**Preliminary Research Document**

**On**

**Debt Risk Analysis and Default Prediction**

Table of Contents

[1. Introduction 2](#_Toc178709633)

[2. Background 2](#_Toc178709634)

[3. Research Objectives 3](#_Toc178709635)

[4. Data Source 3](#_Toc178709636)

[5. Research Questions 3](#_Toc178709637)

[6. Data Analysis 4](#_Toc178709638)

[7. Expected Outcome 5](#_Toc178709639)

[8. Conclusion 6](#_Toc178709640)

[9. References 6](#_Toc178709641)

# **Introduction**

Effective debt risk management is essential to the stability and profitability of the banking industry. As the main providers of credit, Banks are always at danger from borrower defaults, which can cause large financial losses and cause instability in the economy as a whole. Because banks offer so many loans, it is critical that they determine each borrower's risk appropriately. Also, a financial institute plays a pivotal role in economy of the country by facilitating flow of capital, enabling businesses and access of funds to the individuals. Hence, managing its core business, credits, and mortgages, is upmost important in micro as well as macro level.

# **Background**

In the banking sector, preserving financial stability and guaranteeing profitability depend heavily on the efficient assessment and management of debt risk analysis. In the past, banks have used standardized techniques to determine creditworthiness, including evaluating collateral, financial documents, and credit scores. They have also used expert judgment and set underwriting rules.

The research will also address the classification of outstanding loan into various categories for the purpose of provisioning as mentioned in the rules and regulations of Central Bank. As per the Directive issued by Central Bank the outstanding loan shall be classified and provision as per follows:

|  |  |  |
| --- | --- | --- |
| **Loan Categorization** | **Outstanding Days** | **Percentage of Loan to be provisioned** |
| 1. *Performing Loan* | | |
| Pass | up to 1 month | 1.25 |
| Watchlist | 1-3 months | 5 |
| 1. *Non-performing Loan* | | |
| Sub-standard | 3-6 months | 25 |
| Doubtful | 6 months – 1 year | 50 |
| Loss | More than 1 year | 100 |

The major reasons of default by the borrowers are presented below:

1. Inaccurate Credit Risk Assessment
2. Over-reliance on Traditional Models
3. Aggressive Lending Practices
4. Insufficient Loan Monitoring
5. Overexposure to Certain Sectors
6. Failure to Adjust Loan Terms
7. Inadequate Risk Diversification

This capstone project will try to address some of the reasons and also provide the recommendation to the financial institution about the monitoring and classification of the outstanding loan.

# **Research Objectives**

The primary aim of this research is to explore and analyze the data related to Loan Loss Provision to contribute a deeper insight and understanding of loan performance, risk management, and portfolio health. The research is guided by the following objectives to achieve this overarching objective:

1. Understand how much exposure the bank has in case of the overdue loans classified under performing and non-performing categories.
2. Identify factors and features contributing to the loan default through the machine learning model.
3. Determine which branches of the bank have the higher risk of default so that bank focuses on limiting the limit of loan to be sanctioned in such branches.
4. Segment analysis through identifying the exposure of Loan under Retail sector, corporate sector, and foreign sector.

# **Data Source**

In this research, unpublished data from our internal analysis was used to examine the Debt Risk and Default Prediction of loan sanctioned by a bank.

The data is about the Loan Loss Provision to be prepared by bank for fiscal year 2022-23 (from 17th July 2022 to 16th July 2023) with respect to the rules and regulations of Central Bank of country.

# **Research Questions**

1. Classification of outstanding loan into various categories for the purpose of provisioning (Loan Loss Provision) in the balance sheet drafted at the end of each fiscal year as per the rule and regulation of the regulatory body.
2. What are the most important predicting factors (feature vectors) of loan defaults, and how can it be used to create an effective risk assessment model for future loans decisions?
3. Segmentation Analysis of the loan with respect to various sectors.
4. Time Series and Trend Analysis

# **Data Analysis**

This research will employ machine learning models to predict loan defaults using a historical loan dataset. The process will include the following steps:

1. Data Collection

Unpublished dataset, which is like a big spreadsheet that contains past information about loans and borrowers has been selected for the project. This dataset includes details like:

* Branch
* Account Type
* Interest Rate
* Loan Outstanding
* Loan Limit
* Loan Loss Provision
* Segment

These are just 7 examples of columns and we have included full details of columns in separate file as meta data.

1. Data Preprocessing

When working with this data, some of the information might be incomplete, incorrect, or irrelevant. Before using it, data will be fixed in respect to any errors or missing information to make sure everything is correct.

1. Exploratory Data Analysis (EDA)

Some of the exploratory data analysis may include the following but shall not be limited to:

* Exploration of distributions of key columns like INTRATE, OUTSTANDING, and LIMIT.
* Compare the performance of loans across different branches (BRANCHNAME) or states (STATE).
* Investigate loans with high PRINODDAYS or INTOVDDAYS to assess overdue risk.
* Check if any patterns emerge with overdue loans across BRANCHCODE, STATE, or CLIENTCODE.

1. Building Models

We will use different types of models to predict which borrowers might default. The types of models we'll use are:

* Logistic Regression: A basic model that gives us a starting point.
* Decision Trees: A model that works like asking a series of yes/no questions to decide the outcome.
* Random Forests: A model that combines several decision trees to make a stronger prediction.

1. Data Visualization and Reporting

* Visualization tools like Tableau, Power BI, or Matplotlib/Seaborn will be used to present our findings.
* Bar charts for comparing loan classification across segments.

# **Expected Outcome**

The following outcomes can be accomplished with the analysis and predictive modeling efforts:

* 1. Based on branch, Determination of which types of loan are typically assigned a high-risk and categorization according to loan loss provision.
  2. Evaluate loan performance across branches to determine which branches are most likely to have more defaults.
  3. Outcome of machine learning model grouping the features vectors according to their importance to be used in predicting defaults.
  4. Examine loans with large overdue principal and interest to determine the risk of default.
  5. Analyze the relationship between accrued interest and loan classification. High levels of accrued interest can mean financial trouble among borrowers that requires greater attention and possible restructure options.

This analysis will provide changes to debt regulations allowing maintained monitoring of customer behavior and provide a strong basis for long-term financial planning. These results will also help the financial institution to promote financial stability, reduce default losses, and create a more flexible financial environment. Financial institutions will be more comfortable to manage possible defaults and conduct specific strategies if they identify important risk factors, classify debts based on their risk profiles, and develop predictive models.

# **Conclusion**

In conclusion, this Preliminary Research Document has the road map for data analysis, EDA, preparation of machine learning model and presentation of the insights through data visualization so that the research questions mentioned above can be addressed.

# **References**

* Stephanou, C., & Mendoza, J. C. (2005). Credit Risk Measurement under Basel II: An Overview and implementation issues for developing countries. In World Bank, *World Bank Policy Research Working Paper* (Vol. 3556) [Report]. <https://documents1.worldbank.org/curated/en/431871468330983467/pdf/wps35560corrected.pdf>
* Stephanou, C., & Mendoza, J. C. (2005). Credit Risk Measurement under Basel II: An Overview and implementation issues for developing countries. In World Bank, *World Bank Policy Research Working Paper* (Vol. 3556) [Report]. <https://documents1.worldbank.org/curated/en/431871468330983467/pdf/wps35560corrected.pdf>
* <https://www.nrb.org.np/contents/uploads/2024/01/ECL-Guidelines2024-2.pdf>
* Economic Computation and Economic Cybernetics Studies and Research, Issue 2/2019; Vol. 53

<https://ecocyb.ase.ro/nr2019_2/9.%20Coser%20Al.%20Crisan%20Albu%20(T).pdf>

* Credit Risk Analysis Using Machine and Deep Learning Models by Peter Martey Addo, Dominique Guegan and Bertrand Hassani

<https://www.mdpi.com/2227-9091/6/2/38>