced and financially demanding world. Many individuals struggle with credit card debt, unexpected expenses, and a lack of proper financial planning, leading to financial instability. Traditional financial tools often focus on tracking expenditures but fail to provide predictive insights or proactive strategies to help users make informed decisions. Without early risk detection and personalized financial advice, people are left vulnerable to poor spending habits, missed payments, and eventual credit defaults, which can have long-term financial consequences.

To address this growing concern, this project introduces an AI-powered personal finance coach designed to revolutionize the way individuals manage their money. Unlike conventional budgeting applications, this system harnesses the power of machine learning to analyze spending patterns, assess credit risk, and generate personalized financial recommendations in real time. By utilizing a dataset of 30,000 credit card users, the system predicts the likelihood of default with an accuracy exceeding 77%, enabling early intervention and improved financial decision-making. Future enhancements will focus on refining machine learning models to achieve even higher accuracy through advanced feature engineering and model optimization.

What sets this system apart is its ability to go beyond simple predictions. It integrates real-time financial monitoring, risk assessment models, and interactive dashboards to provide users with clear and actionable insights. Instead of relying on static reports, the AI dynamically adapts to user behavior, offering customized savings plans, future expense forecasts, and automated advisory recommendations that align with an individual's financial goals.

The system enhances user experience with interactive visualizations and deep learning, making financial data intuitive and engaging. API integrations for stock data, currency rates, and banking transactions provide real-time insights, while secure cloud storage and predictive analytics ensure seamless, controlled access to financial data.

By integrating artificial intelligence into personal finance, this project aims to empower individuals with data-driven financial guidance, bridging the gap between traditional financial tools and intelligent, automated decision-making. As financial technology continues to advance, solutions like this will play a vital role in shaping a future where individuals can manage their finances with greater confidence, clarity, and efficiency.

Abstract

Managing personal finances, particularly credit card debt, has become increasingly complex in today's fast-paced financial environment. Traditional financial management tools focus on expense tracking but fail to provide predictive insights or proactive advisory support. This project presents an AI-powered personal finance coach, a next-generation solution that leverages machine learning to predict credit card default risks with over 77% accuracy while offering personalized financial recommendations tailored to user spending behaviors.

Utilizing a dataset of 30,000 credit card users, the system applies advanced predictive modeling and real-time analytics to assess financial health dynamically. Unlike conventional budgeting applications, this AI-driven solution continuously adapts to user behavior, generating real-time risk assessments, interactive visualizations, and automated savings strategies that promote long-term financial stability. By integrating deep learning techniques, future expense forecasting, and financial advisory automation, the system goes beyond simple predictions, transforming financial management into an intelligent, data-driven experience.

To enhance user engagement, the platform features an intuitive interface, interactive dashboards, and API integrations for live financial data, including stock market trends, currency fluctuations, and banking transactions. With secure cloud storage, AI-powered insights, and a user-centric design, this project represents a significant advancement in financial technology, empowering individuals to make informed, strategic financial decisions with confidence.

Objective

This project addresses financial mismanagement and credit risk prediction through machine learning and AI-driven financial advisory. Many individuals struggle with credit card debt due to a lack of proactive insights, leading to defaults and instability. Existing tools focus on expense tracking but lack predictive analytics and personalized recommendations. This project bridges that gap by developing an AI-powered personal finance coach that provides real-time risk assessment, tailored financial advice, and interactive financial analytics.

6.0.1 Risk Assessment System Development

Machine learning models will predict credit card defaults with over 77% accuracy. An automated risk scoring mechanism will analyze user spending behavior, while an early warning system will detect financial distress signals. Real-time risk monitoring will help users avoid financial pitfalls.

6.0.2 Financial Advisory Implementation

An AI-driven recommendation engine will provide personalized financial guidance. Spending pattern analysis will identify unhealthy habits, and automated savings plans will align with individual financial goals. Debt management strategies will assist in reducing financial burdens, while real-time financial insights will improve decision-making and literacy.

6.0.3 Interactive Analytics Platform

A user-friendly dashboard will streamline financial management. Advanced visualization tools will display spending patterns and risk levels, while a dynamic reporting system will generate real-time financial summaries. Continuous financial data monitoring will ensure up-to-date insights, and trend analysis will forecast financial stability.

6.0.4 Predictive Analysis System

Future expense predictions will be based on historical transactions. Payment behavior analysis will identify late payment risks, while trend forecasting will aid financial planning. Anomaly detection algorithms will highlight unusual spending behavior, and personalized spending insights will optimize decision-making.

6.0.5 User Experience Enhancement

An interactive input system will allow users to customize financial goals. Personalized financial reporting will present clear insights, while automated alerts and notifications will support budget tracking. Customizable dashboards will tailor analytics to individual needs, and a user feedback system will enhance engagement and adaptability.

Problem Statement

Managing personal finances, particularly credit card debt, has become increasingly challenging in today's dynamic financial landscape. Many individuals rely on traditional financial management tools that primarily track expenses but fail to provide predictive insights or proactive risk assessment. As a result, users often find themselves unaware of potential financial distress until they face serious consequences such as missed payments, excessive debt accumulation, or damaged credit scores. Without early intervention or personalized financial advice, individuals struggle to make informed decisions, leading to long-term financial instability.

Additionally, financial institutions face difficulties in accurately evaluating credit risk, which can result in inefficient lending decisions and increased default rates. The lack of real-time risk assessment and data-driven financial recommendations contributes to poor financial habits and increased financial stress among consumers. Current financial tools do not adapt to individual spending patterns or provide customized strategies to help users improve their financial health.

This project aims to address these challenges by developing an AI-powered personal finance coach that leverages machine learning to predict credit card defaults while offering personalized financial guidance. By analyzing spending behaviors, the system will identify potential financial risks, generate tailored savings plans, and provide real-time financial insights. Through the integration of interactive dashboards, automated advisory features, and predictive analytics, the project seeks to empower users with the knowledge and tools needed to make sound financial decisions.

Dataset Description

”Default of Credit Card Clients” Dataset:
https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients

The project is centered around the ”Default of Credit Card Clients” dataset, which has been sourced from the UC Irvine Machine Learning Repository. This dataset captures the default payment behavior of customers in Taiwan and consists of 25 attributes across 30,000 entries, reflecting a payment history spanning six months. Out of these attributes, 23 serve as features for this study, along with one binary response variable. The dataset is robust and complete, containing no missing values. The primary focus of this project is to identify the combinations of features that most accurately predict a client’s default status, which is the target variable.

Name	**Role**	**Type**	**Description**	**Significance**
ID	ID	Numerical	Identification of each client	Serves no special purpose in training the model.
LIMIT_BAL	Feature	Numerical	Amount of credit limit.	A higher credit limit might indicate lower risk of default.
SEX	Feature	Categorical	Gender of client (1=male, 2=female)	Useful demographic for analyzing default patterns across genders but should not be the sole factor.
EDUCATION	Feature	Categorical	Education level of client (1=graduate, 2=university, 3=high school, 4=others)	Useful demographic for analyzing default patterns across education but should not be the sole factor.
MARRIAGE	Feature	Categorical	Marital status of client (1=married, 2=single, 3=others)	Useful demographic for analyzing default patterns across marital status but should not be the sole factor.
AGE	Feature	Numerical	Age of client in years	Useful demographic for analyzing default patterns across age but should not be the sole factor.

Table 1: Part 1: Description of Attributes of the ”Default of Credit Card Clients” Dataset

Name	Role	Type	Description	Significance
PAY_0 to PAY_6	Feature	Categorical	Payment status over multiple months	Past payment behavior showing how many months the client has been delaying payments. A higher value indicates a higher risk of default (-1 = pay duly; 1 = payment delay for one month; ...; 9 = payment delay for nine months and above).
	Feature	Numerical	Payment status for 2 months ago	
	Feature	Numerical	Payment status for 3 months ago	
	Feature	Numerical	Payment status for 4 months ago	
	Feature	Numerical	Payment status for 5 months ago	
	Feature	Numerical	Payment status for 6 months ago	
BILL_AMT1 to BILL_AMT6	Feature	Numerical	Bill amount across six months	Amount of bill statement shows that the more a client owes, the higher risk of default. Should be related to credit limit and payment history.
	Feature	Numerical	Bill amount for 2 months ago	
	Feature	Numerical	Bill amount for 3 months ago	
	Feature	Numerical	Bill amount for 4 months ago	
	Feature	Numerical	Bill amount for 5 months ago	
	Feature	Numerical	Bill amount for 6 months ago	
PAY_AMT1 to PAY_AMT6	Feature	Numerical	Payment amount for current month	Irregular and partial payments show financial strain. Timely payments by client decrease the probability of default.
	Feature	Numerical	Payment amount for 2 months ago	
	Feature	Numerical	Payment amount for 3 months ago	
	Feature	Numerical	Payment amount for 4 months ago	
	Feature	Numerical	Payment amount for 5 months ago	
	Feature	Numerical	Payment amount for 6 months ago	
default payment next month	Target	Categorical	Default payment for client (0=no, 1=yes)	Response variable predicted based on input features.

Table 2: Part 2: Description of Attributes of the ”Default of Credit Card Clients” Dataset (Continued)

To enhance the training process of the model, additional features have been engineered to conduct a deeper analysis of their behavior and influence on the target variable. These engineered features transform the raw data into meaningful insights, facilitating improved processing and predictions of the model. These features drive the model to become less sensitive and biased to the specific values of raw features and potentially avoid over-fitting the model. They build on the goal of improving the model's ability to make predictions on new data.

Name	**Formula**	**Description**	**Significance**
credit_utilization	credit_utilization = BILL_AMT1 / LIMIT_BAL	Ratio of bill amount to credit limit	Indicates amount of available credit being used by the client. A high ratio implies overspending and possible financial strain.
payment_ratio	payment_ratio = PAY_AMT1 / BILL_AMT1	Ratio of payment amount to bill amount	Reflects the proportion of the bill being paid off by the client. A low ratio shows difficulty in meeting payment requirements.
average_bill	average_bill = mean(BILL_AMT1, BILL_AMT2, BILL_AMT3)	Average of recent bill amounts	Average provides a smoother overview of recent spending behaviors, reducing the impact of short-term fluctuations.

Table 3: Derived Attributes from the "Default of Credit Card Clients" Dataset

Analysis & Data Representation

Exploratory Data Analysis:
https://colab.research.google.com/drive/1o1311tfMefdqZOPkLBXkqtq2iMFOdIVK?usp=sharing

Exploratory Data Analysis (EDA) has been performed for a profound understanding of the ”Default of Credit Card Clients” dataset. Insights from distributions, correlations, and risk clusters aid in creating meaningful features that improve model performance. By selecting the most relevant features and understanding complex relationships, our model strives to become more accurate and reliable in providing meaningful, and interpretive results. Identifying high-risk customer segments enables proactive risk mitigation strategies. Using EDA critiques, stakeholders can make informed decisions related to credit approvals, risk management, and customer segmentation. A deeper understanding of customer behavior and risk patterns stimulates better financial products and risk management, giving a competitive edge. This thorough EDA lays the foundation for building robust predictive models, optimizing credit risk strategies, and driving informed business decisions.

Distribution of Numerical Variables

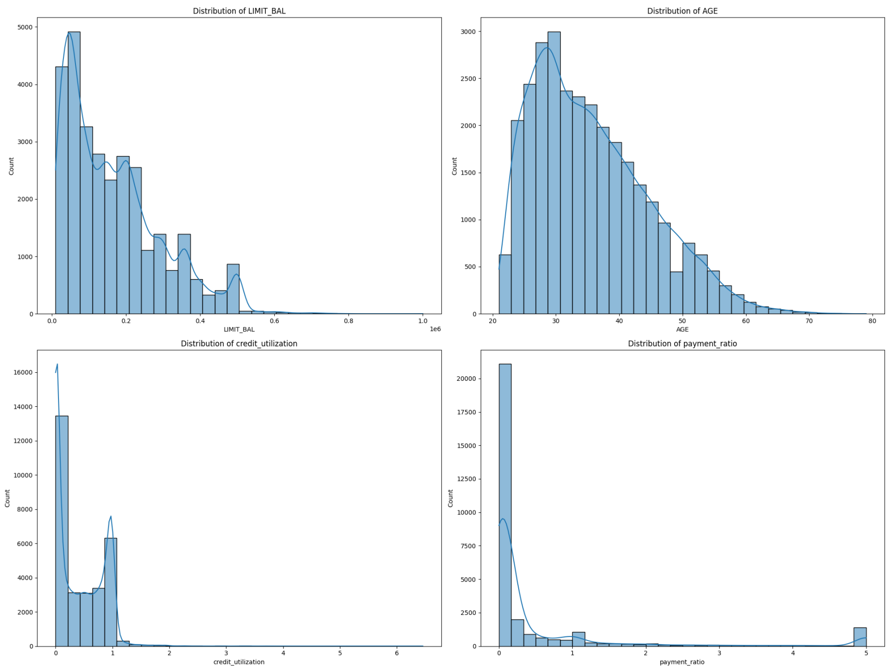

Figure 1: Histograms with KDE (Kernel Density Estimation) of LIMIT_BAL, AGE, CREDIT_UTILIZATION, and PAYMENT_RATIO Columns

Figure 1 illustrates the histograms of numerical variables to aid in comprehending the underlying distribution, central tendency (mean/median), and variability (spread) of these continuous variables. This information helps to make informed decisions about feature scaling, normalization, and potential outliers. Acknowledging the distributions helps decide the appropriate scaling technique, such as standardization for normal distributions or normalization for skewed distributions. Unusual spikes or long tails in the distribution indicate outliers that may skew model performance. Understanding the spending and repayment patterns through these graphs assist in building risk profiles, which enable targeted risk management strategies.

The top left section shows the distribution of credit limits among customers, indicating whether most customers have low, medium, or high credit limits. The right-skewed distribution may reveal credit policies favoring certain limits or customer segments. The top right of the figure is the distribution of age which identifies the age demographics of the customer base, which can help segment users by age groups (e.g., young adults, middle-aged, seniors). The bottom left section shows the distribution of credit utilization ratio. It measures how much of the credit limit is being utilized by customers. High utilization rates may indicate higher credit risk or financial stress, while low utilization suggests better financial management. The bottom right section shows the distribution of the payment ratio. It measures how much of the credit limit is being utilized by customers. High utilization rates could indicate higher credit risk or financial stress, while low utilization suggests better financial management.

Categorical Analysis with Default Rate

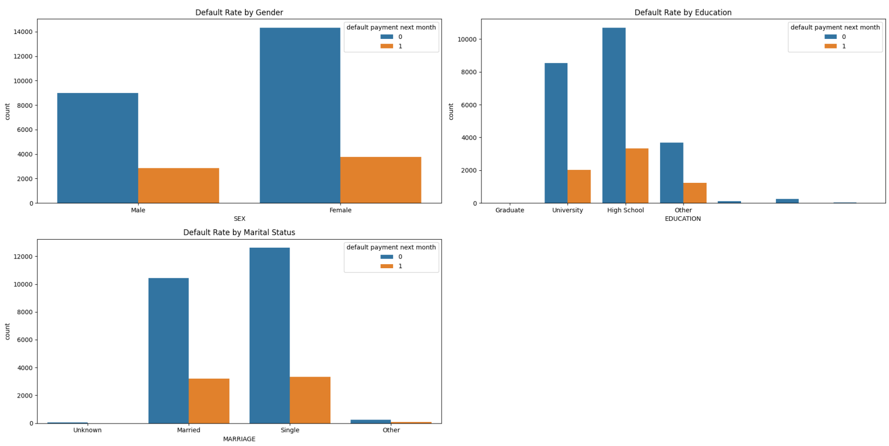

Figure 2: Distribution of SEX, EDUCATION, and MARRIAGE Columns by Default Status

Figure 2 displays the distribution of the SEX, EDUCATION, and MARRIAGE columns by the default status. The figure demonstrates the influence of categorical variables on the likelihood of default payment behavior. This helps in identifying demographic segments with higher credit risk. Demographic segmentation enables the creation of tailored credit risk models that account for demographic differences. These insights could assist in designing targeted marketing campaigns and personalized financial products.

The top left section of the image is the bar graph of the gender of each client by the default status. The graph shows that the female Taiwanese population has a higher count of defaults in comparison to the male population; however, the count of females with non-default status is also much higher than the male. The proportion of defaults within females seems slightly lower than the proportion of defaults within males. This insight may be used to develop gender-specific financial products or risk mitigation strategies. The top right section of the figure is the bar graph showing the education level among clients by the default rate. It shows how educational attainment affects financial discipline and creditworthiness. The ”University” and ”High School” categories have the highest numbers of defaults as well as non-defaults. The default count for the ”University” category is lower than the ”High School” category and the default count for the ”Graduate” category is significantly lower than the ”University” and the ”High School” categories. This portrays that higher education levels might be associated with better financial literacy and lower default rates; however, the proportions of defaults within each category seem identical. The bottom left section of the image shows the bar graph of the MARRIAGE column by the default status. It assesses the impact of marital status on credit risk. The ”Single” category has the highest number of defaults followed closely by the ”Married” category containing the second highest defaults.

Correlation Heat Map of Important Features

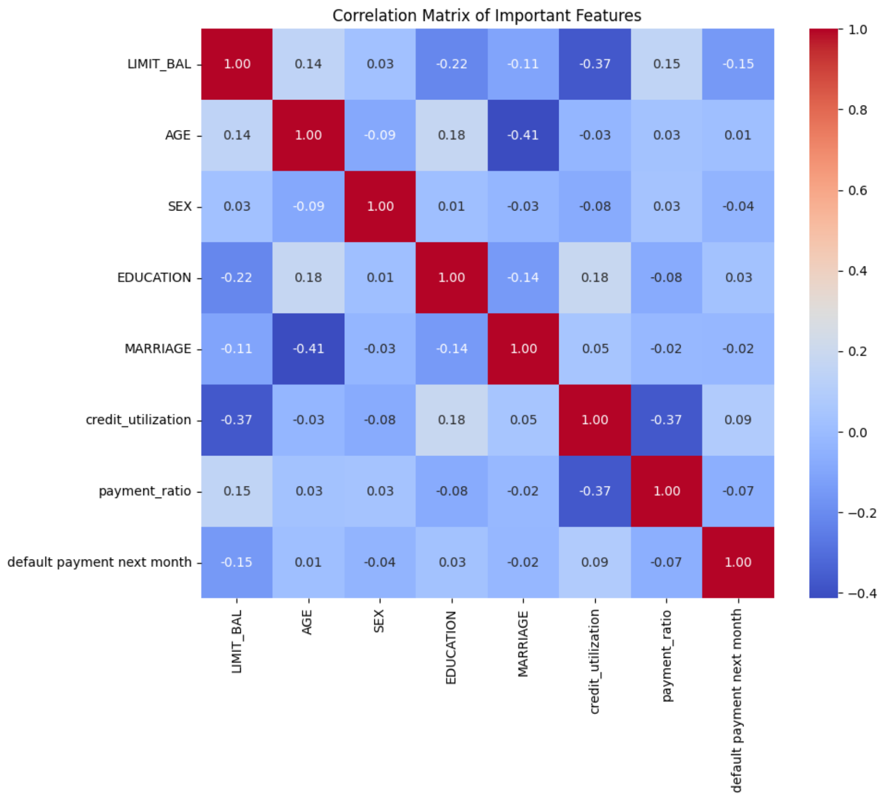

Figure 3: Heat map of the Correlation Matrix showing Pairwise Correlations between Important Features

Figure 3 reveals a heat map of the correlation matrix of each pairwise variable to help understand the linear relationships between numerical variables, exposing the highly correlated features. However, highly correlated features might be redundant, so one can be dropped to reduce dimensionality without significant information loss. Reducing multicolinearity improves model stability and interpretability, especially in linear models. The correlation matrix shows features with a positive correlation with the response variable and others with a negative correlation, indicating an inverse relationship with the response variable.

LIMIT_BAL and credit_utilization show a strong negative correlation (-0.37). As credit limit increases, credit utilization tends to decrease as expected. People with higher credit limits generally tend to use a smaller proportion of their available credit.

credit_utilization and payment_ratio also have a strong negative correlation (-0.37), suggesting that as credit utilization goes up, the payment ratio tends to go down. This indicates that people using a larger proportion of their credit limit are less likely to be paying off a large proportion of their balance each month.

LIMIT_BAL and default payment next month show a slightly moderate negative correlation (-0.15). There is a tendency for clients with higher credit limits to be less likely to default as they have more available credit to manage their expenses.

AGE and default payment next month have a very weak positive correlation (0.01). Age may not have a strong linear relationship with the likelihood of default, but there is a minute tendency for older clients to default less. SEX and default payment next month columns have a very weak negative relationship (-0.04) showing that there is a very small tendency for males to default more.

EDUCATION and default payment next month have a very weak positive correlation (0.03). Similar to age, education doesn't show a strong linear correlation with default, but there may be a slight tendency for those with higher education to default more.

MARRIAGE and default payment next month show a very weak negative correlation (-0.02), there may be a slight tendency for single people to default more.

credit_utilization and default payment next month show a weak positive correlation (0.09). There is a small tendency for higher credit utilization to be associated with a higher likelihood of default. If a client is consistently maxing out their available credit, they might be more likely to struggle with repayments. The correlation is weak proposing that other features also play a considerable role in determining default.

payment_ratio and default payment next month columns also show a weak negative correlation (-0.07). There is a again small tendency for a lower payment ratio to be associated with higher rates of default. If a client is only paying a small portion of their outstanding balance each month, they might be at a higher risk of default. The weakness of the correlation shows similarly that other variables are also important.

Risk Analysis: Credit Utilization vs Payment Ratio

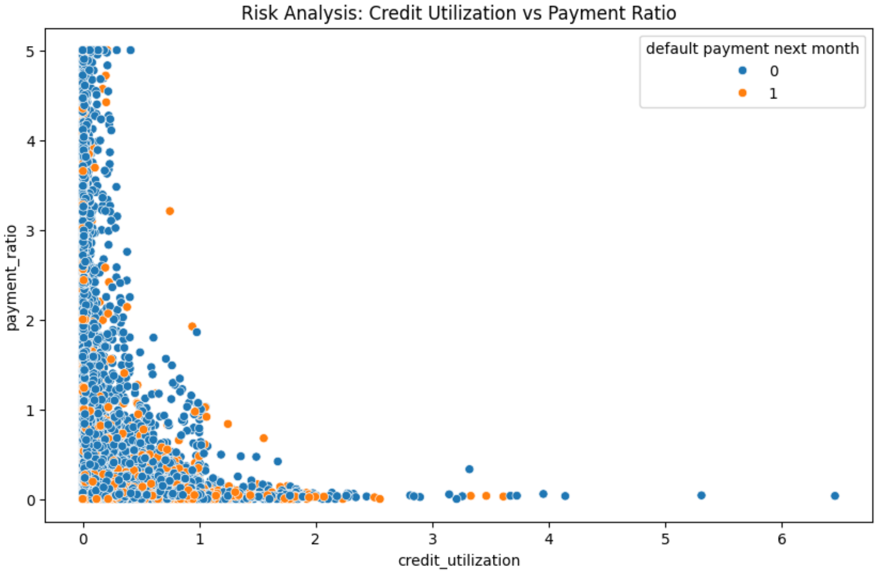

Figure 4: Scatter plot of Credit Utilization vs Payment Ratio for Risk Analysis

Figure 4 is a scatter plot of credit_utilization vs payment_ratio columns, colored by default status. The X-axis contains credit_utilization (the ratio of bill amount to credit limit) and the Y-axis contains payment-ratio (the ratio of payment amount to bill amount). This graph aims to visualize the relationship between the spending and repayment behaviors of clients and how it affects their default risk. High credit_utilization combined with low payment_ratio forms a cluster of high-risk customers more likely to default. Conversely, low utilization and high payment ratios indicate financially responsible customers. This graph assists in identifying decision boundaries for classification models by observing clusters of defaulters and non-defaulters. Conversely, low utilization and high payment ratios indicate financially responsible customers. The graph reveals complex interactions between spending and repayment behavior that may not be evident in uni-variate analysis. High-risk clusters can be targeted with risk mitigation strategies like credit limit adjustments or personalized repayment plans. Insights from this plot help improve model accuracy by incorporating non-linear interactions between variables.

Payment Status Trend Over Time

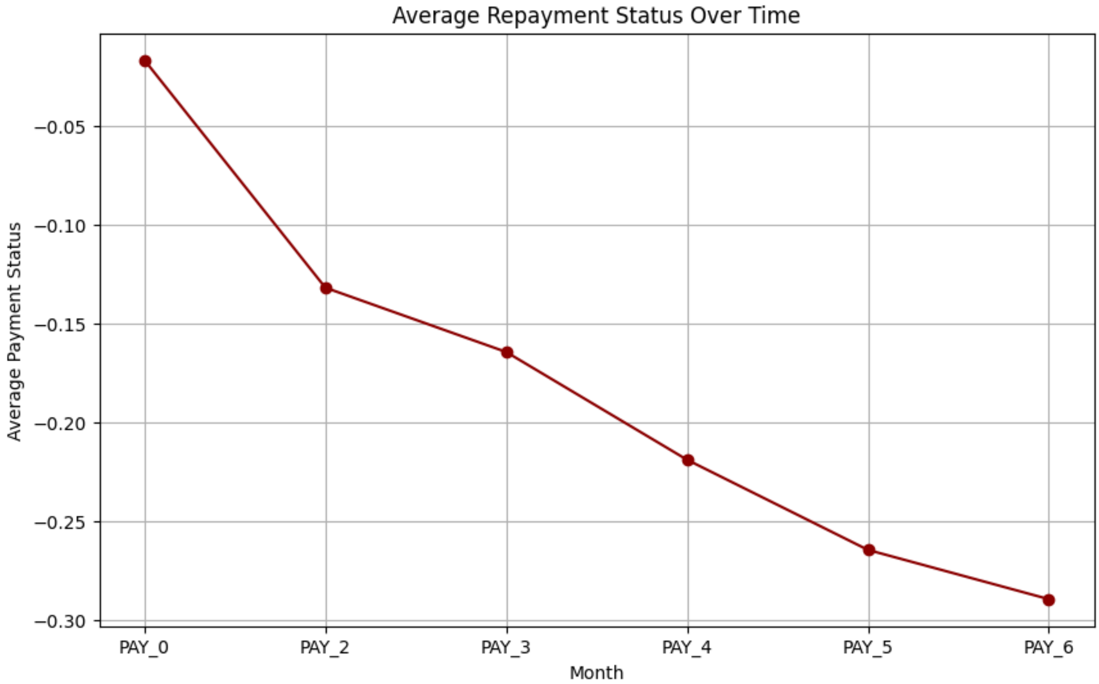

Figure 5: Average Repayment Status Overtime

Figure 5 is a line plot showing the average repayment status (PAY_0 to PAY_6) over six months. The X-axis represents the months, from PAY_0 (most recent month) to PAY_6 (six months prior) in reverse chronological order and the Y-axis represents the average repayment status for the respective month where lower (more negative) values indicate a worse average repayment status. This diagram tracks the changes in client repayment behavior over time, helping to identify trends in financial discipline or distress. The most prominent insight is the clear declining trend of the average repayment status as we move from the most recent month (PAY_0) to the earlier months (PAY_6). This means that on average, clients were more likely to delay or default on their payments in the earlier months compared to the most recent month. There is a gradual deterioration in the average repayment behavior as we go further back in time. This suggests growing financial stress or accumulating payment issues. The significant drop in average repayment status highlights changing payment patterns, crucial for risk modeling and trend analysis.

This analysis enriches the understanding of temporal repayment behaviors, which improves the accuracy of predictive modeling by incorporating time-based patterns. It also assists in Customer Segmentation by identifying consistent vs. inconsistent payers.