Forecasting of Customer's Eligibility for The Bank Loan and, Also Predict Credit Limit, Using ML Approach

Project Id: Research Project - (24-25J-268)

Project Proposal Report

A Project Proposal by

D N PATHIRATHNA

Supervised by

Suranjini Silva

B.Sc. (Hons) in Information Technology Specializing in Information Systems Engineering

Department of Computer Systems Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

August 2024

Declaration

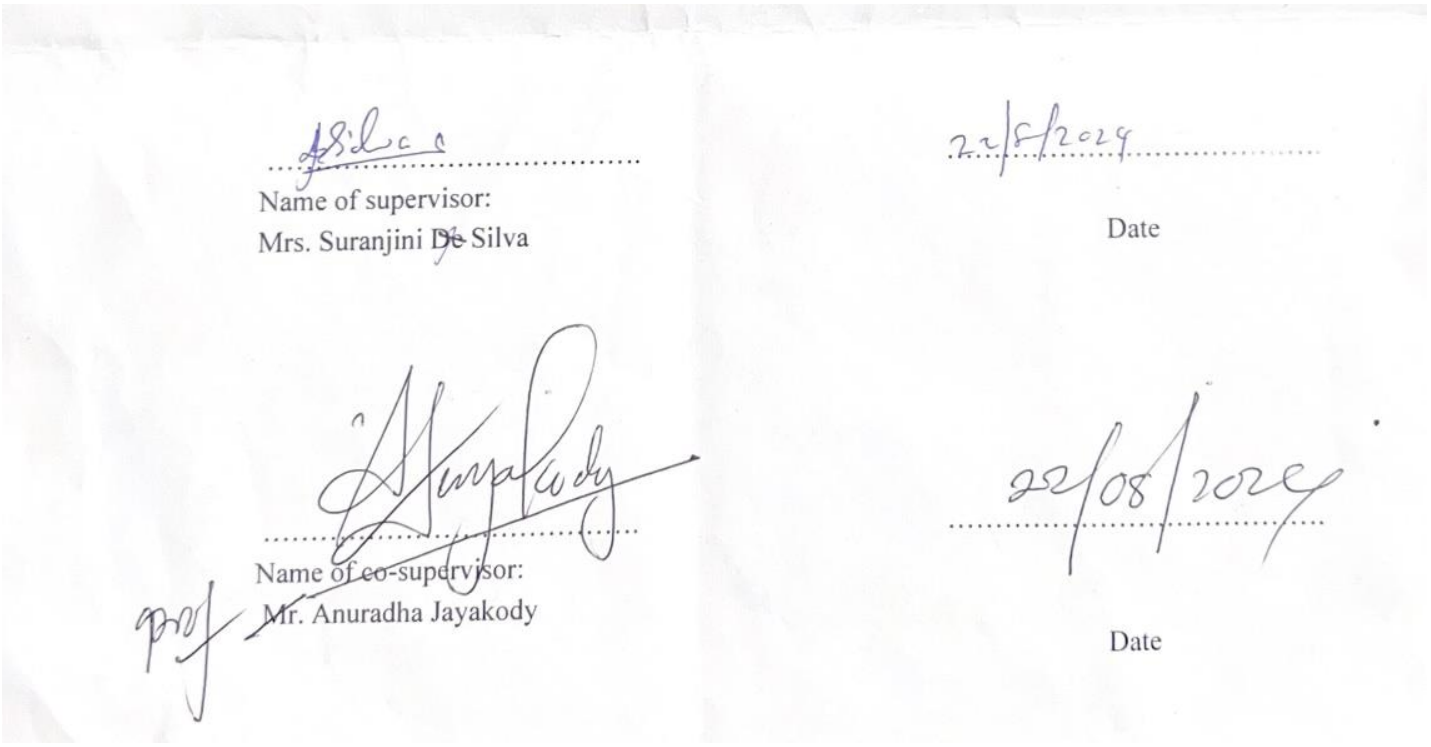
We declare that this is our own work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning, and to the best of our knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Name	Student Id	Signature
D N PATHIRATHNA	IT21164644	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Name of supervisor: Mrs. Suranjini Silva

Name of co-supervisor: Prof. Anuradha Jayakody



I. Table of Contents

II. List of Tables

III. List of Figures

Acronyms

ML – Machine Learning
AI – Artificial intelligence
CRIB - Credit Information Bureau of Sri Lanka

1. Introduction

It's common knowledge that the lending business generates billions of dollars annually. Lenders profit from interest payments (Murfin and Petersen, 2014), so they provide competitive rates, while borrowers look for the most favorable conditions. Borrowers are often expected to disclose their income, assets, as well as credit score when applying for a loan. Lenders and borrowers alike may find this procedure to be tedious and aggravating. This essay will go over the usage of artificial intelligence and machine learning to gauge loan qualifications. As data scientists, it's crucial to have a firm grasp of the difficulties associated with determining loan eligibility (Azzini et al., 2019) , as well as the means by which these obstacles might be overcome via the development of machine learning models. We'll show how machine learning may be utilized to benefit both lenders and applicants in the loan application process (Padimi, .. and Ningombam, 2022).

Eligibility for a loan refers to the simple procedure of figuring out whether a possible borrower satisfies the requirements for that loan. This may depend on things including a person's salary, length of work, as well as credit history. Eligibility for loans is crucial in guaranteeing that borrowers will be able to pay back their debt. The default might have serious implications, such as a drop-in credit score and trouble getting new funding in the future. Lenders thus use extreme caution when determining loan suitability. Borrowers may increase their chances of loan approval by learning more about the criteria used by lenders. Lenders will typically employ a combination of criteria to establish a borrower's suitability for a loan, since there are many factors that might influence this decision. Copying the borrower's credit report is a typical practice. A borrower's current debt load and payment history may be seen in this document. Lenders will also research the borrower's credit history and other public sources (Barron, 2000). Lenders may also ask for proof of income from the borrower, including such tax returns as well as pay stubs.

2. Problem Domain

Loans are the core business of banks. The main profit comes directly from the loan’s interest. Customers who are likely to default on investment and business loans may be identified with the use of feasibility analysis, financial documents, and research into the company's and borrower's reputations in the marketplace. In order to make sound lending decisions, banks offering private loans must collect detailed information on prospective borrowers and keep

close tabs on their behavior throughout the life of the loan also Banks should also consider the relationships between credit risk and other risks. Any banking organization's long-term survival depends on the efficient management of credit risk, which is a vital part of a comprehensive risk management strategy (Alrabiah, 2018).

3. Problem Definition

Customers who are likely to default on investment and business loans may be identified with the use of feasibility analysis, financial documents, and research into the company's and borrower's reputations in the marketplace. In order to make sound lending decisions, banks offering private loans must collect detailed information on prospective borrowers and keep close tabs on their behavior throughout the life of the loan. Banking industry Score Company includes details on the credit history of the clients and moreover helps to prevent default customers via the use of a scoring system (Gallati, 2003). Due to data scarcity and the lack of reliable methods for interpreting client credit histories, this is easier said than done.

3.1 Problem Statement

Financial organizations are providing loans to their customers, so financial organizations must identify the most suitable and relevant customers in their wide customer base and identify the suitable credit limit for each customer according to their financial background.

4. Research Motivation

In Sri Lanka, a customer must wait several days to see if he or she eligible for a loan. As a result, they are not actual idea about they are eligible or not for the loan process also they didn’t have actual idea about loan amount the bank will allocate for them. Nowadays Banking sector collaborate with IT industry and they experienced new technology for their services (Ho et al., 2019). According to the ML and data science, banking sector can implement loan eligibility predictor for their customer.

5. Existing Works

Citation	Description	Improvements	Limitations
Loan Prediction Using Machine Learning and Its Deployment On Web Application (Sujatha et al., 2021)	The loan prediction is developed using machine learning algorithms such as logistic regression. The Python programming language is used or the implementation of the code which has been developed in Colab and the html pages are developed for deployment of website using Visual Studio code.	Logistic Regression and Decision tree algorithms have been used to predict the results The python is used to implement the code in Colab. Flask and Herok app are used to deploy the project. In order to enhance the operating speed of the proposed system.	This work can be extended in order to improve the focus where the high accuracy can be obtained. System only predicts the eligibility of the customer. In this paper predicting the loans according to the dataset taken. Different countries have different attributes as their priority, those attributes can be included in the dataset while training according to their bank's choice
Loan Prediction System Using Decision Tree and Random Forest Algorithms (Chaudhary, 2020)	Logistic Regression, Decision Tree, and Random Forest is applied to this prediction s. Through this system, one can predict if that specific candidate is protected and the entire interaction of approval of highlights is robotized by the ML method.	The outcome of the research is in 79.45% of accuracy achieved. Used two Algorithms to get an accurate prediction.	Consider few utilizes models and also dataset Outdated with considering new patterns. Required larger training sets to give accurate results Only

An Empirical Study on Loan Prediction Using Logistic Regression and Decision Tree (Aditya Sobika et al., 2021)	Logistic Regression, and Decision tree examination used to predict eligibility of loan	According to the paper researchers applied a novel approach for credit scoring by combining genetic programming with deep neural networks.	Required larger training sets to give accurate results there are less than 10 explanatory attributes the accuracy of SVM is reduced.
Prediction of Modernized Loan Approval System Based on Machine Learning Approach (Singh et al., 2021)	In this research paper, using three Machine Learning algorithms. prediction of Data set. XGBoost – XGBoost is a Decision tree based open-source software library. Random forest Decision tree.	proposed model will characterize the behavior of customers on the Basis of their record.	Only show client is eligible or not. According to the paper in some situations like client going through some disaster algorithm cannot predict the appropriate result. Consider few utilizes models and also dataset Outdated with considering new patterns.
Prediction of loan status in commercial bank using machine learning classifier (Arutjothi and Senthamarai, 2017)	According to the paper K-NN credit scoring model was built with the programming language R.	Credit scoring is widely analyzed using classification techniques. Feature selection techniques are used to remove the irrelevant attributes.	Loan applicant should be default or valid customer. Only show client is eligible or not.
Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector (Kumar et al., 2022)	Feature selection method is applied to select more appropriate features to predict the customer loan eligibility according to the research Decision tree, random forest, support vector machine, k-nearest neighbor, and hybrid model with decision tree and adaboost combination are implemented in this work.	Python PYSPARK machine learning package is used to train the model. Used three Algorithms to get an accurate prediction.	Required design an innovative model for customer loan eligibility prediction over banking sector. The model's results are analyzed in terms of accuracy parameter, and performance is represented in graphical format.

Table 1: Existing work

6. Research Gap

Mainly identify three research areas after consideration above existing works,

- All the existing works were given status according to whether they were eligible or not for the bank loan process, didn't implement predicting eligible amount.
- All the existing works consider few datasets and above research papers consider few inputs for the train the model, therefore eligibility predicting accuracy became low.
- All the existing works didn't consider how to change the model and prediction approach with economic patterns.

7. Research Contribution

7.1 Contribution to the Technology

The Vector Machine Support method will be used for this task. In the realm of machine learning, Support Vector Machine is classified as a machine-learning algorithm that is supervised. Additionally, to regression, it could also be utilized for classifying data. In this paradigm, each component is represented by a single coordinate in an in-dimensional space, and the value of each characteristic corresponds to this coordinate's position. The next step is to classify the data by determining the hyper-plane that separates the two groups (Khedr et al., 2021). They have made available an information that can be utilized to automatically determine which customer subsets qualify for certain loan amounts.

7.2 Contribution to the Problem Domain

The banking industry relies heavily on loans as a means of profiting from the financial risk they take on their customers. The interest collected on loans is a significant component of a bank's assets. There is a high degree of danger involved in the business of loaning money, one of the hazards being that the borrower won't pay back the money borrowed by the due date. The term for this is "credit risk." Credit scores were used to determine whether or not a loan applicant was eligible for approval. Thus, this study intends to explain the use of various Machine Learning technique which properly identifies whom to offer loan to and assists banks in identifying the problem loans for significantly decreased creditworthiness.

8. Research Challenges

Individuals desire bank loans in large part due to advancements in the banking sector. However, the bank can only offer loans from a finite pool of funds, hence it is customary for banks to screen applicants for loans before offering them. In this post, they tried to apply ML to predict if a specific consumer would be granted for a bank loan and how much that loan would be. In this scenario loan predicting accuracy should be high according to the prediction. According to the prediction loan eligibility status and loan amount should be accurate to bank requirement. After the considerations of existing works, they identified accuracy issue due to data set training process. So, challenge is research and identify best accurate prediction for the loan eligibility and loan eligibility amount. it's should be helps to bank identifying suitable for customers for the bank loan service.

9. Research Question

1. What is Forecasting of Customer's Eligibility for The Bank Loan and, Also Predict Credit Limit, Using the ML Approach?
2. How to identify Bank Loan Approval Prediction Using Data Science Techniques?
3. What is the solution for Forecasting Customer's Eligibility for The Bank Loan and, Also Predict Credit Limit, Using the ML Approach?

10. Research Aim

"The project's aim is to design, develop and enhance forecasting of customer's eligibility for the bank loan and, also Predict credit limit, using ML approach"

11. Research Objective

Research Objective	Explanation
Literature Survey	Read and study the existing works to clarify the relevant information on related work and critically evaluate them. • **RO1**: Conduct a preliminary study on existing Predicting Systems & Architectures • **RO2:** To clarify the technologies, algorithms, frame works used for developing Opportunistic models.
Requirement Analysis	Defining the project's needs using the relevant approaches and tools in order to solve the expected research challenges and gaps based on prior related research and any domain-specific knowledge sources. • **RO1:** To get the relevant requirements of domain experts and researchers in order to understand how they will expect it to work. • **RO2:** To use surveys to specify the hardware and software requirements.
Design	The proposed prototype system Design and Architect by author based on the researched requirement. • **RO1:** To Increase the accuracy, will be use several algorithms. • **RO2:** To Increase the accuracy, will be consider more patterns according to the data set • **RO3:** To Increase the user experience, will be use graphical user interface to display prediction output.
Development	The proposed prototype system will be implemented by the author based on the researched requirement. • **RO1:** Frontend implementations done with Angular. • **RO2:** Backend implementations done with Angular. • **RO3:** ML and Data Science part implementations done with python. • **RO4:** System deployment done with cloud service.
Testing and Evaluation	Testing the created system and ML models with the relevant data set. Evaluate the prototype with the help of academic experts, industry experts and expected domain expert users of the system. • **RO1:** To Create a test scenario and test individual component. • **RO2:** To Create test plan for unit testing and functional testing. • **RO3:** Arrange session to discuss and evaluate with industry, academy and domain experts.

Table 2: Research Objective

12. Scope of the Research

12.1 In Scope

- Identify the most suitable algorithms and train with the date set and identify best model for the prediction.
- Predict loan amount after consideration of input parameters and using ML algorithms.
- Change the model and prediction approach with economic patterns.
- All the input parameters are highly restricted and consider privacy high improperness

12.2 Out Scope

- Eligible customer can apply loan though the system
- The proposed system can implement or other organizations to consider their employees financial situations.
- Introduce loan guarantors to the system and identify guarantors who are eligible or not (Bloomenthal, 2022).
- CRIB (Credit Information Bureau of Sri Lanka) integrates with the system.
- Implement centralized system and connected all the financial organizations to the system and full fill the customer requirement according to their financial situation.

12.3 Prototype Diagram of the Proposed System

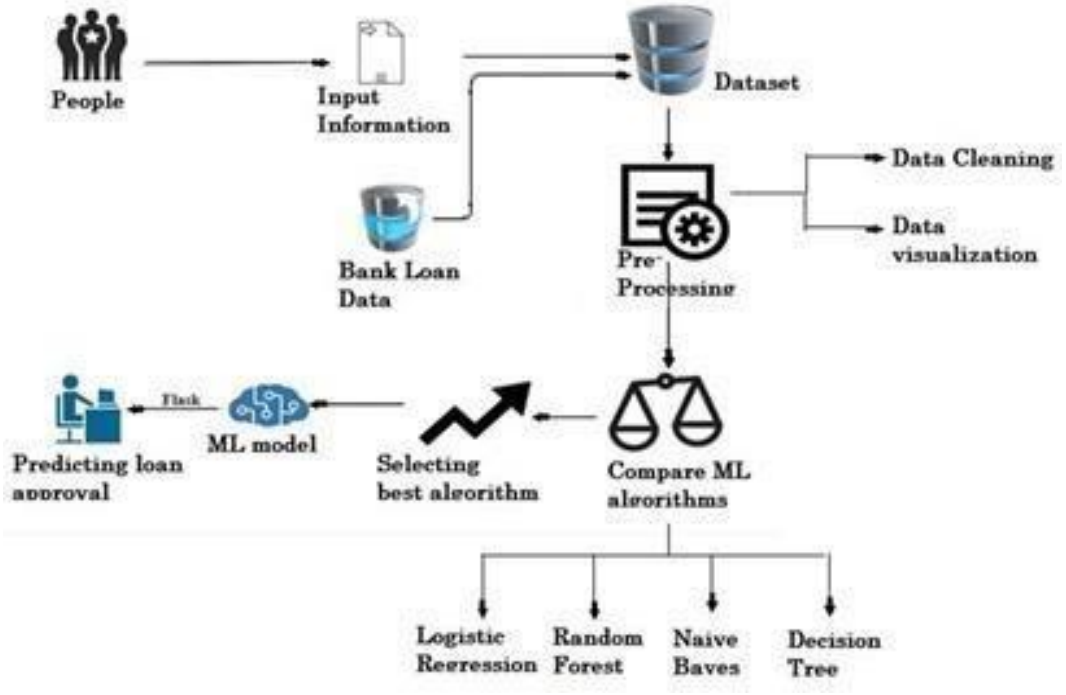


Figure 1: Prototype Diagram of the Proposed System

13. Research Methodology

According to (Boaz and Ashby, 2003) point out four main areas in research "methodological quality, quality of reporting, appropriateness of strategy, and relevance to policy and practice" In addition to that methodologies are doing key role in research. Therefore, researchers can research relevant things to the society in addition to that research methodologies were selected from the predefined Research Onion Model (Saunders, 2019)

Research Philosophy	There are popular data collection methods associated with research philosophy also the philosophy of pragmatism was chosen because research depends on data collecting, analyzing and processing to develop hypotheses.
Research Approach	Research approaches are the project's research strategies and plans. According to the consideration of the research Deductive Approach was selected as a research approach. This is anticipated to be a quantitative approach with the goal of test and proving the hypothesis under consideration.
Research Strategies	By responding to the research questions, the research strategy establishes how the methodology is to be used. Surveys, interviews, questionaries and experiments based on evaluation metrics will be utilized to narrow down the sets of potential participants.
Research Choice	Basically, there are three major options. Mono, Mixed and Multi-method are mentioned three things and Mixed was chosen for this scenario. In addition to that quantitative results are the focus; it is acknowledged that the quality of the data used to design the system will also be a crucial factor that will influence the quantitative results.
Time Horizon	The duration of the investigation is determined by the Time Horizon. In this scenario the research's time horizon was selected as longitudinal since data will be acquired and used for testing and evaluation over a lengthy period of time.
Techniques and Procedures	Techniques like observations, documentation, discussions, evaluation reports, interviews, and questionnaires will be employed for data collecting and analysis.

Table 3: Research Methodology

14. Development Methodology

14.1 Software Development Methodology

In this scenario, the author is anticipated to use prototypes for the research project as well. Software engineering methodology is a process used under the development methodology. As a result, authors are anticipated to use prototype models after considering the testing and outputs, make alterations, and test them again and again until a positive consequence is achieved and an achievable outcome is gathered.

14.2 Design Methodology

Author anticipated to for Design Methodology, Object Oriented Analysis and Design (OOAD). In addition to that author expected to reuse the models and intractability among objects. Its support to future developments and future enhancement of the research and system.

14.3 Evaluation Methodology

The system's evaluation is a crucial step in determining whether the suggested solution meets the stated issue. As a result, consideration of literature and identify the fresh techniques for related problem domain. Therefor researches able to evaluate the system using identified evaluation approach.

14.4 Requirement Elicitation

In this section, explained how the project's needs will be obtained. The following elicitation techniques will be employed for this approach.

Method	**Justification**
Literature Review	According to the literature review, author can identify the pervious works, existing systems and similar technologies.
Observations	Observation helps to verify requirements and deliver instant requirements worthy of consideration also observations obtained

	during the experimentation phase will be used for the completed project.
Surveys	People who have the necessary expertise in machine learning and embedded systems will receive surveys. This will ensure that the data collected is of the greatest standard.
Interview	Interviewing domain experts, industry experts, and stakeholder groups is more efficient for gathering requirements and resources.
Questionnaire	Questionnaire is best solution to identify and clarify the requirements and resources. Also interview data is qualitative, it will take thematic analysis to gain crucial information from the participants.

Table 4: Requirement Elicitation

15. Project Management Methodology

Machine Learning is the branch of AI that has seen the most progress in the recent decade. This is a group of algorithms and statistical techniques for analyzing data in order to make predictions or enhance performance. The software, known as a model, "learns" from historical data without being specifically programmed by a human, giving rise to the notion that the machine is capable of autonomous learning (Burke, 2010).

The demand for executives with a foundational grasp of Machine Learning and Artificial Intelligence has grown rapidly in tandem with the rapid adoption of Machine Learning in the business and the employment of thousands of technical professionals like Data Scientists and Machine Learning Engineers.

Machine learning iterations are very congruent with Agile processes. Agile was developed for circumstances that are both highly dynamic and difficult to predict in advance. It points the way to the solutions, but it's flexible enough to accommodate and even promote change along

the way. Regardless of whether there are gaps in the planning, the development team will have more structure and a better framework to work within (Bourne, 2016).

15.1 Resources Requirements

15.1.1 Skill Requirements

Skill Requirement	Justification
Research Skills	Ability to doing proper research.
Programming Skills	Ability to implementing required and proposed system and follow best practices.
Testing skills	Ability to testing implementing required and proposed system.
Project Management	Ability to manage project and complete the research in proper manner.
Presentation skills	Ability to present problem and research information.

Table 5 : skill Requirements

15.1.2 Software Requirements

Software Requirement	Justification
Operating System (macOS /Windows 10 /Windows 11)	The OS is required to manage all processes and able to run required programs such as IDEs, Libraries, and Tools
Python/Angular/.NET	Python will be used as the primary programming language to implement and develop prediction model using ML and data science Angular will be used as the primary programming language to implement front end. .NET will be used as the primary programming language to implement Backend end
PyCharm IDE / Jupiter IDE	IDEs used for implement Code for the systems

IntelliJ IDEA/ Visual Studio Code/Note Pad	
Google Drive/Git Hub/Git Lab	Will be used for Document and Code backups
SQL Database - MYSQL	SQL Database will be used as the primary Data base management system.
MS Office/ Mendely/ Google doc	A documentation, reporting and researching tools use manage documentation and backup purpose.
Figma/Draw.Io	A designing tool used to create diagrams, wire frames and data visualization charts.
Google Colab	Google Colab will be used as cloud development environment to build, train & test ML & Deep Learning models.

Table 6 : Software Requirements

15.1.3 Hardware Requirements

Hardware Requirement	**Justification**
Core i5 Processor or above	To perform long running, Intensive processing power.
16GB RAM	To manage data-sets & Setup local development environments.
Disk space 100GB or above	To store datasets, files, applications, IDE and frameworks necessary for development purposes.
Screen resolution 1360*768	To visualized data and display outputs

Table 7: Hardware Requirements

15.1.4 Data Requirement

Data Requirements	**Justification**
Customer Data sets	Kaggle and Google Dataset Search.

Table 8: Data Requirement

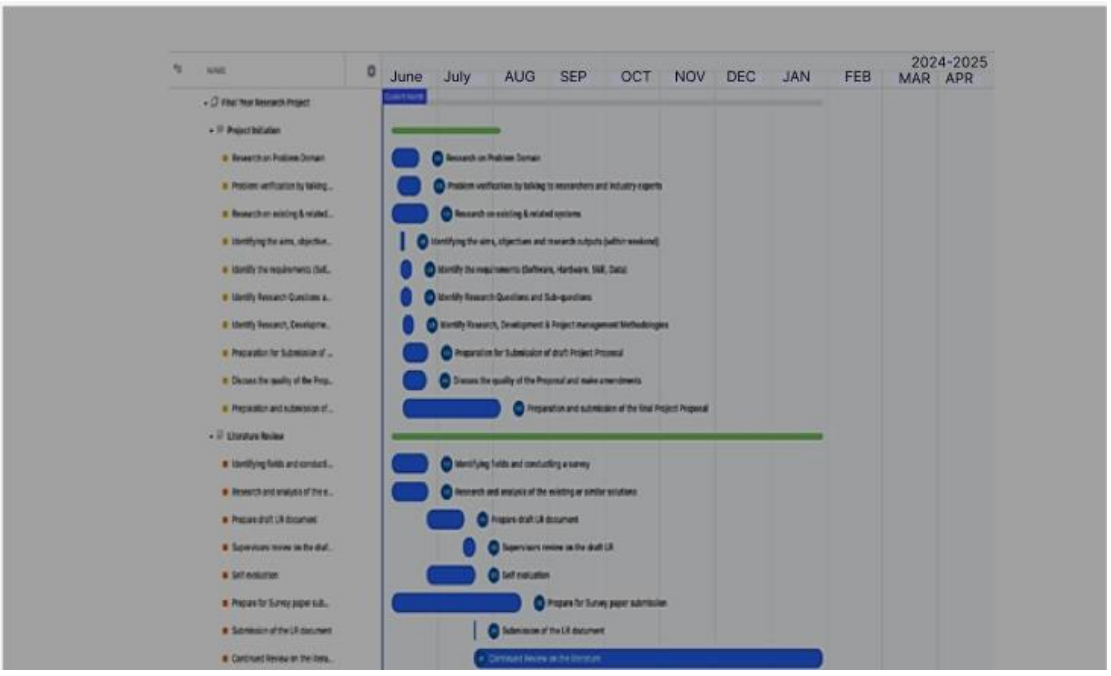
15.2 Deliverables

Deliverable Component	Tentative Delivery Date
Project Proposal	23/08/2024

Table 9: Deliverables

15.3 Grant Chart

Forecasting of Customer's Eligibility for The Bank Loan and, Also Predict Credit Limit, Using ML Approach



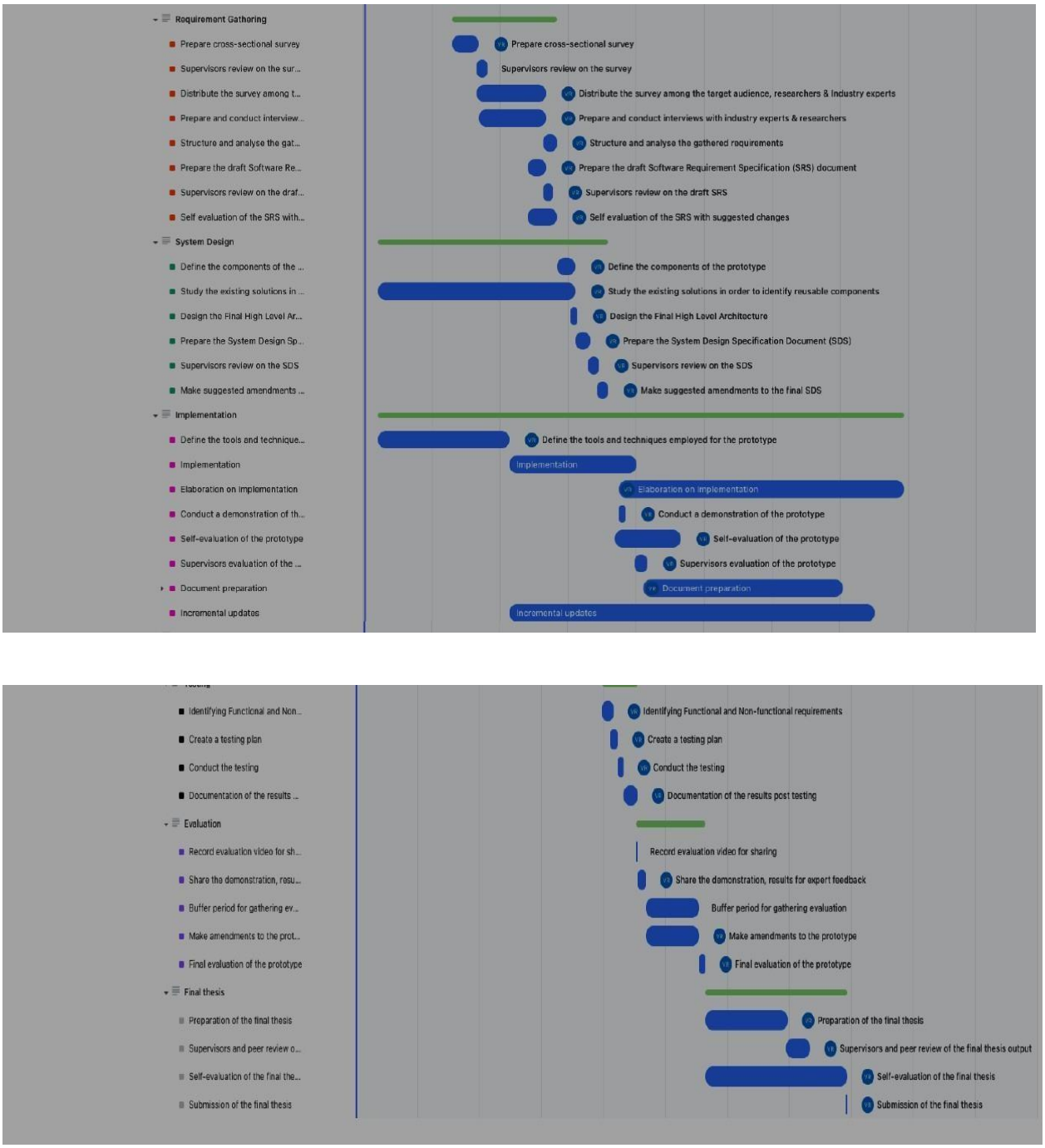


Figure 2: Forecasting of Customer's Eligibility for The Bank Loan and, Also Predict Credit Limit, Using ML Approach

15.4 Risk Management

Risk mitigation planning is the process of coming up with ideas and steps to enhance prospects and lessen risks associated with project issues. Monitoring the development of risk mitigation requires identifying obstacles, learning about new hazards, and gauging the effectiveness of risk management procedures throughout the project.

Risk	Severity	Frequency	Mitigation Plan
Insufficient domain Knowledge.	2	1	Get the assistance of domain experts, read journals, books, related to domain, interviewing domain experts and study the previous related researches also study the existing works.
Unable to finish all anticipated deliverables by the dates specified.	4	3	Give priority to deliverables and used time tables and Trello boards to mange the works. Also focus the major tasks.
Delays brought on by unavoidable things, such as inevitable circumstances such as illnesses or difficulties, have an impact on the project's success.	4	2	Prepare in advance and allow enough time for a buffer to ensure that unforeseen issues won't significantly alter the analysis's main course.
Power interruptions and connectivity issues	2	1	Used extra power sources and always try to keep daily or weekly goals.
Privacy and policies of the banking industry is very high. therefore, customers data and bank policies are very sensitive and risky.	2	1	Discussed with domain experts and identify secure methods also try to maintains privacy and ethical things.

Documentation Issues	5	4	Use word documents backups daily and also use the cloud backups for the documentation things

Table 10: Risk Management

16. Summary

The paper published presents a proposed project proposal related to upcoming research and a proposed solution. Research area, problem background, technical background and author's contribution discussed in this paper.

17. Reference

1. R. Manglani and A. Bokhare, “Logistic Regression Model for Loan Prediction: A Machine Learning Approach,” 2021 Emerging Trends in Industry 4.0 (ETI 4.0), May 2021, doi: https://doi.org/10.1109/eti4.051663.2021.9619201.

2. C. N. Sujatha, A. Gudipalli, B. Pushyami, N. Karthik, and B. N. Sanjana, “Loan Prediction Using Machine Learning and Its Deployement On Web Application,” IEEE Xplore, Nov. 01, 2021. https://ieeexplore.ieee.org/abstract/document/9696448

3. K. Aditya Sobika, M. Jamuna, H. Limitha, M. Saroja, and S. Shivappriya, “An Empirical Study on Loan Prediction Using Logistic Regression and Decision Tree,” 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Oct. 2021, doi: https://doi.org/10.1109/icaeca52838.2021.9675505.

4. G. Arutjothi and C. Senthamarai, “Prediction of loan status in commercial bank using machine learning classifier,” IEEE Xplore, Dec. 01, 2017. https://ieeexplore.ieee.org/abstract/document/8389442

5. Ch. Naveen Kumar, D. Keerthana, M. Kavitha, and M. Kalyani, “Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector,” 2022 7th International Conference on Communication and Electronics Systems (ICCES), Jun. 2022, doi: https://doi.org/10.1109/icces54183.2022.9835725.

6. Azzini, S. Marrara, Amir Topalović, Mirjiana Pejic Bach, and M. J. Rattigan, “Opportunities and Risks for Data Science in Organizations: Banking, Finance, and Policy - Special Session Overview,” Oct. 2018, doi: https://doi.org/10.1109/dsaa.2018.00078.

7. Padimi, Venkata Sravan, and D. D. Ningombam, “Applying Machine Learning Techniques To Maximize The Performance of Loan Default Prediction,” Journal of Neutrosophic and Fuzzy Systems, pp. 44–56, 2022, doi: https://doi.org/10.54216/jnfs.020204.

8. Alrabiah and M. M. Elgammal, “Optimal regulation of banking system’s advanced credit risk management by unified computational representation of business processes across the entire banking system,” Cogent Economics & Finance, vol. 6, no. 1, p. 1486685, Jan. 2018, doi: https://doi.org/10.1080/23322039.2018.1486685.