CREDIT CARD DEFAULT PREDICTION

Prediction of client default

Ian Macharia

Data Science

May 23, 2023

Table of Contents

[Get Started Right Away 1](#_Toc135733394)

[Make It Gorgeous 1](#_Toc135733395)

[Give It That Finishing Touch 1](#_Toc135733396)

[Add a Table of Contents 1](#_Toc135733397)

[Add a Bibliography 1](#_Toc135733398)

[Abstract 3](#_Toc135733399)

[Introduction 4](#_Toc135733400)

[Stakeholder Audience: 4](#_Toc135733401)

[Relevance of Credit Card Default Prediction for Stakeholders: 5](#_Toc135733402)

[Impact of Accurate Predictions on Business Decisions: 5](#_Toc135733403)

[The Dataset 6](#_Toc135733404)

# Get Started Right Away

When you click this placeholder text, just start typing to replace it all. But don’t do that just yet!

This placeholder includes tips to help you quickly format your report and add other elements, such as a chart, diagram, or table of contents. You might be amazed at how easy it is.

# Make It Gorgeous

* Need a heading? On the Home tab, in the Styles gallery, just click the heading style you want. Notice other styles in that gallery as well, such as for a quote or a numbered list.
* You might like the cool, blue ice pond on the cover page as much as we do, but if it’s not ideal for your report, right-click it and then click Change Picture to add your own photo.
* Adding a professional-quality graphic is a snap. In fact, when you add a chart or a SmartArt diagram from the Insert tab, it automatically matches the look of your document.

# Give It That Finishing Touch

Need to add a table of contents or a bibliography? No sweat.

## Add a Table of Contents

It couldn’t be easier to add a table of contents to your report. On the Insert tab, click Cover Page to see cover page designs that include a table of contents page — look for TOC.

Just click to insert one of these and you’ll be prompted to update the TOC. When you do, text you formatted using Heading 1, Heading 2, and Heading 3 styles is automatically added.

## Add a Bibliography

On the References tab, in the Citations & Bibliography group, click Insert Citation for the option to add sources and then place citations in the document.

When you’ve added all the citations you need for your report, on the References tab, click Bibliography to insert a formatted bibliography in your choice of styles.

And you’re done. Nice work!

# Abstract

The increasing prevalence of credit card defaults poses significant challenges for financial institutions, necessitating the development of accurate prediction models to mitigate risks. In this project, we address the problem of credit card default prediction using a data-driven approach. By leveraging machine learning techniques and a comprehensive dataset, our objective is to build a classification model capable of accurately identifying individuals who are likely to default on their credit card payments.

To achieve this goal, we performed extensive data pre-processing, including handling missing values, feature engineering, and data manipulation. We then applied various machine learning algorithms, such as logistic regression, random forest, and support vector machines, to train and evaluate the predictive models. The performance of these models was assessed using metrics such as accuracy, precision, recall, and F1-score.

Our results indicate that the developed credit card default prediction model exhibits promising performance, achieving high accuracy and robust predictive capabilities. By effectively identifying individuals at risk of defaulting, financial institutions can take proactive measures to manage their credit portfolios, reduce potential losses, and optimize decision-making processes.

The findings from this project have important implications for the financial industry, providing insights into credit risk assessment and enabling the implementation of proactive strategies to mitigate default risks. However, it is important to note that the model's effectiveness is dependent on the quality and representativeness of the dataset used. Future research should focus on incorporating additional features and exploring advanced machine learning algorithms to further enhance the accuracy and reliability of credit card default prediction models.

Overall, this project contributes to the growing body of knowledge in credit risk management and offers practical solutions to address the challenges posed by credit card defaults. The insights gained from this study can empower financial institutions to make informed decisions, minimize risks, and ensure sustainable financial stability.

# Introduction

The goal of this project is to develop a credit card default prediction model using a given dataset. The dataset contains information about credit card clients, including their demographics, credit history, bill statements, and payment records. By analysing this data, we aim to build a predictive model that can accurately predict whether a credit card client will default on their payment or not.

## Problem Statement:

The problem statement revolves around predicting credit card default, which refers to the failure of a borrower to make timely payments on their credit card. Credit card default prediction is crucial for financial institutions, such as banks and credit card companies, as it helps them assess the creditworthiness and risk profile of their clients. By accurately predicting credit card default, financial institutions can take proactive measures to mitigate potential risks and make informed decisions regarding credit approvals, setting credit limits, and debt collection strategies.

## Stakeholder Audience:

The stakeholder audience for this project includes:

1. Financial Institutions: Banks, credit card companies, and other financial institutions are directly impacted by credit card default. They have a vested interest in accurately predicting default to manage risk, protect their financial assets, and optimize their lending practices.
2. Risk Management Professionals: Risk management professionals within financial institutions play a vital role in assessing and mitigating credit risk. They rely on accurate credit card default predictions to develop risk management strategies and make data-driven decisions.
3. Credit Underwriters: Credit underwriters are responsible for evaluating creditworthiness and making decisions regarding loan approvals. Accurate default predictions assist underwriters in assessing the risk associated with a credit card applicant and determining the terms and conditions of credit offers.

## Relevance of Credit Card Default Prediction for Stakeholders:

Accurate credit card default prediction is highly relevant for stakeholders due to the following reasons:

1. Risk Mitigation: Predicting credit card default allows financial institutions to identify high-risk borrowers and take appropriate measures to mitigate potential losses. This includes setting appropriate credit limits, adjusting interest rates, or declining credit applications from clients with a higher likelihood of default.
2. Profitability: By accurately assessing the creditworthiness of clients, financial institutions can optimize their lending practices. This enables them to allocate resources more efficiently, reduce default rates, and maintain a profitable portfolio of credit card clients.
3. Customer Satisfaction: Predicting credit card default helps financial institutions identify clients who may be facing financial difficulties. Proactive measures can then be taken, such as providing financial counselling or offering flexible repayment options, to assist clients in managing their credit obligations and improving overall customer satisfaction.

## Impact of Accurate Predictions on Business Decisions:

Accurate credit card default predictions have a significant impact on business decisions, including:

1. Credit Approval and Limit Setting: Accurate default predictions enable financial institutions to make informed decisions when approving credit applications and determining the appropriate credit limits for clients. This ensures responsible lending practices and reduces the risk of default.
2. Collection Strategies: For clients at a higher risk of default, accurate predictions aid in designing effective debt collection strategies. Financial institutions can prioritize collection efforts, tailor communication approaches, and offer alternative repayment options to maximize recovery rates.
3. Risk Management and Portfolio Optimization: Accurate default predictions assist risk management professionals in assessing and managing credit risk. They can optimize the composition of the credit card portfolio, monitor risk exposure, and implement strategies to maintain a healthy balance between risk and profitability.

By leveraging the insights gained from accurate credit card default predictions, stakeholders can make data-driven business decisions that minimize risk, improve profitability, and enhance customer satisfaction.

# The Dataset

The given dataset is called "default of credit card clients" and was provided by Yeh, I-Cheng (Yeh, (2016). It is available in the UCI Machine Learning Repository. The dataset focuses on predicting credit card default payments, using a binary variable where "Yes" represents default (1) and "No" represents non-default (0).

The dataset comprises 30,000 rows and 23 columns including the target column of default payment the next month, offering a diverse range of information for analysis.

The dataset consists of 23 explanatory variables (features) and the response variable (default payment). Here's a summary of the variables:

1. X1: Amount of the given credit (NT dollar) - includes both individual and family credit.
2. X2: Gender (1 = male; 2 = female).
3. X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
4. X4: Marital status (1 = married; 2 = single; 3 = others).
5. X5: Age (year).
6. X6-X11: History of past payment - Monthly payment records from April to September 2005. (Repayment status scale: -1 = pay duly, 1 = payment delay for one month, 2 = payment delay for two months, ..., 8 = payment delay for eight months, 9 = payment delay for nine months and above.)
7. X12-X17: Amount of bill statement (NT dollar) - Bill statement amounts from April to September 2005.
8. X18-X23: Amount of previous payment (NT dollar) - Previous payment amounts from April to September 2005.

These variables provide information about the credit amount, demographics (gender, education, marital status, age), payment history, bill statements, and previous payment amounts for credit card clients.

**Response Variable:** The response variable in this dataset is "default payment," which indicates whether a credit card client defaulted on their payment or not. It serves as the target variable for our classification model. A value of 1 denotes default, while a value of 0 represents non-default

**Explanatory Variables:** The dataset contains 23 explanatory variables that are potential predictors of credit card default. These variables encompass a range of client demographics, credit history, bill statements, and payment records. Each variable plays a unique role in predicting credit card default and contributes to the overall predictive power of the model.

By incorporating a comprehensive set of variables, the predictive model can capture the complex relationships and patterns that contribute to credit card default, enabling proactive risk management strategies and informed decision-making in the financial industry.