**CREDIBILITY ANALYSIS OF CUSTOMERS IN BANKS**

**Financial Industry -Midterm Project**

**Ban 240NBB-Business Analytics Consulting Capstone**

Prof. Adenike Adesanmi

Ashil James 171974215

Gullu Mariyam Varghese 171992217

Jithin Philip Mukkadan 167456219

Macquina Tuscano 163773211

Nivetha Murugan 165624214

11th December 2022

Shray Kalra 160876215

**TABLE OF CONTENTS**

Contents

[1. INTRODUCTION 3](#_Toc121673193)

[**1.1 Research Aim:** 3](#_Toc121673194)

[**1.2 Research Objectives:** 4](#_Toc121673195)

[**1.3 Research Questions:** 4](#_Toc121673196)

[2. INDUSTRY OVERVIEW 4](#_Toc121673197)

[**2.1 Industry Introduction:** 4](#_Toc121673198)

[**2.2 Industry Trends:** 4](#_Toc121673199)

[**2.3 Industry Statistics:** 5](#_Toc121673200)

[**2.4 Covid -19 Impact:** 6](#_Toc121673201)

[**2.5 Key Players:** 7](#_Toc121673202)

[**2.6 Entry and Exit Barriers:** 7](#_Toc121673203)

[3. LITERATURE REVIEW 7](#_Toc121673204)

[4. DATA DESCRIPTION AND METHODOLOGY 10](#_Toc121673205)

[**4.1 Data description:** 10](#_Toc121673206)

[**4.2. Source and method of data collection:** 12](#_Toc121673207)

[5. DATA ANALYSIS AND INTERPRETATION 13](#_Toc121673208)

[**5.1 Data Analysis:** 13](#_Toc121673209)

[**5.2 Interpretation:** 16](#_Toc121673210)

[**5.3 Discussion:** 17](#_Toc121673211)

[6. CONCLUSION AND RECOMMENDATIONS 18](#_Toc121673212)

[**6.1 Conclusion:** 18](#_Toc121673213)

[**6.2 Recommendations:** 18](#_Toc121673214)

[WORKS CITED 20](#_Toc121673215)

# **INTRODUCTION**

The banking sector plays a crucial role in the financial industry. For banks, taking on a loan is their main source of revenue and financial risk. An important source of a bank's assets is the interest generated by disbursed loans. Loan lending has several hazards, one of which is the borrower's failure to repay the loan within the allotted time. "Credit risk" is how people refer to it.

It has become unavoidable for banks to take the borrower's creditworthiness into account and have a credit risk management mechanism in place because of the post-pandemic era's increased credit requirement. High levels of non-performing loans on bank balance sheets would be detrimental to the nation's economic recovery, hence proactive and effective credit risk management procedures are essential. Financial institutions must now conduct underwriting and monitoring with a new configuration that includes sector analysis, borrower resilience, and high-frequency analytics.

On looking at the available dataset, the study will be focusing on determining the various factors a bank would look at before approving a loan application. Gender, marital status, no. of dependents, education level, self-employed, applicant and co-applicant income, loan amount, credit history, property area etc. are some of the variables used for analysis. Tableau tool will be further used in this project to visualize the dependency of these various factors on credit risk analysis. The final objective is to find the most relevant factor that banks would focus on before approving bank loan applications.

## **1.1 Research Aim:**

The aim of the project is to help the banks to understand the credibility of individuals seeking the loan/credit so the banks can safeguard themselves from incurring losses due to non-performing loans.

## **1.2 Research Objectives:**

* To identify the parameters/attributes which can lead to a non-performing loan.
* To identify customers who are ineligible for the repayment of loans.

## **1.3 Research Questions:**

* What will be the customer’s fate, getting approval or rejection for their loan applications?
* What factors (e.g., education, employment status) will increase the chances of getting a loan approval?

# **INDUSTRY OVERVIEW**

## **2.1 Industry Introduction:**

The economy of any nation is a contribution of multiple sectors, but the financial sector is the most important contributor and influential sector. According to the finance and development department of the International Monetary Fund (IMF), financial services are the processes by which consumers or businesses acquire financial goods. Companies in the financial industry manage money. When a financial system in an economy starts to break the economy starts to suffer and the country, also the world at times must face recession. The financial industry has seen a significant transition over the last few decades, it is becoming obvious that the accelerating pace of technological change is the most creative force-and also the most destructive -in the financial services ecosystem today.

## **2.2 Industry Trends:**

Customers have put up with the outdated technology at traditional banks for a very long time. Customers now lean more toward these new Fintech businesses than conventional brick-and-mortar banks because of their instant services provided by technological improvements in the industry. Following this trend, conventional banks are also trying to meet these demands by utilizing innovative technologies. New technologies are currently being practiced in a lot of the financial industries.

* **Blockchain technology:**

Blockchain enables the creation of digital representations of physical assets, modernizing the procedures used for their transfer, trade, and management. Paper-based financial systems can be converted into entirely digital procedures that have lower costs and higher efficiency because of the trust, security, and automation that blockchain delivers.

* **Embedded finance:**

By using BaaS (Bank-as-a-Service), so-called "embedded finance" enables any type of business or online shop to incorporate banking software into their websites or mobile apps as an additional service within their range of services without having to direct customers to external websites. Therefore, a firm can, among other things, offer insurance, and integrate payments into its website so that customers don't have to enter their credit card information for each transaction or issue their own credit cards. This way, customers can pay for their online purchases over time.

## **2.3 Industry Statistics:**

According to the Wall Street Journal, the current personal loan holdings in the US alone are around $178 Billion and it seems to be the highest in the last 17 years. Customers depend on banks for multiple types of loans like business loans, student loans, mortgages etc.

Chart, bar chart

Description automatically generated

“The global lending market is expected to grow from $7,070.08 billion in 2021 to $7,833.88 billion in 2022 at a compound annual growth rate (CAGR) of 10.8%. The market is expected to grow to $11,285.05 billion in 2026 at a compound annual growth rate (CAGR) of 9.6%”. (Lending Global Market Report 2022)

## **2.4 Covid -19 Impact:**

COVID-19 was the most recent and greatest difficulty the financial sector and every other industry faced. Maintaining the flow of funding to the real economy while maintaining financial resilience is a twofold challenge for the financial system in this situation. Market volatility and instability have increased significantly because of COVID-19. With declining bank valuations in every nation on earth, the financial sector is most adversely affected.

## **2.5 Key Players:**

The top banking institutions in the North American region are JP Morgan Chase, Bank of America, Wells Fargo, and Citigroup. These banks dominate the landscape because they generate significant financial gains.

## **2.6 Entry and Exit Barriers:**

The relationship between competition and entry restrictions in the financial services industry is particularly complex. The major factor driving the idea that banks and other financial intermediaries are what determine whether the economy is stable or unstable is one factor. A second element is a widely held belief among policymakers that "excessive competition" in the financial services industry is harmful to sector efficiency generally. Smaller businesses and start-ups bear an unfair share of the costs associated with compliance and licensing because they may lack the scale to offset high fixed costs and sunk costs. In a report submitted by Infosys, it is estimated that global banks would be required to comply with over 120,000 pages of regulations. All these factors including the multiple regulatory obligations and their stringent implementation timelines form barriers while entering the finance industry.

# **LITERATURE REVIEW**

Commercial banks can reduce the risk of defaulters when they adhere to a specific document checklist for each loan that is prepared; make sure that all the details on collateral, credit history, and specific personal information is on file. The findings imply that loan appraisals are considered necessary, but efficient loan re-payment management is more likely to support its necessity. Financial institutions can improve loan repayment management through the implementation of credit management practices and subsequently attain long-term profitability. With the same view, another study claims that financial institutions should be mindful of loan repayment as the type of credit management strategy that should be implemented is also determined by loan repayment management. Credit management strategies focus on enhancing loan repayment management because it barely determines whether lending institutions can turn a profit. The results also imply that loan appraisals are required, though efficient loan re-payment management is more likely to support its relevance. (Wanjira, Nyasaka, Kagoyire & Shukla  
)(Mwinlaaru et al., Vinh)

Understanding how to manage money effectively and efficiently and making intelligent financial decisions based on one's available resources is defined as financial literacy. Financial literacy studies reveal that people who are more financially educated are more likely to repay their loans successfully which reduces the delays in loan repayments. Increasing people's financial literacy and awareness will enable them to manage their finances effectively, which will likely enable them to repay loans they obtain within the predetermined loan duration and prevent loan default. The socioeconomic and financial institution factors influencing non-performing debts in commercial banks are discussed in this article. Non-Performing Loans (NPLs) are loans that are non-recoverable for the period specified by a country's laws or as agreed upon between the lender and the borrower at the time of service. Due to the variances, lenders must be careful to keep to established strict standards and make sure that certain risks are eliminated or, if not, at least reduced. (Ghosh, Atkinson & Messy, Fernandes et al., Lusardi & Mitchell, 2011a, 2011b, OECD, PACFL). The development of non-performing loans (NPLs) can be explained by several factors, including global and domestic macroeconomic, bank-level, and institutional indices, according to the economic literature. Numerous studies carefully document the impact of macroeconomic indicators. More recently, more detailed bank-level statistics have been used to account for the impact of changes in external funding conditions. Domestic macroeconomic conditions link the business cycle to banking performance since shifts in these conditions are likely to have an immediate impact on a borrower's capacity to pay back debt. For instance, evidence of the anti-cyclical characteristics of NPLs shows that GDP growth frequently exhibits a negative association with NPLs. In some circumstances, global variables are employed to account for changes in the external financial environment. Several studies also include oil prices as a factor in determining NPLs Higher oil prices would be detrimental in countries that export oil decreasing the frequency of loan defaults, as evidenced by increasing earnings and income levels; Higher oil costs, however, could result in borrowers' income being reduced in countries that import oil greater NPLs. (Kimberly Beaton, Alla Myrvoda (IMF), and Shernnel Thompson)

As part of the Fed's quantitative easing programs, lending banks offered longer loan maturities, and larger loans, and loosened more covenants for companies whose long-term bond ratings were low and were lower than those with investment-grade bond ratings. This happened after the Federal Funds Rate dropped to 0.00-0.25% on December 16, 2008. Furthermore, we discover that fresh bank loans at this time were linked to a decline in a firm's value and an increase in the likelihood of default. These findings suggest that during the 2008 quantitative easing, banks increased their level of risk-taking by lowering their lending rules for borrowers who posed a higher degree of risk. The contribution of existing social capital in target communities to the effectiveness of group lending is a crucial question for microfinance. Data from field studies were conducted in South Africa and Armenia, where participants played trust and microfinance games. When it comes to groups the factors affecting not repaying are different from the factors affecting in the case of individuals. (Cassar et al.; Chen et al.). It is important to note that risk management practices are not developed and aimed to eliminate risks altogether, but they aim at controlling opportunities and hazards that may result in risk. Moreover, risk management practices also ensure that financial institutions must have a strong and rational framework for decision-making by which a firm’s objectives can be attained. Effective credit risk management practices have never been successful to eliminate the human element in making decisions about controlling risk. (Nikolaidou & Vogiazas., Ross et al., Frank et al., Ross et al., García et al., Ross et al.,)

# **DATA DESCRIPTION AND METHODOLOGY**

## **4.1 Data description:**

The aim of the study is to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience and helping the banks to safeguard themselves from the losses incurred by loan defaulters. When a bank gets a loan application, the bank must choose whether to approve the loan or not depending on the applicant's profile. Here, the bank's choice is accompanied by two different forms of hazards. "If the applicant has good credit risk, or is likely to repay the loan, then the bank loses business if the loan is not approved. If the applicant has credit risk, meaning they are unlikely to pay back the loan, the bank will suffer a loss by authorizing the loan." The strategy entails assisting banks in developing credit scoring models for business borrowers, assisting in the forecast of the borrower's credit score and repayment capacity in conjunction with the borrower's asset situation, and using machine learning to complete the data prediction. Python programming language is used for analyzing the dataset and a tableau dashboard for the data visualization to get an understanding of what factors will be the ones which will have the most important while deciding the loan credibility of an individual for banks. The plan is to try different machine learning models like logistic regression, decision tree or support vector models and check the accuracy of the model to decide the best-suited model for the research questions.

Below are the predictors that may potentially influence the credibility of the individuals’ requesting loans:

|  |  |
| --- | --- |
| **Attributes** | **Explanation** |
| Loan\_ID | Unique loan ID numbers |
| Gender | Gender of the individual applying for a loan |
| Married | Marital status |
| Dependents | Number of dependents on the applicant |
| Education | Education qualification |
| Self\_Employed | Employment status |
| ApplicantIncome | Income |
| CoapplicantIncome | Income of the co-applicant |
| LoanAmount | The total loan amount applied for |
| Loan\_Amount\_Term | The tenure of the loan |
| Credit\_history | Any previous loan is taken (yes/no) |
| Property\_area | Size of the property owned |

## **4.2. Source and method of data collection:**

The method used for the project was quantitative because to predict the loan approval status the requirement was to investigate the different parameters like LoanAmount, number of dependents, loan\_amount\_term and so on. The approach is positivist for the project i.e., it is quantifiable, and the factors that will be responsible for getting approvals or rejection of the loans from the banks will be examined. The dataset is from a secondary source Kaggle because a primary approach to collecting data couldn’t be used as it would lead to privacy issues as the information related to the finances of any individual is most critical.

Data Collection

Exploratory Data Analysis

Data Cleaning & Structuring

Dealing with Categorical Variables

Building the Model

Evaluating the Model

Credit Risk Dataset: <https://www.kaggle.com/datasets/imkushwaha/credit-risk-dataset>

# **DATA ANALYSIS AND INTERPRETATION**

## **5.1 Data Analysis:**

After the data collection the next step was the data pre-processing step. It is an essential step required to clean the data so that it is made useful to make the correct analysis and obtain the necessary predictions. In the pre-processing step, the missing values were handled in which the categorical attributes were taken care by use mode function and the numerical were dealt using the median. The dataset was skewed and had to be normalized so that the model doesn’t get biased. The next phase, after data pre-processing was to divide the dataset into training and testing on which the different models were applied. In this research two different models were used to predict the loan approval status of customers using the collected dataset. The models that were used as below:

* **Decision tree:**

It is a classification model that resembles a tree, with branches and nodes formed based on data gathered for each attribute during the model learning phase. The number of variables described in the dataset determines how the decision tree's branches and nodes are connected. The number of values assigned to each attribute is used in the forwarding process. Additionally, each branch and node reached to the decision for each transaction was followed by the rules described on each one. Finally, the class label will be applied to the record in accordance with the decision node. The iterative process repeats itself until each transaction has a class category assigned to it. To choose one of the attributes as the decision node, also known as the class label, this technique turns the attributes into branches and nodes (Gupta).

* **Naïve Bayes’ algorithm:**

The Naive Bayes classifier was the next one to be employed in this study. Additionally, it uses supervised learning to categorize the data by calculating the likelihood of independent variables. The high probability class is assigned for the entire transaction once the probabilities of each class have been calculated. Naive Bayes is a popular method for class prediction in a variety of datasets, including those used in medical and educational data mining. This approach helps categorize many dataset types, such as sentiment analysis and virus identification. It operates by predicting a predefined class for each record using the independent variable values (Zhang). As demonstrated in the following equation, it evaluates the likelihood that A will occur given that B. After that, since none of the other variables in this situation are reliant on one another, focus on determining the distinct class for each attribute. In Naive Bayes, the probability is calculated using the following equation:

After implementing the above two ML models on the testing dataset, we calculated the accuracy of the models.

The accuracy of the Decision Tree model achieved was: 70.3%

The accuracy of the Naïve Bayes’ model was 82.9%.

Looking at the above accuracy numbers the training dataset was trained on the Naïve Bayes’ model and the results were predicted for the customers getting an approval or rejection for the bank. The fig (1) shows the predicted output ‘1’ & ‘0’ which is further represented as eligibility ‘yes’ and ‘no’. If the eligibility of the corresponding loan\_id is ‘yes’, then it means the customer has received an approval for this loan application and it it's a no then the customer’s loan approval is rejected (Bank loan prediction using Machine Learning).  
A picture containing table

Description automatically generated

Fig(1). Table showing the results of the testing data after applying Naïve Bayes’ classifier

Dataset of Loan

Pre-process Data

Training Dataset

Classification Algorithms

Predictive Models

Test Models

Results

Compare & Analyze Results

Testing data

Fig (2). Flowchart for the data analysis.

## **5.2 Interpretation:**

The below figures are the Tableau results for the dataset wherein different factors were analyzed to find the most important ones that had significant impact on the approval of the loans.

**Chart, diagram

Description automatically generated**  **Chart, bar chart

Description automatically generated**

**Figure. 3.1 Figure. 3.2**

Chart, pie chart

Description automatically generated

**Figure. 3.3**

From figure 3.1 it is evident that customers’ marital status also plays a vital role in determining loan eligibility of the customers. The figure illustrates that the if the customer is married there is a higher chance of getting a loan when compared to the unmarried customer.

Figure 3.2 signifies the importance of the credit score for the customer. If the customer has a good credit score the chance of getting a loan (99.64%) which is superior to that of a customer who has lesser credit score (0.36%) or no credit history.

Figure 3.3 indicates that customers who are a graduate have a higher likelihood to get a loan which is 77%. On the other hand, when a customer has not graduated the probability is 23% of getting loan approval.

## **5.3 Discussion:**

It is clear from the data above that a married individual is more responsible and will make sure that the liabilities are paid on time. Another side of it is that a married individual will have a spouse to help pay for the loans because there is a potential that both partners are working and their salaries will rise, which will enhance the likelihood getting the loan approval from banks. The next characteristic of the study that was focused on was the credit history, and it seems obvious that someone with a good credit history will have a higher probability of being approved. The reason for this is that people with a good credit history have demonstrated themselves, and always maintained a sense of dignity and paid off their previous debts on time, which has led to a good credit score. This will reassure the financial institutions that such individuals won't default on their obligations. Furthermore, having a decent education is vital for receiving approvals since graduates are more likely to get good positions and earn good salaries that will enable them to pay off their loans. On the other hand, it is more likely that a person without a college degree won't land in a decent position and earn less money, which will make it harder for him to repay this loan.

# **CONCLUSION AND RECOMMENDATIONS**

## **6.1 Conclusion:**

The project's aim was to comprehend the variables that influence the loan acceptance procedure. So, an individual with a good credit history has the highest chance to get an approval from the bank. In this project we have tried to device a ML model that would aid banks in understanding client’s loan repayment potential and protect banks from suffering losses brought on by NPLs (Non-performing Loans). The analytical steps in developing the model include data gathering, exploratory data analysis, data pre-processing, model construction, and model testing. As per the study the credit history of a customer proves to be the most important decision factor while approval of loans.

## **6.2 Recommendations:**

Based on above research and study, below recommendations are suggested:

* While checking the credibility of the person, the purpose for the loan should be considered as an attribute. Such supervisions can become a key factor for the client’s better productivity and the assurance of repayment. As an e.g. If a person is taking a loan for education the probability of the repayment is higher because the student after completing his education will earn and plan his repayment of the loan.
* Another factor that can be considered while approval of loans is the collateral, as it will reduce the risk of the financial institutions.
* The financial institutions should also adopt a longitudinal study so they can track changes in the borrower's business or employment status throughout the credit duration.
* They can also advice the borrower on how to utilize the money they borrowed. This will aid the institutions to reduce the risk of the loan becoming a nonperforming one.

# **WORKS CITED**

Baidoo, Samuel Tawiah, et al. “Improving Loan Repayment in Ghana: Does Financial Literacy Matter?” *Cogent Economics & Finance*, vol. 8, no. 1, 2020, p. 1787693., [https://doi.org/10.1080/23322039.2020.1787693.](https://doi.org/10.1080/23322039.2020.1787693)

Bellotti, Anthony, et al. “Forecasting Recovery Rates on Non-Performing Loans with Machine Learning.” *International Journal of Forecasting*, vol. 37, no. 1, 2021, pp. 428444., [https://doi.org/10.1016/j.ijforecast.2020.06.009.](https://doi.org/10.1016/j.ijforecast.2020.06.009)

Chen, Hsuan-Chi, et al. “Bank Loans during the 2008 Quantitative Easing.” *Journal of Financial Stability*, vol. 59, 2022, p. 100974., <https://doi.org/10.1016/j.jfs.2022.100974>.

Wang, Z., Jiang, C., Ding, Y., Lyu, X. & Liu, Y. A novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending. Electron. Commer. Res. Appl. 27, 74–82 (2018). <https://www.semanticscholar.org/paper/A-Novel-behavioral-scoring-model-for-estimating-of-WangJiang/46f0d066348d6747c99c8ca74140d677ae88ab1f>

Cassar, Alessandra, et al. “The Effect of Social Capital on Group Loan Repayment: Evidence from Field Experiments.” *The Economic Journal*, vol. 117, no. 517, 2007, <https://doi.org/10.1111/j.1468-0297.2007.02016.x>.

Nkusu, M. (2011). Non-performing Loans and Macro financial Vulnerabilities in Advanced Economies. IMF WP/11/161. <https://econpapers.repec.org/paper/imfimfwpa/2011_2f161.htm>

Espinoza, R., & Prasad, A. (2010, October). Nonperforming Loans in the GCC Banking System and their Macroeconomic Effects. International Monetary Fund. <https://doi.org/10.5089/9781455208890.001>

Journal, IRJET. “Bank Loan Prediction Using Machine Learning.” *Academia.edu*, 21 Apr. 2022, <https://www.academia.edu/77162007/BANK_LOAN_PREDICTION_USING_MACHINE_LEARNING>

Gupta, Prashant. “Decision Trees in Machine Learning.” *Medium*, Towards Data Science, 12 Nov. 2017, <https://towardsdatascience.com/decision-trees-in-machine-learning641b9c4e8052>

Zhang, Zixuan. “Naive Bayes Explained.” *Medium*, Towards Data Science, 14 Aug. 2019, <https://towardsdatascience.com/naive-bayes-explained-9d2b96f4a9c0>

“Case Study: Will a Bank's New Technology Help or Hurt Morale?” *Harvard Business Review*, 27 Aug. 2021, <https://hbr.org/2021/07/case-study-will-a-banks-new-technology-help-or-hurt-morale>