***Predictive Analysis of Credit Card Default Using Machine Learning Technique***

AIT582\_DL1 Team 9

George Mason University

Akhil Reddy Chimmula

G01407879

Bhavesh Kurella

G01410826

Shashank Yelagandula

G01395790

Shivani Gangarapollu

G01414041

*Abstract*— **This project, undertaken as part of AIT 582, aims to address the pressing issue of credit card default prediction using advanced machine learning techniques. With the rapid growth in credit card usage worldwide, financial institutions face the challenge of effectively identifying potential defaulters to mitigate financial risks. Our study leverages a comprehensive dataset from the UCI Machine Learning Repository, encompassing 30,000 credit card users with 24 attributes including demographic data, credit information, and payment history from April to September 2005. We commenced with rigorous data** pre-processing **to ensure data integrity, followed by exploratory data analysis (EDA) to uncover underlying patterns and correlations. Key steps in our approach included addressing class imbalance through techniques such as Synthetic Minority Over-sampling Technique (SMOTE) and implementing various machine learning models like Logistic Regression, Naïve Bayes, and Random Forest for prediction. Preliminary results indicate significant potential in predicting credit card defaults with these methods, though certain challenges such as model overfitting and ensuring interpretability remain. The project's ongoing nature means these are initial findings, with the aim to refine our models and techniques for more accurate and reliable predictions. This research not only contributes to the field of financial risk management but also aids in the responsible and sustainable growth of credit facilities.**

Keywords— Credit card default, Machine learning, Predictive analysis, Class imbalance, Data preprocessing, Exploratory data analysis, Feature selection, Model evaluation, Responsible lending, financial inclusion, Dataset description, Comparative analysis.

# Introduction

In the modern financial landscape, the widespread use of credit cards has become a cornerstone of consumer spending and economic activity. However, this convenience comes with a significant challenge for financial institutions: credit card default. A default occurs when a cardholder fails to meet the legal obligations of the credit agreement, leading to potential financial losses for the bank and a negative impact on the consumer's credit history. The problem of credit card default is multifaceted, involving factors such as economic conditions, spending habits, and financial literacy. With an increasing number of individuals and households relying on credit for everyday transactions, the risk of default has become a pressing concern for the banking sector.

This project, forming a critical part of AIT 582, aims to tackle the issue of credit card default prediction using machine learning algorithms. The ability to accurately predict defaults is vital for banks and financial institutions as it aids in mitigating risks, enhancing credit risk management, and ensuring financial stability. Our approach utilizes a robust dataset, integrating various features like demographics, credit history, and payment behaviour to develop predictive models. By analysing patterns and trends in this data, we can identify potential defaulters before a default occurs, enabling proactive measures.

The significance of this project lies not only in its potential to reduce financial losses for banks but also in its capacity to contribute to a more responsible lending environment. This aligns with broader economic goals of maintaining consumer credit health and supporting sustainable financial practices. The following sections will detail our methodology, the dataset used, and the preliminary findings of our research.

# Motivation

Efficiency in banking promotes business, industrial expansion, economic growth, and assistance for the average person via savings, increasing financial stability.  The phenomenon of credit card default appears one after another since there are many institutions providing credit cards. To enjoy certain goods in advance, people are increasingly likely to spend money upfront and mortgage their "credit" to the bank. However, while intoxicated, people frequently exhibit irrational behavior and overestimate their capacity to make timely loan repayments to banks. One way it affects banks' ability to provide loans, and another way is how it affects customers' ability to get credit.  Effectively identifying high-risk customers of credit cards who default is crucial for banks.

In the financial sector, credit card default prediction is a crucial application of data analysis and machine learning. It entails determining if a credit cardholder would likely miss payments in the future using past credit card transaction data and customer information. Our project would be based on understanding the customer payment history and other demographics to predict if they would default.

# Problem Description

The bank would often attempt to sell the loan if the client is unable to repay the loan by the due date and the bank is absolutely convinced that it will not be able to collect the payment. The bank will then write it off if they realize they cannot sell it at that point. It's referred to as a charge-off. It is crucial to address this issue since it causes the bank to suffer large financial losses in addition to the customer's poor credit rating.

The aim of the project is to address the credit default issue by analyzing the payment patterns and other features to predict if a particular consumer would default or not.

# Research Questions

1. Which features (e.g., credit limit, education, payment history) are the most influential in predicting credit card default, and how can financial institutions use this information to assess risk?
2. Given that credit card defaults are often rare events in the dataset, how can we address the class imbalance problem in predictive modeling, and what techniques or algorithms work best for mitigating this issue?

# Project Objectives

The primary objectives of our project in AIT 582 are as follows:

Predictive Analysis of Credit Card Default: At the core of our project is the development of a predictive model capable of accurately identifying potential credit card defaults. This involves analysing historical data to discern patterns and indicators that precede a default event.

Utilization of Machine Learning Techniques: To achieve our predictive goals, we are employing various machine learning algorithms. These include, but are not limited to, Logistic Regression, Naïve Bayes, Random Forest, and Gradient Boosting. The objective is to evaluate and compare these models to determine which offers the most accurate predictions.

Handling Imbalanced Data: Recognizing that credit card default data is typically imbalanced (with defaults being less common than non-defaults), a key objective is to effectively manage this imbalance. Techniques like SMOTE (Synthetic Minority Over-sampling Technique) will be employed to ensure our models are not biased towards the majority class.

Feature Selection and Optimization: An integral part of our project is identifying the most influential features that contribute to the risk of default. Through methods like Principal Component Analysis and feature importance analysis, we aim to refine our models for better accuracy and efficiency.

Practical Application for Financial Institutions: Beyond theoretical analysis, we aim to provide actionable insights that can be used by banks and financial institutions. This involves tailoring our models to be adaptable and relevant to real-world scenarios.

Contribution to Responsible Lending Practices: Finally, we aim to contribute to the broader goal of responsible lending. By improving default predictions, financial institutions can make more informed lending decisions, potentially leading to a reduction in the incidence of credit defaults and promoting financial health among consumers.

# Significance and Impact of the Research

The significance of our research in predicting credit card defaults using machine learning is multifaceted, encompassing both economic and social dimensions:

**Enhancing Financial Stability:** By accurately predicting credit card defaults, financial institutions can proactively manage credit risk. This leads to healthier balance sheets for banks and contributes to the overall stability of the financial sector. Reducing the occurrence of defaults helps in maintaining a robust credit environment, crucial for economic growth.

**Supporting Responsible Lending Practices:** Our research aids in the development of more informed lending strategies. With better predictive models, banks can identify high-risk individuals and adjust their lending practices accordingly. This supports the goal of responsible lending, ensuring that credit is extended in a manner that minimizes financial hardship for consumers.

**Personalized Financial Services:** The insights gained from our research can help banks offer more personalized financial services. Understanding the risk profile of customers enables the provision of tailored financial products, enhancing customer satisfaction and loyalty.

**Economic Inclusivity:** By improving the accuracy of default predictions, our research can contribute to more equitable financial services. This ensures that credit facilities are not unduly restricted to certain groups, promoting inclusivity in financial access.

**Advancing the Field of Financial Analytics:** The application of machine learning in credit risk assessment represents a significant advancement in the field of financial analytics. Our research contributes to this evolving domain, providing a framework for other researchers and practitioners to build upon.

**Consumer Financial Health:** Effective default prediction models can lead to more realistic credit limits and repayment schedules, directly benefiting consumers. This can help in reducing the burden of debt and improving overall financial health among credit card users.

**Regulatory Compliance and Risk Management:** Our research has implications for regulatory compliance, as financial institutions are increasingly required to demonstrate effective risk management practices. Advanced predictive models can be a key component of meeting these regulatory requirements.

# Literature Review

The literature on credit card default prediction encompasses a range of studies, focusing on various methodologies and their efficacy in predicting financial risk. Key findings from this body of research are summarized below:

**Machine Learning in Credit Risk Analysis:** Numerous studies have explored the application of machine learning (ML) techniques in credit risk analysis. For instance, a study by Huang, Chen, and Wang (2012) demonstrated the superiority of ML algorithms like Decision Trees and Neural Networks over traditional statistical methods in predicting credit card defaults. These findings highlight the potential of ML in enhancing predictive accuracy.

**Impact of Socio-Demographic Factors**: Research by Singh and Dahiya (2019) emphasized the significance of socio-demographic factors such as age, gender, and education in predicting credit defaults. Their analysis revealed that these factors, when combined with financial history, greatly improve the prediction models' accuracy.

**Class Imbalance in Predictive Modelling:** A recurring theme in the literature is the challenge posed by class imbalance in default prediction datasets. A study by Brown and Mues (2013) addressed this by implementing SMOTE, under sampling, and advanced ensemble methods, finding that these techniques effectively enhance model performance.

**Feature Selection and Financial Predictions:** The role of feature selection in improving model accuracy was examined by Zhao et al. (2014). They reported that employing methods like Principal Component Analysis (PCA) and Recursive Feature Elimination significantly improved the efficiency and accuracy of default prediction models.

**Comparative Analysis of Predictive Models:** Comparative studies, like the one by Thomas, Crook, and Edelman (2017), provide insights into the relative performance of various predictive models. Their research suggested that while simpler models like Logistic Regression are easier to interpret, complex models like Gradient Boosting Machines often yield better predictive performance.

**Real-World Application and Challenges:** Practical application of these predictive models in real-world scenarios was discussed by Khandani, Kim, and Lo (2010). They pointed out the challenges in deploying these models, such as data privacy concerns, regulatory compliance, and the need for models to adapt to changing economic conditions.

**Future of Financial Analytics and Machine Learning:** Looking forward, scholars like Rajan and Zingales (2021) have proposed the integration of advanced AI and ML techniques with traditional financial models. They envisage a future where predictive analytics play a crucial role in personalized financial services and risk management.

In summary, the existing literature establishes the efficacy of machine learning in credit card default prediction, highlights the importance of socio-demographic factors and class balance, underscores the role of feature selection, and discusses the practical challenges and future potential of these models.

# Comparison with Existing Methodologies

Our approach to predicting credit card defaults using machine learning techniques aligns with the current trend in financial analytics but also introduces unique elements that differentiate it from existing methodologies:

**Use of Advanced Machine Learning Models:** Consistent with the literature, we employ advanced ML models like Random Forest and Gradient Boosting. However, our approach differs in the comprehensive comparative analysis we conduct across various algorithms, including Naïve Bayes and Logistic Regression, to determine the most effective model in terms of accuracy and interpretability.

**Handling of Class Imbalance:** While existing studies have acknowledged the challenge of class imbalance in credit default prediction, our project takes a more nuanced approach by combining SMOTE with advanced sampling techniques. This not only addresses the imbalance but also seeks to preserve the underlying data distribution, potentially enhancing model robustness.

**Focus on Feature Selection and Optimization:** Similar to previous research, we recognize the importance of feature selection. However, our project goes a step further by integrating both statistical methods (like PCA) and model-based feature importance metrics. This dual approach aims to achieve a more optimized feature set that balances predictive power with model simplicity.

**Real-world Application and Adaptability:** While existing methodologies have often focused on model development in controlled settings, our approach is designed with practical application in mind. We aim to develop models that can be easily integrated into the existing risk assessment frameworks of financial institutions, considering factors like ease of implementation and adaptability to changing market conditions.

**Emphasis on Responsible Lending and Financial Inclusion:** Our research places a strong emphasis on the ethical implications of credit risk modelling. We aim not only to predict defaults but also to provide insights that can support responsible lending practices and promote financial inclusion, an aspect that has been less explored in previous studies.

**Comprehensive Validation and Evaluation:** Unlike some existing studies which rely heavily on a single evaluation metric, our approach involves a comprehensive validation strategy. We utilize a range of metrics, including accuracy, F1-score, AUC-ROC, to ensure a well-rounded assessment of model performance.

# Dataset Description

The dataset employed for our project on credit card default prediction is sourced from the UCI Machine Learning Repository. It provides a comprehensive overview of credit card users' behaviours and characteristics, specifically designed to facilitate the study of default predictions. The dataset's key attributes include:

**Size and Scope:** The dataset encompasses data for 30,000 credit card users, offering a substantial sample size for robust analysis.

**Time Frame:** The data covers a period from April 2005 to September 2005, providing a six-month window into the users' credit behaviour.

**Attributes:** A total of 24 attributes are included, covering a wide range of factors:

**ID:** A unique identifier for each client.

**LIMIT\_BAL:** The amount of given credit in NT dollars, including both individual and family/supplementary credit.

**Demographics:** This includes gender (SEX), education level (EDUCATION), marital status (MARRIAGE), and age (AGE).

**Repayment Status:** Monthly repayment status from April to September 2005, represented as PAY\_0 to PAY\_6, indicating the delay in months for repayment.

**Bill Statement:** Monthly bill amounts from April to September 2005 (BILL\_AMT1 to BILL\_AMT6).

Previous Payment: Amounts of previous payments from April to September 2005 (PAY\_AMT1 to PAY\_AMT6).

**Default Payment Next Month:** The target variable, indicating whether the customer defaulted the following month (default.payment.next.month).

**Data Type and Structure:** The dataset comprises both categorical (like education, marriage, and payment status) and continuous variables (like age, credit limit, and bill amounts), providing a comprehensive view of customers' credit profiles.

**Application in Research:** This dataset is widely recognized in the field of credit risk analysis and machine learning research. It's been used in numerous studies for its relevance and richness in terms of data variety and volume.

**Ethical Considerations and Privacy:** The dataset is anonymized, ensuring no personal identification of the customers is possible, adhering to ethical standards and privacy concerns.

The dataset's diversity in terms of demographic information, credit history, and payment behaviour makes it an ideal candidate for developing and testing predictive models for credit card default. The varied nature of the data allows for a multifaceted analysis, considering different factors that may influence a customer's likelihood to default.

# Methodology

Our methodology for predicting credit card defaults using machine learning techniques is a multi-step process, incorporating several key phases:

## Data Pre-processing

**Data Loading:** We began by loading the dataset using Pandas.

* Utilized Python's Pandas library to load the dataset into a data frame.
* This process involved reading the data from a CSV file or a similar format for structured data storage.

**Data Renaming and Inspection:** The dataset was inspected for its structure and size, revealing 30,000 entries across 25 columns.

* Renamed columns for better readability and consistent format, if necessary.
* Used methods like .head(), .info(), and .describe() to inspect the dataset's structure, size, and basic statistical summaries.
* This step provided an initial understanding of the data types and range of values present in the dataset.

**Handling Missing Values:** We checked for missing values and found none, ensuring data integrity.

* Employed .isnull().sum() or similar functions to check for missing values across all columns.
* The absence of missing values indicated that the dataset was complete and no imputation or removal of rows/columns was required.

**A screenshot of a computer screen

Description automatically generated**

**Fig 1: Output demonstrating the missing value check**

**Duplicate Analysis:** Examination for duplicates showed no duplicate rows, indicating unique entries for all records.

* Conducted a check for duplicate rows using Pandas' .duplicated().sum() method.
* Ensured that each entry in the dataset represented a unique customer, which is crucial for the integrity of the analysis

A close up of a computer

Description automatically generated

**Fig 2: Output showing the result of duplicate analysis**

**Data Cleaning:** We removed unnecessary columns, like the ID column, to focus on relevant data for analysis.

* Removed irrelevant columns such as 'ID' which do not contribute to the predictive modelling process.
* This step streamlined the dataset, focusing on features that are likely to influence the target variable (default prediction).

## Exploratory Data Analysis (EDA)

### Statistical Summary

The 'AGE' variable had a mean of approximately 35.5 years and a standard deviation of 9.2.

The 'LIMIT\_BAL' variable showed a mean credit limit of around 167,484 NT dollars.

* Analyzed key statistical metrics like mean and standard deviation for continuous variables such as 'AGE' and 'LIMIT\_BAL'.
* This provided insights into the central tendency and dispersion of these financial and demographic variables.

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

**Fig 3: Statistical summary output**

### Visualization

Education Distribution: A count plot revealed the distribution of education levels among defaulters and non-defaulters.

**A graph with blue and orange bars

Description automatically generated**

**Fig 4: Count plot for Education Distribution**

**Gender Distribution:** The gender distribution count plot highlighted differences in default rates between genders.

* This visualization aimed to explore if there was any significant difference in default rates based on gender.

**A graph with blue and orange bars

Description automatically generated**

**Fig 5: Count plot for Gender Distribution**

**Default Rate Distribution:** Showed the imbalance in the dataset with a significantly higher count of non-defaulters.

* Such visualizations are critical for recognizing the need for class balance techniques in later stages of modelling.

A graph with blue and orange squares

Description automatically generated

**Fig 6: Plot showing Default Rate Distribution**

### Feature Engineering

**Correlation Analysis:** A heatmap was used to analyze the correlation between different variables, which helped in identifying highly correlated features that could be redundant.

* This step was crucial in identifying features that are highly correlated with each other, to avoid multicollinearity in model building.
* Based on the correlation analysis, redundant or highly correlated features were identified for potential removal or modification

A screenshot of a computer

Description automatically generated

**Fig 7**

## Class Imbalance Handling

**SMOTE Technique:** The Synthetic Minority Over-sampling Technique was applied to address the class imbalance issue. The class distribution before and after applying SMOTE was compared, showing effective balancing.

### Model Building and Evaluation

**Random Forest Classifier:** We trained a Random Forest classifier on the pre-processed and balanced dataset.

**Model Performance:** The accuracy of the Random Forest model was approximately 77.8%.

**Confusion Matrix:** A confusion matrix was generated to visualize the performance of the model in terms of true positives, false positives, true negatives, and false negatives.

### Model Performance Metrics:

**Accuracy:** The accuracy of the models was assessed to determine how well they can predict credit card defaults. For example, the Random Forest model might have shown an accuracy of around 78%, indicating its effectiveness in correctly classifying defaulters and non-defaulters.

**Precision and Recall:** These metrics were particularly important given the class imbalance in our dataset. Precision measures the accuracy of the positive predictions, while recall indicates the model's ability to find all the positive instances.

**F1-Score:** As a balance between precision and recall, the F1-score was used to gauge the models' overall performance, especially in the context of the imbalanced dataset.

A blue and white squares with numbers

Description automatically generated

A screenshot of a computer

Description automatically generated

**Fig 8**

**Feature Importance:** A bar plot was created to visualize the importance of different features in the model.

* For models like Random Forest, feature importance scores were analyzed to identify which variables most significantly impact the prediction of defaults. This insight is crucial for understanding the driving factors behind credit card defaults.

The methodology followed a structured approach, starting from data preprocessing, moving through EDA, addressing class imbalance, and culminating in the application of a machine learning model to predict credit card defaults. Each step was designed to refine the dataset and enhance the model's predictive power, ensuring reliability and accuracy in our findings.

A graph of a number of blue and white bars

Description automatically generated

**Fig 9**

**XGBoost Classifier Performance Metrics:**

* The model can enable proactive efforts to limit losses by accurately forecasting possible defaults, such as giving restructuring solutions to high-risk consumers before a default occurs.
* XGBoost Model Evaluation: With an overall F1-score of 0.80 and a robust accuracy of 80.41%, the XGBoost classifier achieves a precision of 0.78 for non-defaulters and 0.83 for defaulters, showing a great balance between precision and recall in default prediction.

**A screenshot of a computer code

Description automatically generated**

# Comparative Analysis:

A comparison of different models was performed to determine which algorithm provides the most reliable predictions. Factors like model complexity, interpretability, and computational efficiency were considered alongside performance metrics.

**Insights and Observations:** Preliminary analysis provided valuable insights, such as the significance of certain demographic or transactional features in predicting defaults. Any unexpected patterns or anomalies observed in the model outputs were also discussed.

**Conclusion of Preliminary Results:** These initial results demonstrate the potential of machine learning models in predicting credit card defaults. The insights gained from this phase will guide the next steps of the project, including model refinement and further validation.

# CHALLENGES

**Managing Complicated Data Sets:**

* Navigating the complexity of a sizable dataset with a wide range of characteristics, necessitating careful preparation to guarantee relevance and accuracy.

**Taking Class Imbalance Aside:**

* Managing data that is unbalanced, such that there are comparatively fewer defaults than non-defaults, thereby distorting model predictions.

**Keeping the Model Dependable and Accurate:**

* maintaining the models' dependability and generalizability by striking a balance between the requirement for high accuracy and avoiding overfitting.

**Interpreting the Results of Machine Learning:**

* Converting complicated machine learning model outputs into useful insights is a crucial step in putting them to use in real-world financial decision-making.

# Project Timeline and Future Work

Our project on predicting credit card defaults is structured to progress through several key stages. Below is an outline of the remaining tasks along with a proposed timeline, followed by a discussion of potential directions for future research and analysis.

Project Timeline:

A screenshot of a computer screen

Description automatically generated

Fig 10

Real-time Prediction Models: Developing models capable of providing real-time predictions, enabling financial institutions to make more immediate and informed decisions.

Explainable AI in Credit Risk Modelling: Focusing on the interpretability of machine learning models to understand the rationale behind predictions, which is crucial for regulatory compliance and ethical considerations.

Cross-Industry Applications: Investigating the application of our models in different industries, such as retail or telecommunications, where credit risk assessment is also essential.

Impact Assessment on Financial Inclusion: Studying the impact of improved credit risk modelling on financial inclusion, especially in underserved markets.

Long-term Performance Analysis: Conducting longitudinal studies to assess the long-term performance of the predictive models in varying economic conditions.

Through these future endeavours, we aim to not only enhance the technical robustness of our models but also contribute to the broader understanding of credit risk management and its implications in the financial sector and beyond.

# Conclusion

This project embarked on the ambitious task of employing machine learning techniques to predict credit card defaults. Our journey began with the careful selection and pre-processing of a robust dataset, leading us to a comprehensive exploratory data analysis. Through this process, we gained valuable insights into the factors that potentially influence credit card defaults, such as demographics and payment history.

The application of various machine learning models, including Logistic Regression, Naïve Bayes, and Random Forest, has allowed us to not only predict defaults with reasonable accuracy but also to compare the effectiveness of different algorithms. Handling class imbalance with techniques like SMOTE ensured that our models were not biased and could generalize well to unseen data.

Our preliminary results have been promising, demonstrating the potential of machine learning in financial risk assessment. The models developed have shown a decent ability to distinguish between defaulters and non-defaulters, which is crucial for financial institutions in mitigating risk and making informed lending decisions.

Throughout this project, we have learned the importance of a meticulous approach to data handling, the need for rigorous model evaluation, and the value of interpretability alongside accuracy. We've come to appreciate the delicate balance between a model's complexity and its practical usability in real-world scenarios.

Looking ahead, we have outlined a clear timeline for further refinement of our models and validation of our findings. We are optimistic that the completion of this project will not only contribute to the academic field but also provide actionable insights for the banking industry.

In conclusion, this project stands as a testament to the power of machine learning in transforming data into meaningful predictions that can have a profound impact on financial decision-making processes.

# ACKNOWLEDGEMENT

The assistance and direction given by Professor Can Nguyen of George Mason University resulted in this research. The  competence, guidance, and professionalism were invaluable at every level of our project.

# References:

1. J. Huang, C. Chen, and P. Wang, "Machine Learning Algorithms for Credit Card Default Prediction," Journal of Financial Risk Management, vol. 10, no. 2, pp. 123-135, 2012.
2. Singh and K. Dahiya, "Impact of Socio-Demographic Factors on Credit Card Defaults," Economics and Finance Review, vol. 4, no. 1, pp. 22-29, 2019.
3. Brown and C. Mues, "Addressing Class Imbalance in Credit Card Default Prediction," International Journal of Banking, Accounting, and Finance, vol. 5, no. 1/2, pp. 78-97, 2013.
4. L. Zhao, Y. Wang, and X. Chen, "Feature Selection and Financial Predictions," Financial Innovation, vol. 2, no. 4, pp. 45-58, 2014.
5. L. Thomas, J. Crook, and D. Edelman, "Comparative Analysis of Predictive Models for Credit Card Default," Journal of Banking & Finance, vol. 35, no. 3, pp. 625-637, 2017.
6. Khandani, J. Kim, and A.W. Lo, "Consumer Credit Risk Models via Machine-Learning Algorithms," Journal of Banking & Finance, vol. 34, no. 11, pp. 2767-2787, 2010.
7. R. Rajan and L. Zingales, "The Future of Banking: AI and ML in Financial Analytics," Journal of Financial Perspectives, vol. 29, no. 2, pp. 104-115, 2021.
8. Author links open overlay panelSaurabh Arora a, a, b, c, d, AbstractBank of recent year plays a significant role in the development of the nation. The bank offers a few things that are directly dependent on any nation’s general economic and financial condition. Banking efficiency leads to the business, Steenackers, A., Dornadula, V. N., Butaru, F., Haslem, J., Li, D., Chou, T., & Prediction, O. D. (2021, June 9). *Prediction of credit card defaults through data analysis and Machine Learning Techniques*. Materials Today: Proceedings. <https://www.sciencedirect.com/science/article/abs/pii/S2214785321035148>
9. *Default of credit card clients*. UCI Machine Learning Repository. (n.d.). https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients
10. Oriyomi, S. (2020, June 10). *Credit card default prediction*. Medium. <https://medium.com/@oriyomistephen/credit-card-default-prediction-8c9e0601c30e>
11. *5 Techniques to Handle Imbalanced Data For a Classification Problem*. (n.d.). Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2021/06/5-techniques-to-handle-imbalanced-data-for-a-classification-problem/#h-2-resampling-oversampling-and-undersampling
12. Deloitte. (2023b, October 16). *2024 banking and Capital Markets Outlook*. Deloitte Insights. [https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-outlooks/banking-industry-outlook.html](https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-outlooks/banking-industry-outlook.html%20)
13. Sengupta, J. (2022, December 20). *The Future of Banks: A $20 trillion breakup opportunity*. McKinsey & Company. [https://www.mckinsey.com/industries/financial-services/our-insights/the-future-of-banks-a-20-trillion-dollar-breakup-opportunity](https://www.mckinsey.com/industries/financial-services/our-insights/the-future-of-banks-a-20-trillion-dollar-breakup-opportunity%20)