B213025B

BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE

AN EXPLORATION OF CREDIT RISK ASSESSMENT ON LOAN APPROVAL PROCESS USING MACHINE LEARNING TECHNIQUES

SUBMITTED BY JULIET R DOKA (REG NO. B213025B)

SUPERVISOR: MR H CHIKWIRIRO

YEAR: JUNE 2025

A project proposal is submitted to Bindura University of Science Education in partial fulfillment of the requirements of the Bachelor of science Honors degree in Computer science

B213025B

APPROVAL FORM

The undersigned certify that they have supervised the student Juliet R Doka's dissertation entitled an exploration of credit risk assessment on loan approval process using machine learning techniques submitted in Partial fulfilment of the requirements for the Bachelor of Computer Science Honours Degree of Bindura University of Science Education.

Juliet R Doka	J.Doka	02/07/2025
STUDENT	SIGNATURE	DATE
Mr. H. Chikwiriro		16/10/2025
SUPERVISOR	SIGNATURE	DATE
Mr. P. Chaka		16/10/2025
CHAIRPERSON	SIGNATURE	DATE

DECLARATION

As the sole author of this dissertation, I, Juliet R. Doka, certify that it is the product of my own independent work. I agree that Bindura University of Science Education may use this dissertation for research or academic purposes, including lending it to other people or organizations for scholarly purposes.

ACKNOWLEDGEMENTS

I am very grateful to the Almighty Lord for giving me the chance, wisdom, and strength to finish this dissertation. By His grace, I was able to successfully finish my academic path while maintaining a healthy physical and mental state.

I want to sincerely thank Mr. H. Chikwiriro, my supervisor, for his tremendous advice, knowledge, and continuous support during this research endeavor. The composure, gratification I have in this day, looking at the remarkable success that the project has, has been brought up by the assistance received. I have not only learnt project related concepts from him, but this also includes the art of doing things in time and persistent consistence to do the work, deliver quality and achieve academic excellence.

Special thanks to Mr. G. Chiridza for guiding me, your insight and expertise were invaluable, and I appreciate the time and effort invested in helping me to achieve my goals.

Furthermore, distinguished appreciation also goes to the management of Bindura University of Science Education (BUSE) and the participants who generously shared their time and experiences. Many thanks also go to Mr. Zano, Mr. P. Chaka and Mr. D. Ndumiyana (Industrial attachment supervisor) for their unending support.

Finally, I want to express my gratitude to my colleagues and support system for making this journey possible. It is impossible to overlook their spiritual and moral assistance. We would especially like to thank Linda Satimburwa, the CEO of Global Hand of Hope International (GHOHI), and her team for their ongoing financial assistance.

DEDICATION

This effort is dedicated to my mother and my siblings, Collen Bheurani, Faith Doka, Shanon Doka, Shakemore Doka, and Ruth Doka. We greatly valued the encouragement and assistance provided.

ABSTRACT

A loan is a sum of money that can be obtained from financial institutions with the understanding that the borrower will repay the loan within a predetermined time frame and with interest. After accruing debt, the borrower is required to pay it back in full, typically with interest. The model will be trained to predict whether the applicant will be able to pay back the loan. There are certain features that have to be taken into consideration before approving the loan. Automating the procedure is the primary goal in order to process loan applications more quickly. Train and test data sets were used as the input. To determine the accuracy of the model, the training data was used, and to output the predictions, test data was used. To reach the scope of our goal, the researcher has used machine learning tools and algorithms as well as Python, Jupyter notebook and Django to develop the system.

Table of Contents

CHAPTER 1: PROBLEM IDENTIFICATION

1.1 Introduction

The banking division is an essential element of each economy. Banks act as intermediaries between borrowers and savers, delivering banking facilities to people, businesses, and public sectors. Loans are crucial especially in economic growth (King, et al., 2001). People and businesses can invest in new opportunities, buying houses and cars, and fund to other purchases that they cannot afford without borrowing. Taking out loans can help the economy to expand business operations and employing more staff. Aside from allowing access to capital (credit) banks are also responsible for risk management. Borrower's creditworthiness and the possibility of loan repayment can be taken into consideration by banks through thorough investigation of the public and private data that is accessible (Samreen,2012). This lowers the risk of non-payment by allowing debtors to identify the danger of failure with a yield. Based on the status of applicant, loans are approved according to lending policies created by banks.

According to the applicant's credit policy, the loan application is assessed. Through thorough verification of submitted documents, banks can approve loan applications. However, there is no guarantee that all applicants are the best (Sheikh et al., 2020). By investigating large amounts of data and processes, tracking of patterns and trends, machine learning can help in automating loan approval process and eliminating human error.

A key benefit of machine learning is its ability to leverage prior data to learn and predict future results (Srivastava et al., 2018). By training machine learning models on previous loan data and results, based on a variety of factors, including a candidate's qualifications, gender, banking, job position, and other relevant information, the loan approval process may become easy for financial institutions. These machine learning models can also identify important risk factors that can help clarify the application process, by highlighting high-risk applications requiring human review (Al Mamun et al., 2022). In this way, machine learning can make faster and accurate loan decisions without the need for a manual review, provide personalized recommendations based on applicants, and improve the client's overall experience.

1.2 Background of the Problem

Lending business faces difficulties inaccurately assessing loan applications, which can lead to significant non repayment rates and financial losses. According to a report by Companies and Markets in September 2012, Lobels Limited was indebted to five local banks FBC Bank, CBZ Bank, NMB Bank, Metbank, and Capital Bank with outstanding loans totalling approximately $14 million. The company reportedly refused to repay the debt. Methods that are used for loan evaluations on the existing system are time consuming and biased. In the era of digitalization when loan applications are becoming numerous, there is a constantly growing need for automatic systems that would be able to evaluate loan possibilities both faster and more precisely. Number of factors, such as income, and credit history and economic factors affected loan eligibility. The existence of such broad ranges of influencing aspects can cause inconsistency in the lenders' decisions and jeopardize their business as long as the factor of risk is concerned. Hence, the purpose of the research is to apply modern algorithms that predicts the loan decision process to ensure efficient, bias-reducing, and able to assist in preventing any loan default.

1.3 Statement of the problem

Many banks still sit down with stacks of paper and check boxes to decide who qualifies for a loan. That hands-on grind slows everything, invites mistakes, and leaves too many borrowers in the danger zone. Old-school scoring formulas also miss the tangle of traits-an extra income source here, a short credit dip there- that actually tip the balance. The crush of new applications and the constant churn of credit data just keep turning that problem up to eleven. This project swaps ledgers for algorithms, building a machine-learning model that spots good risks and bad ones faster, cuts the default bleed, and frees lenders to make sharper calls.

This proposed system will use demographic data, credit history, financial and employment information in order to predict the odds of loan repayment. By using clever data processing techniques that uncover hidden insights and subtle relationships in borrower information, the system enhances credit evaluation. Financial institutions can streamline their lending operations, reduce risk and optimize resource allocation because of this system’s scalability and interpretability. Ultimately, this project aims at narrowing down the difference that exists between traditional credit scoring methodologies as well as advanced predictive analytics which ultimately leads to a more informed and effective lending climate that is advantageous to both lenders and borrowers.

1.4 Research Aim and Objectives

To develop and evaluate a forecasting model based on machine learning techniques to accurately predict loan eligibility based on applicant characteristics, financial history and creditworthiness, thereby improving the productivity of the loan approval process.

1.4.1 Research Objectives

The study aims to:

- Investigate and analyze the relationships between borrower attributes and loan approval rates.
- Develop a machine learning model that predicts loan eligibility based on a diverse set of borrower attributes.
- Evaluate the model's effectiveness for loan applications.

1.5 Research Questions

- How accurately can a machine learning model predict the eligibility of loan applicants?

- Can an automated prediction system improve the speed and efficiency of loan processing compared to traditional methods?

- How effective is the model in minimizing loan default risks by identifying high-risk applicants?

1.6 Research Hypothesis/propositions

- H_0: Using selected features, the model cannot predict loan eligibility rightly.
- H1: Using selected features, the model can predict loan eligibility rightly.

1.7 Justification/Significance of the research

This study is noteworthy because it offers a data-driven remedy for existing problems with the loan approval procedure, which benefits financial institutions and consumers alike. Customers will benefit from quicker processing times and more equitable assessments, while financial institutions will experience increased operational efficiency and lower loan default rates. Additionally, by decreasing discriminatory practices in loan decisions, the model will conform to regulatory standards. Because it demonstrates how machine learning can improve financial services, this is significant for the financial technology (FinTech) industry.

1.8 Assumptions

- This study will use feature engineering to select and transform relevant features for credit risk assessment including demographic information, credit history, income levels and financial behaviors.
- Existing regulatory frameworks and ethical standards in financial lending will be adhered to when using machine learning to the credit risk assessment during the loan approval process. This includes laws pertaining to data protection and fair lending practices.

1.9 Limitations/ Challenges

- Adhering to regulatory requirements and guidelines is a critical challenge. Financial institutions must ensure that their use of machine learning complies with legal standards, which can be complex and vary by jurisdiction.
- While machine learning models can be highly effective, scaling these solutions across various loan products and borrower segments can be challenging. Each segment may require tailored approaches, complicating the implementation of a one size fits all solution.
- New technologies are challenging to incorporate and apply within current financing systems.

1.10 Scope/Delimitations of the Research

In particular, this study will focus on the default risk assessment component of the loan approval process and investigate two algorithms, such as logistic regression and linear regression. This study's scope encompasses data gathering, attributes processing and model development. The study will not go into great detail regarding other risk elements, such as operational market and liquidity issues, despite its concentration on credit risk assessment. Again, a comprehensive examination of all regulatory frameworks in various jurisdictions will not be carried out by the study.

1.11 Definition of terms

Big data-refers to large and information set data that keep growing rapidly over time.

Regulatory compliance-laws, regulations and guidelines that govern financial practices and lending standards.

Predictive Modeling- the process of using statistical ways to create a mock-up that predicts future outcomes based on historical data.

Default evaluation- the process of checking the probability of the loan receiver loan repayment in order to determine the likelihood of default.

Machine learning- is a branch of computer science that concentrates on creating algorithms capable of analyzing data and making predictions or decisions based on that data.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Credit risk assessment is a critical component of the loan approval process in the banking and financial industries. To reduce losses and increase earnings, borrowers' creditworthiness must be accurately estimated. As machine learning algorithms have advanced and data has become more accessible, researchers have looked into a variety of methods for evaluating credit risk. Credit risk analysis based on machine learning has been studied by various scholars. Random forest and logit boosting were employed by Belhadi et al. (2021) to forecast small and medium businesses' creditworthiness. According to ZIMSTAT (2021), the formal banking sector has been slow to adopt advanced technological solutions for risk assessment, emphasizing the need for an innovative approach. This study mainly focuses on the application of linear and logistic regression models, benefits and challenges of loan approval prediction.

2.2 Importance of machine in credit risk assessment

Machine learning has revolutionized the way we assess credit risk globally. These powerful algorithms can process vast amounts of data, identify patterns, and accurately predict loan default rates. While there are many advanced techniques, logistic regression remains a popular choice due to its simplicity, interpretability, and ability to classify borrowers as high or low risk. Cutting-edge machine learning models can analyze large datasets to uncover hidden trends related to a borrower's creditworthiness, leading to more accurate loan predictions. This has proven to be a valuable management tool, particularly for evaluating the credit risk of small and medium-sized businesses. With its game-changing capabilities, machine learning has transformed the financial industry, making credit assessment more efficient and effective than ever before.

Unlike more complex models such as Random Forests or Deep Neural Networks, logistic regression provides a clear understanding of the relationship between borrower attributes such as income, employment, and repayment history and loan default probability. This transparency is essential for compliance and decision justification in regulated financial environments.

High levels of credit risk have resulted from Zimbabwe's economy's severe problems, which include macroeconomic instability, unemployment, and inflation. Effective credit risk assessment is essential to reducing the risks associated with loan defaults, according to scholars like Nyoni (2020). Current manual operations, as described by Chiri et al. (2021), often result in subjective decision-making that can jeopardize financial institutions' overall stability.

2.3 Models used

2.3.1 Linear regression

The connection among each input variables and an outcome variable can be identified based on the statistical technique According to Anyanwu (2021), linear regression may overlook financial datasets that contain complex relationships, even though it can function well for simple relationships. Although its role in credit scoring systems has been highlighted by scholars such as

Menzies (2014), its standalone applicability is frequently limited by its inability to model nonlinear relationships.

2.3.3 Logistic Regression

Loan default probabilities can be accurately predicted using this algorithm, regression approach that predicts two possible outcomes. According to Maliyamkono & Zimba (2020), logistic regression makes it possible to model events with a categorical dependent variable. The application of logistic regression in Zimbabwe's financial setting is covered in great detail by Ndlovu (2021), who highlights how its probabilistic nature might improve loan decision-making.

2.3.4 Advantages of predicting default risk

Increase Prediction Accuracy: Chen et al. (2019) claim that the problem of accuracy being jeopardised by conventional methods has been resolved with the advent of machine learning models. A more accurate evaluation of credit risk results from their capacity to analyse sizable datasets and spot patterns and trends.

Reducing Bias: Human biases in manual evaluations can be reduced by using machine learning methods (Ravi & Ramesh, 2021). This is necessary to promote equity in lending practices..

Data Utilization: ML models can integrate a lot of data from alternative data sources, such as utility payments and mobile money usage, which are crucial in Zimbabwe (Mudavanhu, 2022).

2.4 Gaps in the Literature

Although there is a wealth of research from around the world supporting the benefits of machine learning in default risk, there is a dearth of literature that specifically addresses the particular difficulties faced by financial institutions in Zimbabwe. Most local institutions are yet to transition from manual to automated systems, and few studies have tested the practical implementation of logistic regression within this environment. Furthermore, not much is known about incorporating non-traditional data sources into logistic models, like informal employment records or mobile money transactions. This gap presents an opportunity for the current study to develop and validate a logistic regression model tailored to the Zimbabwean context, using realistic borrower data and aiming to enhance objectivity and processing efficiency in loan approvals.

2.5 Debates and Evaluation

The implementation of machine learning-based systems in loan approval processes generates multiple debates about Accuracy and fairness, Automation, Regulatory Compliance and Privacy. According to the ProPublica Report (2016), the issue of fairness in machine learning decision-making continues because these systems may reinforce or preserve discriminatory biases present in their training data. Because machine learning systems need to collect a lot of data, including personal and alternative data sources, their use in loan eligibility approval systems raises privacy concerns. The evaluation of these controversies helps to understand the wider implications of machine learning adoption in loan approval processes.

2.6 Relevant Theories of the subject matter

Logistic and linear regression have been used by numerous studies to evaluate credit risk, which has helped predict loan eligibility. However, it may be difficult to integrate traditional methods with machine learning (ML) frameworks, nevertheless, because of model scalability, processing speed, and durability in real-world situations. In their thorough research, Smith and Johnson (2020) emphasized the efficiency of logistic regression in forecasting borrower default. Because logistic regression takes into account non-linear relationships in the data, they discovered that it performed better than conventional scoring techniques. However, their study did not evaluate the impact of logistic regression on decision-making speed or processing efficiency, evaluating the model's effectiveness in reducing loan processing time. Kim and Lee (2019) explored logistic regression as a predictive tool for loan default risk, identifying key variables such as income, debt-to-income ratio, and credit history that significantly influenced loan approval outcomes. Their results showed that adding these variables improved the accuracy of the model. However, they did not made comparison on logistic regression with more complex models. In their study, Liu et al. investigated various statistical techniques, including logistic regression, when testing credit risk models. These authors note that with respect to their findings, logistic regression managed to exhibit relatively high performance, mainly in the case of significant class imbalance in datasets. This allowed for improved patterns in the decision-making process of loans. Nonetheless, the authors have not investigated if these patterns may be applied to real-time data or if using these models should have resulted in speedier processing. Regarding the evaluation of credit risk, Chen and Zhang contrasted logistic regression with machine learning. They claimed that while more complex machine learning methods tended to show better prediction capabilities, the fact that logistic regression was more interpretable meant that the lenders preferred it to other methods. Nevertheless, they did not include a model in the comprehensive machine learning pipeline that can ensure a balance between interpretability and predictive potential when automating the process.

Furthermore, they did not incorporate feature selection and model tuning, nor did they assess the relative effectiveness of logistic regression. Adams (2022) used both logistic regression and linear regression to perform a multifactor study of loan approval criteria. The study found that by precisely defining the approval versus rejection dichotomy, the logistic regression model provided lenders with valuable information. He did not, however, apply model tuning or assess performance indicators for various algorithms. The primary goal of Torres and Dias (2020) was to apply logistic regression to forecast borrower defaults. Their findings supported the idea that logistic regression performs better than conventional techniques, particularly in a range of economic scenarios. Nonetheless, the study did not examine broader contexts. Martinez (2019) looked at what is new in credit risk modeling which he also put out there that although we still use traditional models which are for the most part relevant, when we integrate them with machine learning that is when we see an improvement in what we predict. However, he failed to provide an empirical study or implementation strategy, which made it difficult to assess the veracity of his assertions. Also in

2021, Hernandez and Santos looked at how machine learning models do against the traditional regression techniques in credit scoring. What they found is that although some of the more complex machine learning models do report high accuracy, logistic regression is still very much the go to for institutions that need to be able to explain what is going on. Also, they didn't look at how these models do in terms of speed or with very large data sets. In their 2023 study on the application of machine learning to credit risk, Patel and Gupta made the simple but accurate claim that, with the right adjustments, logistic regression is an excellent tool for predicting loan eligibility. Moreover, their research showed model's output must aligns with best practices within the industry. It lacks thorough model comparisons and consideration of other tuning techniques makes their report ineffective.

2.7 Limitations of the existing system

The existing processes often led to a series of limitations that can adversely affect both financial institutions and their clients. Here are several key limitations associated with reliance on manual procedures in the loan approval process:

Inefficiency and Delays: Loan pipelines still move at a human pace. Under a manual model, someone has to collect papers, check them twice, calculate once again, and finally rubber-stamp the green light. Borrowers who thought they had heard back in a week often sit in silence for a month or more, and that wait wears on their patience. When the clock stretches, many people simply go across town and sign with a rival lender-or post on social media that their bank blew it.

Increased Human Error: A person typing numbers is never infallible. A stray finger can turn a six into an eight, and suddenly an applicant's income looks ten percent higher than it is. File clerks may skip a pay stub or misread a tax line, which tilts the risk equation toward a default or an unfair rejection. Every slip not only bruises customer trust; it also adds a sneaky layer of exposure the bank must quietly absorb.

Limited Data Utilization and Analysis: One problem is that manual processes do not allow for full utilization, and analysis, of the vast amounts of data available. Instead, decisions are made on handpicked information or out-of-date records. The lack of data-driven decision making can affect the accuracy in assessing credit risks that a financial institution may have. Furthermore, lenders miss out on potential insights about borrower behaviors by failing to use advanced analytics that could incorporate alternative data sources such as bank statements and patterns of income, which could lead to defaults or missed lending opportunities.

Scalability Issues: Often manual loan approval procedures cannot be scaled up easily specifically during periods of high demand or economic recovery when there is an influx of loan applications. Financial institutions may be overwhelmed by the number of applications causing delays and backlogs in processing times. By being unable to handle increased workloads efficiently, companies might lose business because borrowers would rather go to other institutions that provide quicker responses

Ineffective Risk Management: Traditional system is usually not advanced enough to carry out sufficient risk assessments. It becomes hard for financial institutions to identify high-risk borrowers without systematic approaches. This can lead to higher default rates as lenders may unknowingly approve loans for people who will not repay them. Eventually, such an outcome can significantly affect the stability and reputation of the institution. Poor record keeping and data management: Manual methods often result in disorganized records where physical documents get lost or misplaced while digital records are not effectively integrated.

Poor data management: makes it difficult for loan officers to quickly retrieve important borrower history and documents due to limited access. This has the potential of prolonging processing times as well as complicating compliance with regulatory requirements concerning record-keeping and audit trails

Financial institutions in Zimbabwe should be aware of these constraints and look for ways to get around them by embracing technology and implementing more sophisticated and automated loan approval procedures. They can increase operational effectiveness, boost customer satisfaction, and maintain their competitiveness in the ever-changing banking sector by switching to more automated and data-driven systems.

2.8 Conclusion

To sum up, frameworks and languages were identified to support the system's successful development before the model was developed. By using the previously developed techniques to identify the gaps that need to be filled, the author would be able to emphasise how much more efficiently the new system will operate. The author has examined the prior systems that were in place, reviewed the problems they were meant to address, and noted any gaps in the research in this chapter. The researcher assessed the benefits to users in the future as well as the advantages and disadvantages of the suggested solution.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This chapter serves as the backbone of any scholarly inquiry, providing a structured framework that guides researchers in their quest for knowledge. It delves into the essential components of research methodology, elucidating the systematic processes that underpin the collection, analysis and interpretation of data. Researchers can guarantee the validity and reliability of their findings by establishing a clear methodological approach, which contributes to the body of knowledge in their field. The researcher will examine a variety of research designs in this chapter, including population and sample, research instruments, qualitative and quantitative approaches, and data analysis techniques to be employed.

3.2 Population and sample

The study population includes all the people and businesses who have applied for loans from financial institutions. It varies with credit profiles, income, and loan requirements.

3.2.1 Sample

A sample of loan applicants was selected from publicly accessed dataset that is Kaggle and Claxon microfinance. It was stratified based on credit score, loan amount, and loan type to ensure representation from segments.

3.3 Data collection approaches

The investigator employed a variety of techniques to collect data regarding the current system, including surveys, observations, and interviews. The author used the information and data collected to design and develop the system.

3.3.1 Interview

Two accountants from Bank BancABC and CBZ were privately interviewed, during which the researcher posed questions, and the interviewees provided their responses. This process helped the researcher gather insights into the existing system. Both accountants shared their concerns regarding the current setup and suggested improvements to enhance their operational efficiency.

3.3.2 Observations

The researcher was able to obtain beneficial information by observing the current system in the National Building Society where she conducted her internship. The observation involved how loan information was handled and the procedures that were used in loan application. This enabled the researcher to determine the most frequent challenges that the applicants experienced.

3.3.3 Questionnaires

The researcher managed to obtain feedback on the questionnaires from numerous individuals by sending them through online media. The questions applied in the survey are presented in the appendix entitled "Loan application process." Through this survey, the author managed to obtain thorough details regarding the loan application process.

3.4 Research Design

It is a roadmap on which research is being carried out. According to Saunders, et al. (2012), research design is a plan aimed at answering a specific research question.

The researcher will use python language with machine learning techniques to build the system. With the aid of documentation and record to gather data about loan eligibility prediction model and machine learning algorithms, the author was able to carry out the research.

3.4.1 Requirement Analysis

It is required to identify the most significant data points to ascertain a borrower's creditworthiness as well as documenting functional and non-functional system requirements. The study also identified potential constraints, including dataset availability and computational power, that might jeopardize the success of the proposed methodology.

3.4.2 Functional Requirements

- Generate a probability score indicating loan approval likelihood.
- Clear display of loan eligibility prediction results.
- Identify high risk loan applications based on predictive analysis.

3.4.3 Non-Functional Requirements

- Protect sensitive financial data with robust encryption.
- Ensure data accuracy and fast response time.
- Ability to integrate with existing loan origination systems.

3.4.4 Hardware Requirements

- A 64-bit processor
- 8 GB RAM

3.4.5 Software Requirements

- Windows 10 Operating system
- Pycharm2024

3.5 System Development tools

Python 3.6.3-The researcher chose to use Python for its vast number of tools that are well-suited for machine learning. According to the Python Software Foundation (2021), Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it highly attractive for rapid application development, as well as for use as a scripting or glue language to connect existing components.

PyCharm IDE- According to Wikipedia (2018), PyCharm is a software application that provides a better environment for programmers which supports Python language.

Jupyter Notebook-It is web application that allows programmers to create and share documents.

SQLite database- SQLite is a lightweight, self-contained database that is simple to use.

Django- it is an open and free platform that excels in user authentication and database support.

3.6 System Design

This section comprises of UML (Unified Modelling Language) design for the system to clearly show the static view of the system.

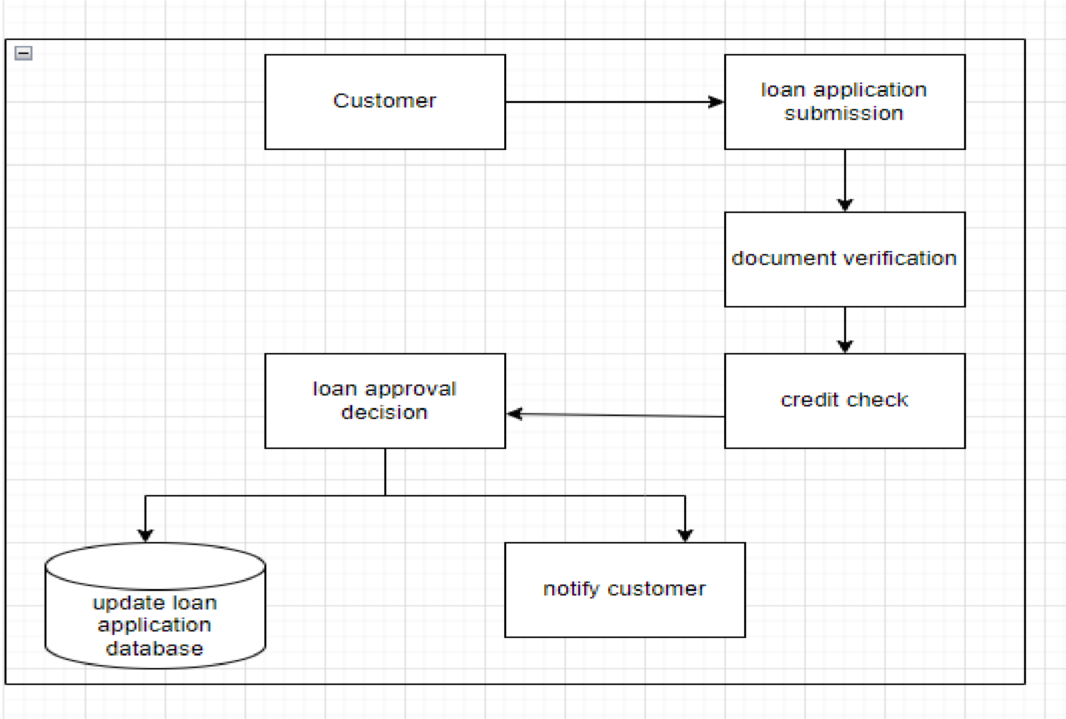


Figure 3.3: Data flow diagram for the model

3.6.1 Description

Customer: The individual or business applying for the loan

Loan Application filing: The customer submits their loan application along with required documents.

Document verification: The bank verifies the submitted documents for accuracy and completeness.

Credit check: A credit check is performed to assess the customer's creditworthiness.

Loan approval decision: The system reviews the application and makes decision on approval or rejection.

Notification of the decision: The customer is notified of the loan approval or rejection.

Loan application database: Contains records of all submitted loan applications and their statuses.

3.6.2 Flow chart for the proposed system

Programmers and end users can effectively communicate with each other by using flowcharts. They are flowcharts specialized in distilling a significant amount of data into comparatively few symbols and connectors.

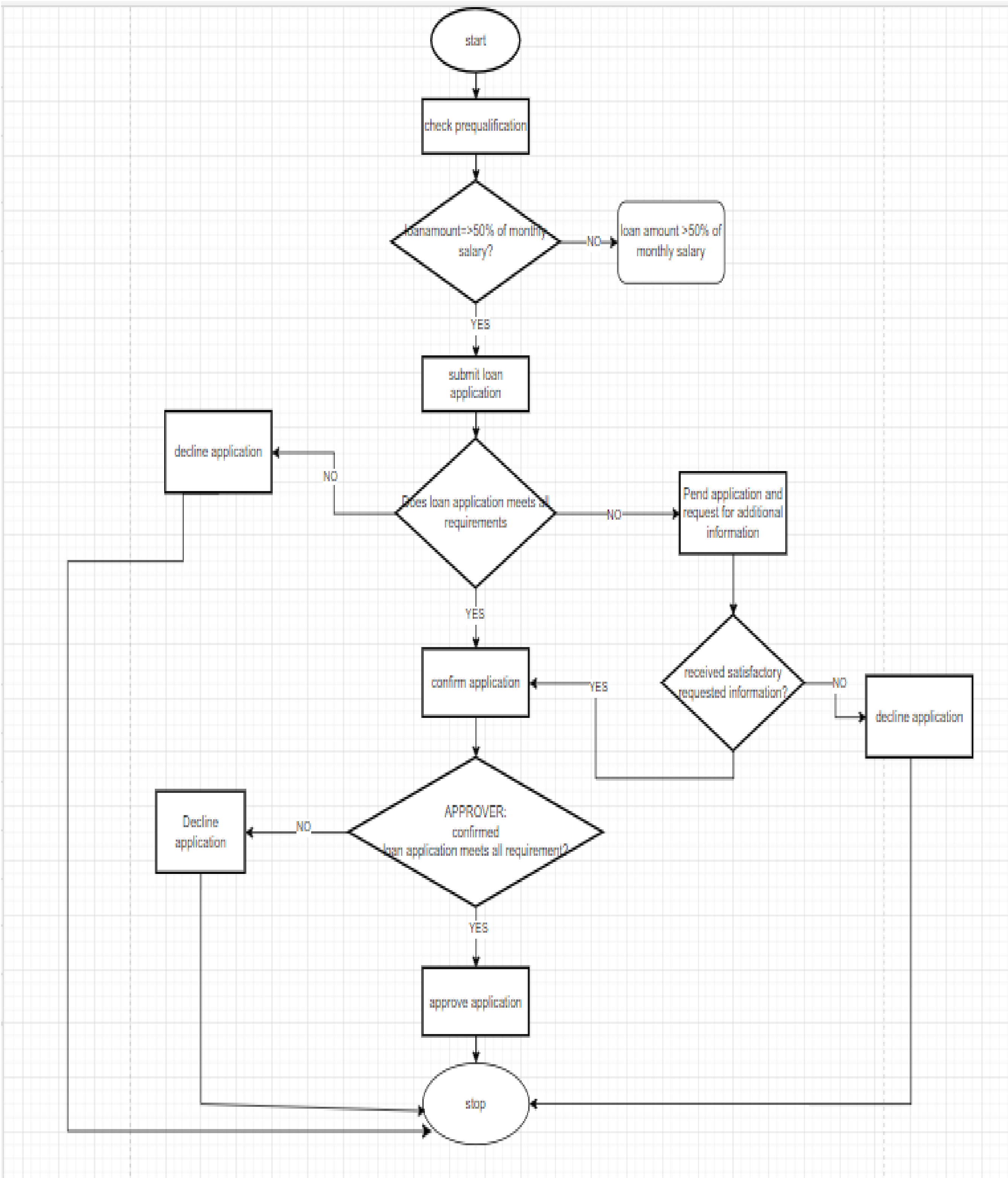


Figure 3.4: Flow chart for the mode

B213025B

3.6.3 Use case diagram for the proposed system

A use case diagram is diagrammatic presentation where users interact with the system and show the link between the user and the different activities in which the user has been acted upon. Below is the system's use case diagram.

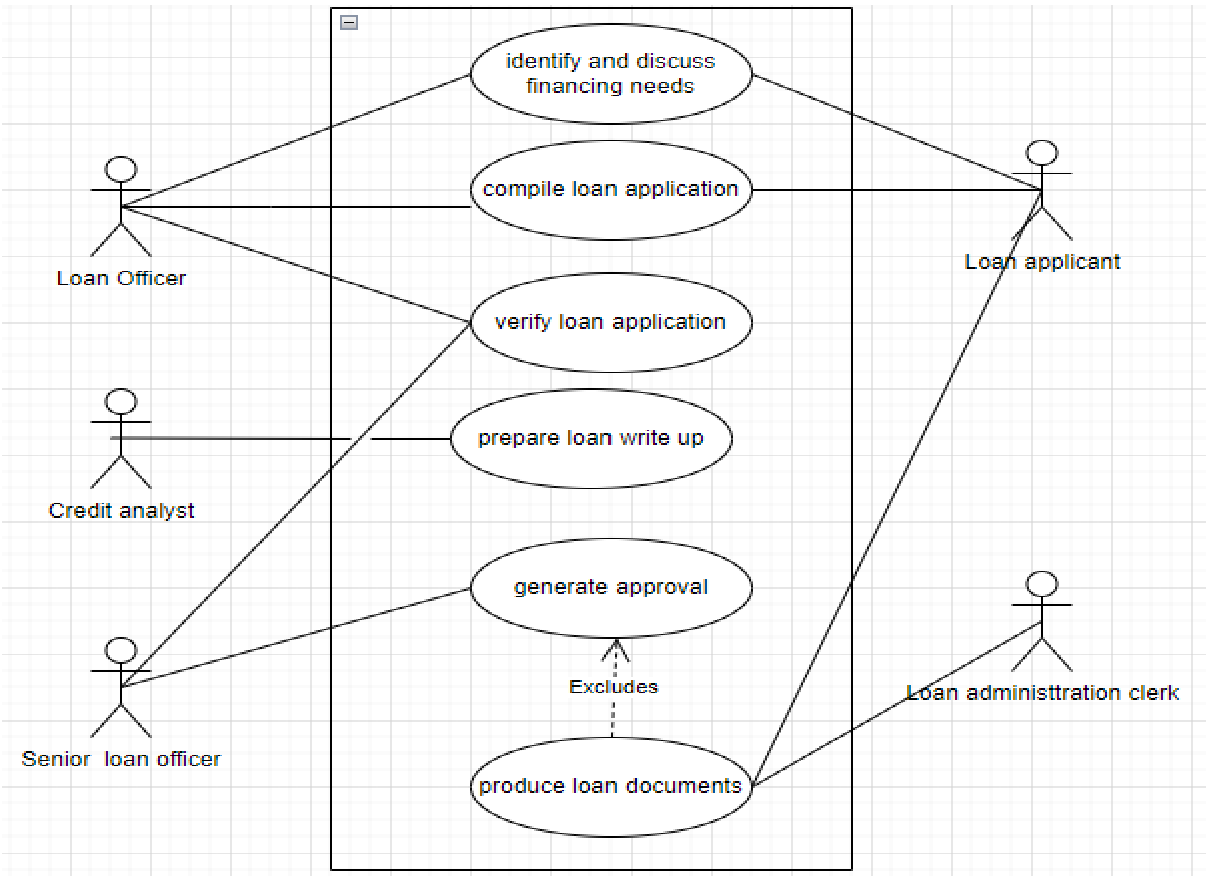


Figure 3.5: The Use case diagram for loan application process

3.6.4 Class diagram

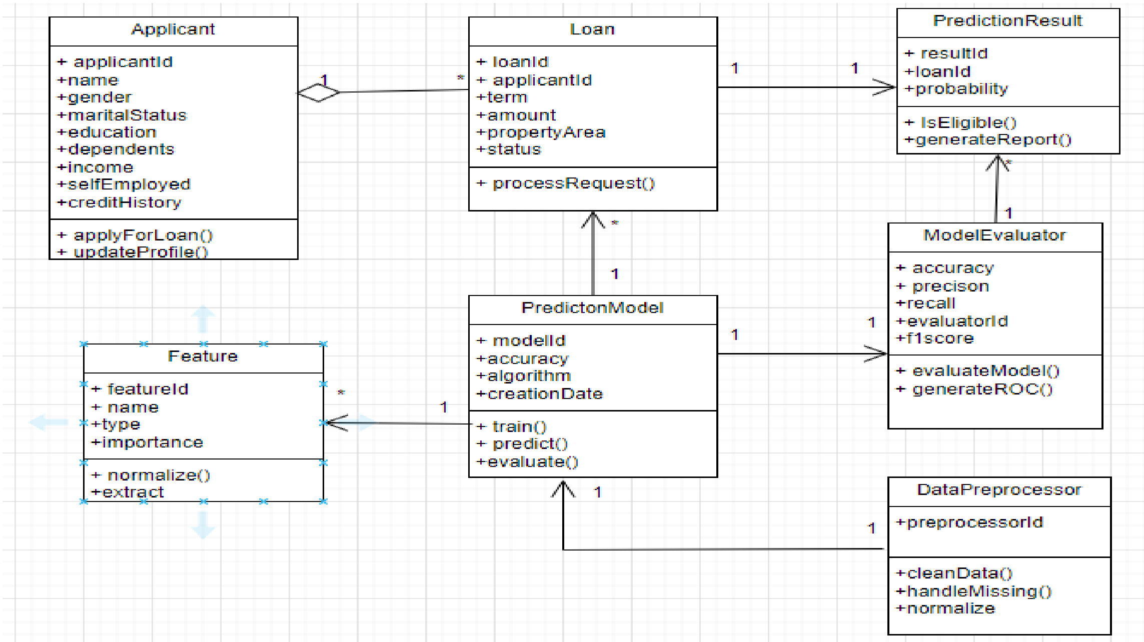


Figure 3.6: The class diagram showing relationship and association between objects

3.6.5 Description

1. Applicant to Loan (one-to-many relationship)
   - One applicant can apply for multiple loans.
   - Type: Aggregation (Loans can exist independently of the Applicant)
2. Loan to Prediction Result (one-to-one relationship)
   - Each loan application has one prediction result
   - Type: Direct Association
3. PredictionModel to Feature (one-to-many relationship)
   - A prediction model uses multiple features
   - Type: Direct Association
4. PredictionModel to PredictionResult (one-to-many relationship)
   - A prediction model generates multiple prediction results
5. **5.** DataPreprocessor to PredictionModel (one-to-one relationship)
   - The data preprocessor prepares data for the prediction model

- Type: Direct Association

6. ModelEvaluator to PredictionModel (one-to-one relationship)

- The model evaluator assesses the prediction model
- Type: Direct Aggregation

3.7 Development model

The researcher is going to use Agile development as it provide a room for the developers to adapt changes in requirements while ensuring that user needs are continuously met through regular feedback cycles.

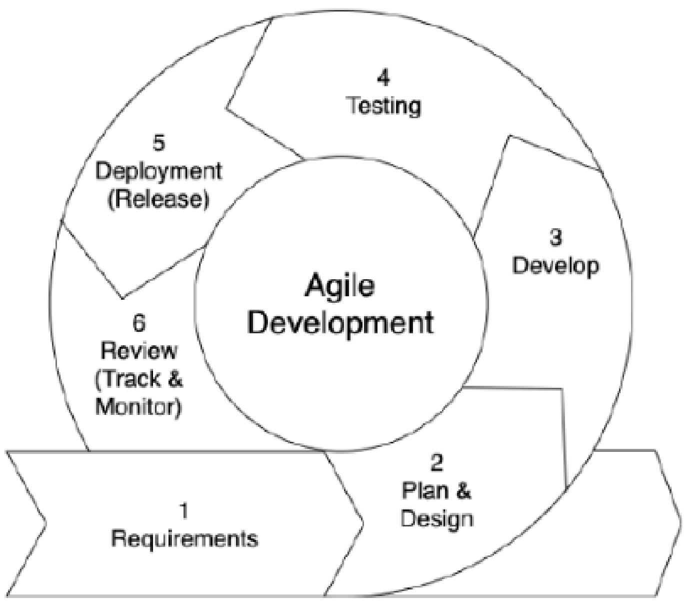


Figure 3.7: Agile Development process

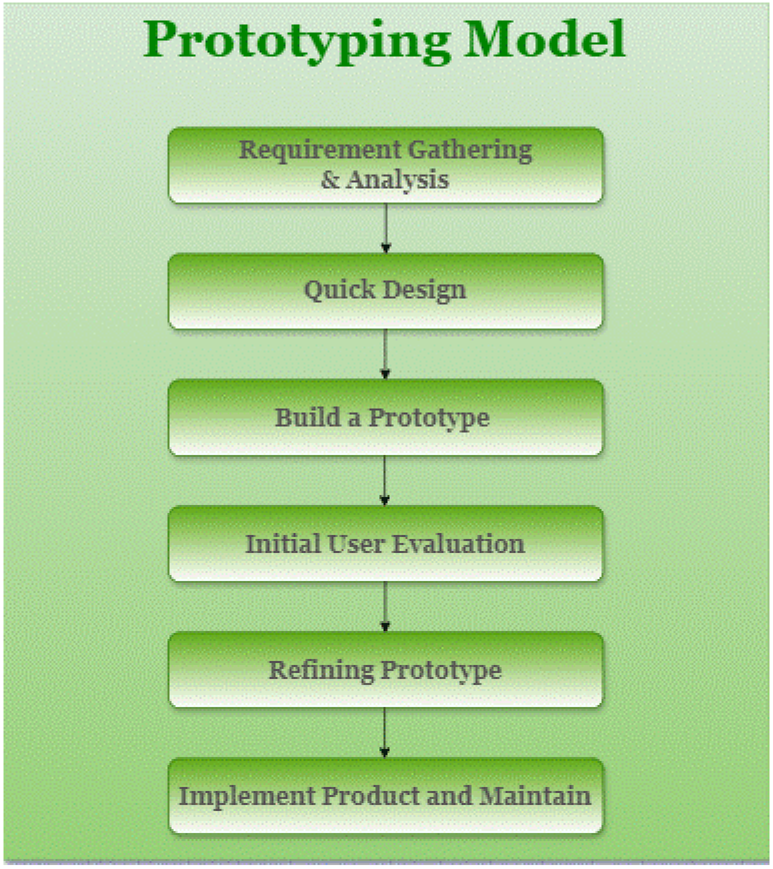


Figure 3.8: Prototyping Model

Requirement gathering and Analysis

Interviews were conducted with system end users in order to get all user requirements. Once requirements were gathered, the author compiled the expectations of system users (Pressman, 2014).

Quick Design

The second stage is rapid or preliminary which gives the user an overview of the system (Sommerville, 2011).

Build a prototype

In this phase, an actual prototype was designed based on the information gathered from quick design and this small working model helped the researcher to see the feasibility of the study (Kendall and Kendall, 2010).

Initial user evaluation

After the prototype building is finished, the client must perform an initial evaluation to find any possible defects. The developer could use this information to fix the bugs.

Tweaking/refining prototype

If the user is not satisfied with the previous model, the author should make changes depending on their comments and ideas. Once the customer is happy with the mock-up, a final architecture will be developed depending on the results (Sommerville, 2011).

System enhancement

The final architecture will be tested, deployed and maintained to minimize system unavailability (Martin, 2021).

3.8 Feature Engineering on model development

On the feature engineering process, the author used visualizations charts in order to understand how each variable contributed to the model's decisions. The following charts are used.

Using matplotlib

Below is bar chart of total number of loans defaulted versus none defaulted.

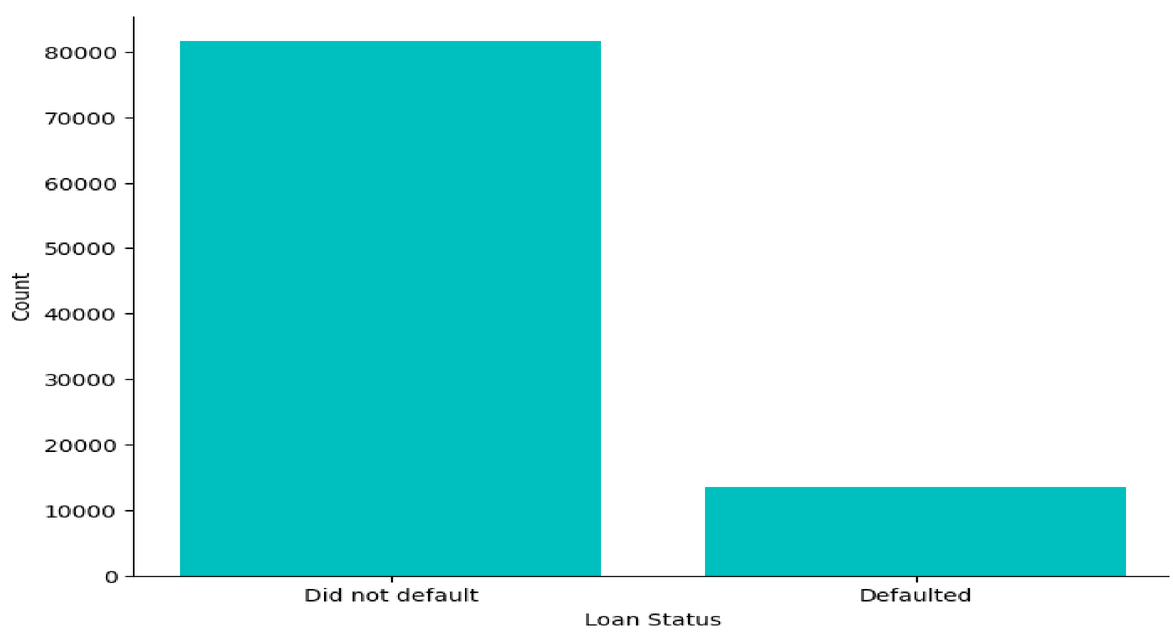


Figure 3.9: Bar chart showing total number of loans defaulted vs none defaulted

This bar chart clearly shows that more loans were not defaulted compared to those that were defaulted. The significant class imbalance highlights that most applicants in the dataset were good payers.

Pie Chart

The author visualizes the percentage of those who defaulted versus those who didn't. The flag autopct to add the percentage and specify the decimal floating-point number.

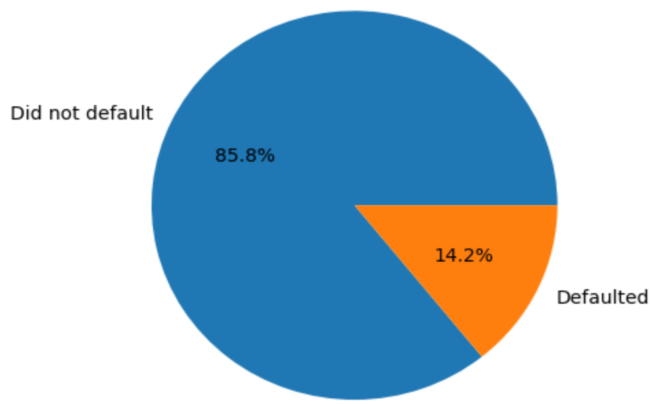


Figure 3.10: Pie chart showing the percentage of defaulted vs none defaulted

It complements the bar chart by visually confirming that the majority of applicants did not default, possibly over 70%.

Using seaborn

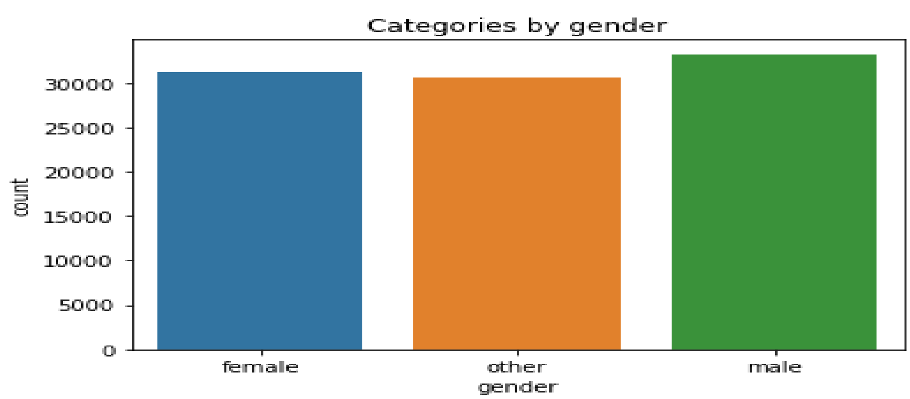


Figure 3.11: Loan approval rate by gender

This graph shows the loan approval distribution across gender categories. Male applicants had the highest approval rate, followed by female applicants, while 'Other' (non-binary or unspecified gender) entries if present were minimal or underrepresented. This visual trend suggests that gender may have an influence on loan decisions, though further fairness evaluation is necessary to ensure unbiased model prediction.

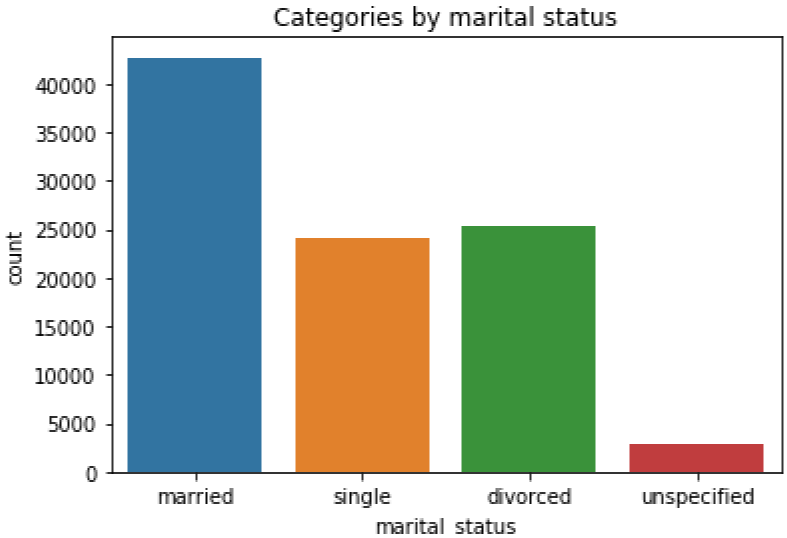


Figure 3.12: Loan approval rate by marital status

Married applicants showed a higher approval rate than single applicants, suggesting that marital status is a predictive feature in loan assessment. This could be due to perceived financial stability or dual-income potential often associated with married individuals. However, it's important to ensure that this feature does not unfairly disadvantage single applicants, especially if unrelated to financial capacity.

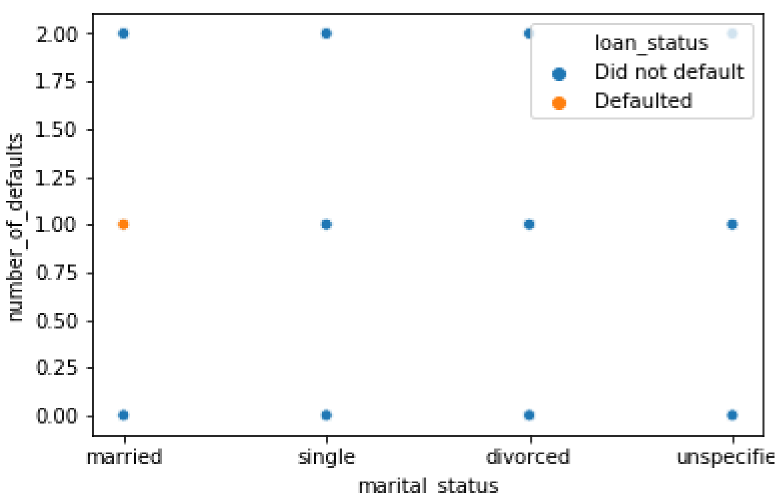


Figure 3.13: Scatterplot showing correlation variables involved

The above displays a scatterplot comparing marital status with the number of defaults across loan applicants. The plot reveals that the majority of applicants across all marital categories did not default, with only a single default observed among married individuals. No default cases were recorded for single, divorced, or unspecified statuses. This visualization suggests that marital status alone does not show a strong correlation with loan default behavior and is not a dominant factor in predicting default risk. However, it may contribute to prediction accuracy when evaluated alongside other variables.

3.9 Interface Designs

Malvik (2020) defines interface design as the perspective that users will be given and interact with through inputs to the computer, which can include buttons, typefaces, icons, and text inputs. Interface design makes sure that the interface provides a better experience without making system use more challenging overall by looking at the screens that allow users to accomplish tasks they may need to complete within the system. In order to create software that is easy to use and doesn't confuse users, user interface design is therefore essential. The proposed loan eligibility prediction system aims to provide an easy-to-use interface for all parties involved, enabling them to accomplish their tasks efficiently and effectively. It's important to keep in mind the design.

3.9.1 Menu Design

A group of options that assist users in navigating to various areas of a system to obtain the information they require is referred to as menu design. According to Tozzi (2021), the menu serves as a navigation hub linking all of the accessible pages or views and is typically found in the top bar of the program interface. Links protected by authentication procedures may also be displayed, along with content pertinent to the current user interface. For example, the system will direct the user to the login page if they attempt to access a protected page, like the application dashboard, without first logging in.

3.9.2 Main menu

This menu shows the option links to all the system sections. At the web application landing page, it offers a standardized list of navigation options and might also have more extensive drop-down options.

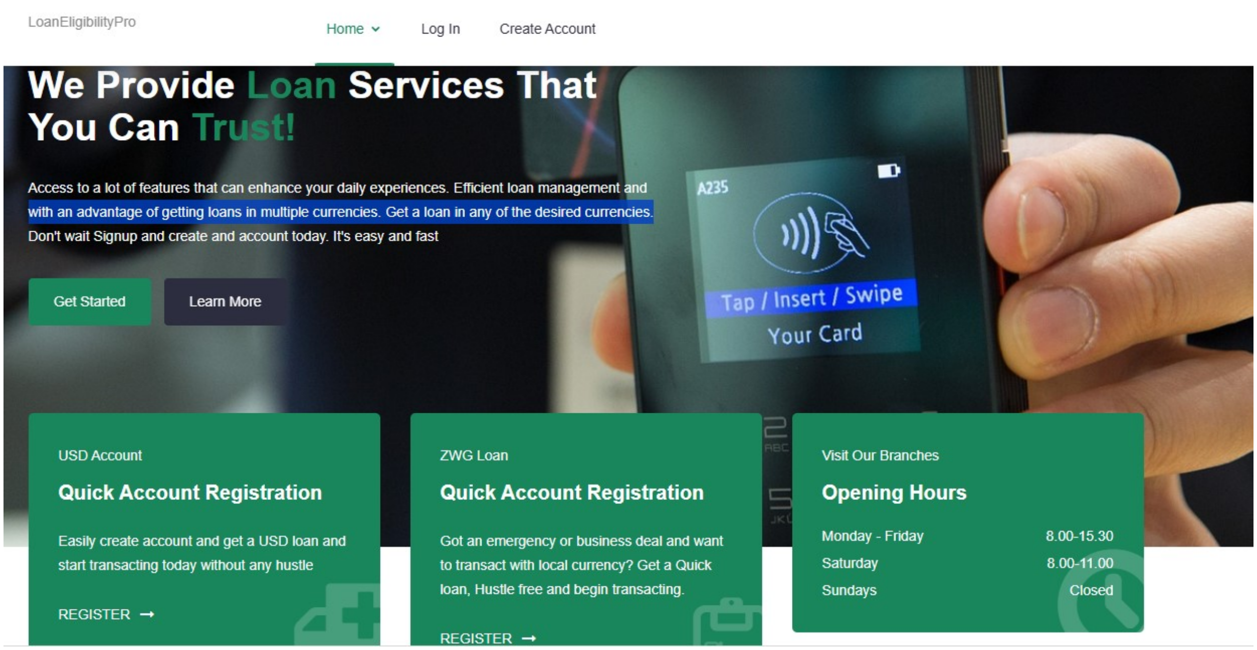


Figure 3.14: Main menu

3.10 Input Design

The process entails transforming user input into a computer-understandable format. It collects error-free user input and send it to the appropriate computer functions with the logic to change the received data. When the data gathered matches the data types specified by the storage, validation checks should be performed. It is also crucial to make sure that all of the proposed system's form inputs are simple, within the parameters of their design, and enable users to perform only the tasks they need to. Additionally, they should provide a brief description of the input labor and be appropriately colored.

The system's input screens are as follows:

3.10.1 Signup and Sign in Forms

The purpose of these forms is authentication. Signing up is the first step for both clients and businesses, and the credentials are stored in the system. Once the credentials have been saved, they can both log in to the system to access features like client registration.

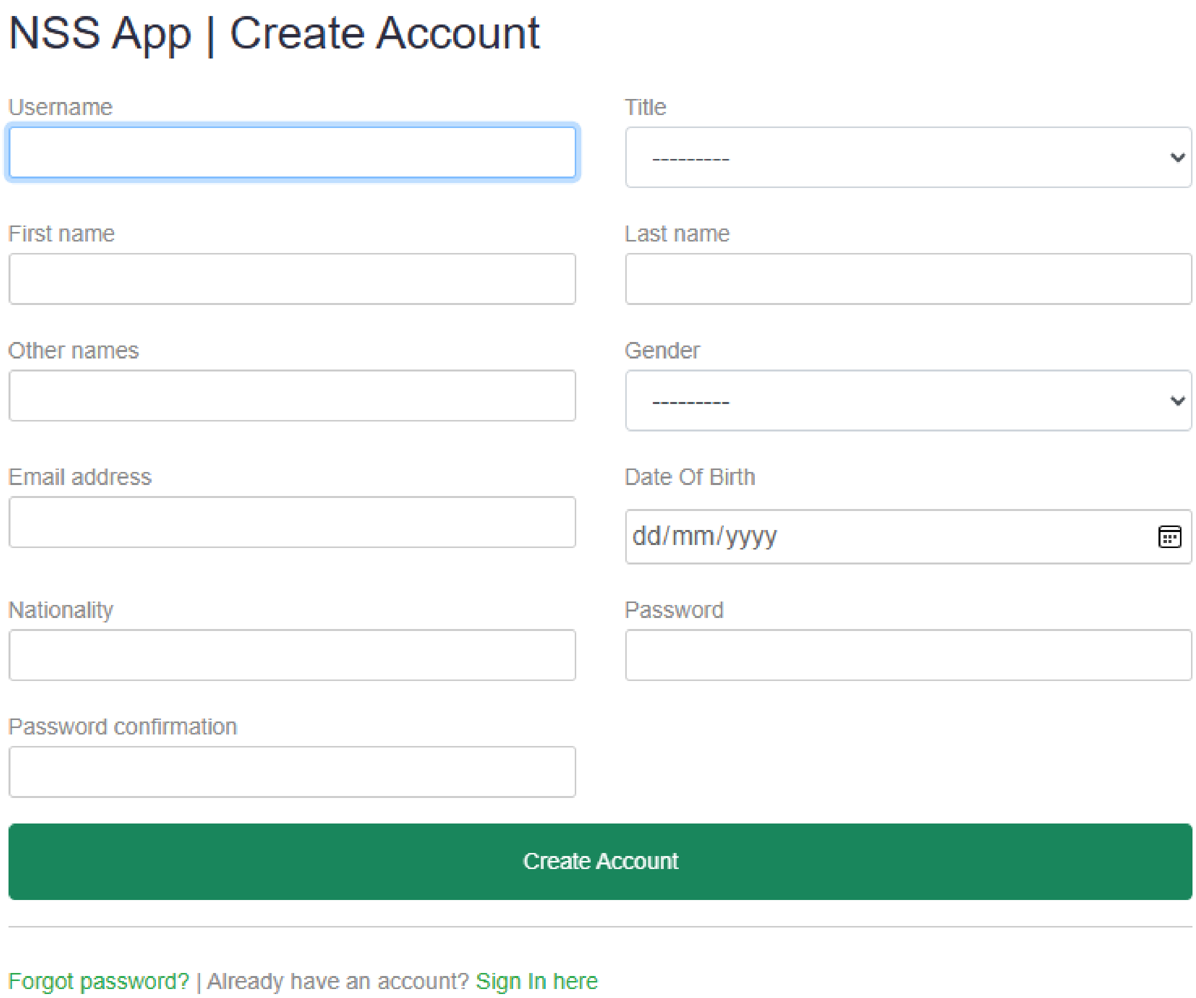


Figure 3.15: signup form

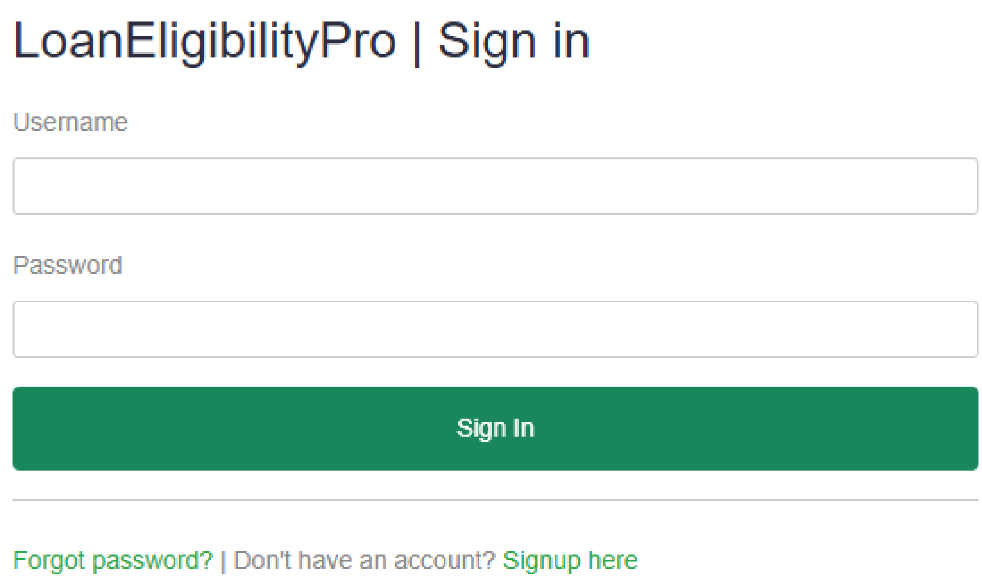


Figure 3.16: sign in form

3.11 Summary

This chapter introduced a summary of the methodological strategy adopted in developing a loan eligibility forecasting system using machine learning techniques. The use of interviews, observations, and surveys made it possible for the researcher to gather relevant system requirements from real situations, while technical implementation of the model was made possible using tools like Python, Jupyter Notebook, Django, and SQLite. Moreover, the chapter covered functional and non-functional requirements, system design through UML diagrams, as well as user interface issues to enable usability by users of all levels. The inclusion of the Agile development model enabled iterative refinement of the system through user input, whereby the end solution is correct and user-friendly. The methodology in general presented a good base for creating a smart, data-based loan eligibility forecasting system to enable decision-making in financial institutions.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.1 Introduction

In creating a Loan Eligibility Prediction System, the act of representing data, analysing data, and making interpretations of data is an important function that impacts the usability and accuracy of the prediction results. This section, elaborated below, is dedicated to describing the process and outputs from the exploratory data analysis (EDA) stage of the inquiry up until the model building process. The dataset was subject to an exploratory data analysis (EDA) exercise with five machine learning models we chose specifically to leverage the strengths of to identify patterns, isolate anomalies, and quantify feature significance - the model fit assisted in discerning which features are influential in making loan approvals and how much influence, so at this early stage we were setting the stage for the predictive modelling stage. After we had moved past EDA, we drew upon logistic regression and linear regression as our models to functionally predict loan eligibility. Logistic regression was relevant because it is a binary classifier that allowed us to predict loan eligibility at a fundamental level with an "approved" or "not approved" classification and linear regression was useful to examine the relationship between its quantitative variables of applicant income, loan amount, and credit history.

4.2 Testing

Testing is a sort of analysis carried out against a system to make sure it performs precisely as intended and that all errors are verified prior to deployment, according to Yasar (2022). Since the main goal of testing is to provide usable systems, we can verify and validate the system by conducting system tests, which allows the software engineer to address any issues. Through testing, we can find system flaws, enhance quality, win over customers, and fix any software vulnerabilities that could pose a risk or allow cyberattacks.

The suggested system can be tested at the following levels:

4.2.1 Unit testing

This kind of testing is carried out against every single software component with the goal of making sure that each unit of code or component functions as intended (Yasar, 2022). The developer must test each component separately in order to create test cases that can verify the dependability of the tested code units. By testing separate software components, we can lower the likelihood of errors occurring during the merging process

```
.
======================================================================
FAIL: test_intentional_fail (loan_application.tests.LoanApplicationTestCase.test_intentional_fail)
----------------------------------------------------------------------
Traceback (most recent call last):
  File "C:\Users\840 g3\PycharmProjects\loan_eligibility_pro\loan_application\tests.py", line 15, in test_intentional_fail
    self.assertEqual(1, 2)  # This will FAIL on purpose
    ~~~~~~~~~~~~~~~~^^^^^^
AssertionError: 1 != 2

----------------------------------------------------------------------
Ran 5 tests in 3.199s

FAILED (failures=1)
Destroying test database for alias 'default'...
PS C:\Users\840 g3\PycharmProjects\loan_eligibility_pro>
```

Figure 4.1: failed unit testing

```
.
----------------------------------------------------------------------
Ran 4 tests in 2.110s

OK
Destroying test database for alias 'default'...
PS C:\Users\840 g3\PycharmProjects\loan_eligibility_pro>
```

Figure 4.2: successful unit testing

4.2.2 Integration Testing

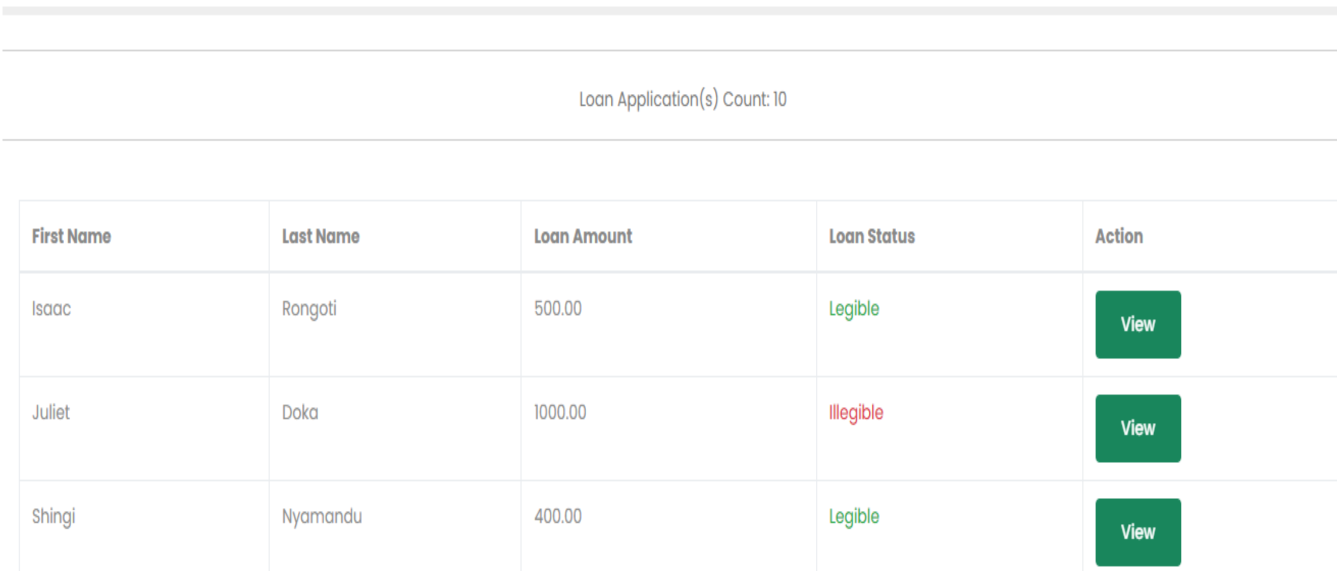
System components are combined to ensure that they are working as intended (Martin,2017). We can identify all data flow flaws during integration testing when module components are interfaced, allowing us to promptly address problems in their early stages.

```
======================================================================
FAIL: test_view_requires_login (loan_application.tests.TestViewRendering.test_view_requires_login)
----------------------------------------------------------------------
Traceback (most recent call last):
  File "C:\Users\840 g3\PycharmProjects\loan_eligibility_pro\loan_application\tests.py", line 117, in test_view_requires_login
    self.assertEqual(response.status_code, 302)  # Should redirect to login
    ~~~~~~~~~~~~~~~~^^^^^^^^^^^^^^^^^^^^^^^^^^^
AssertionError: 200 != 302

----------------------------------------------------------------------
Ran 8 tests in 20.257s

FAILED (failures=1)
Destroying test database for alias 'default'...
PS C:\Users\840 g3\PycharmProjects\loan_eligibility_pro>
```

Figure 4.3: Integration testing

4.2.3 Black box testing

A popular technique in software testing is called "black-box" testing, in which the tester is not aware of the internal organization, architecture, or operation of the system under test. The goal is to test the solution's response (the resulting output) through evaluating the input and input behavior to see if it meets specifications or expectations (Jazayeri & Alfayez, 2022). Black-box testing is useful in developing systems like Loan Eligibility Prediction, in which models like Linear Regression and Logistic Regression are employed. The tester provides combinations of inputs, such as: income (amount), credit history (good, fair, poor), and employment status (employed, unemployed, retired), to ascertain cases where the test entity assigns users to be "Eligible" or "Not Eligible" based on rules or predictions that are predefined (Mahmood et al., 2021). Black-box testing is a practical method to test for functional errors, missing requirements, and user interface discrepancies, and should be considered during system-level and acceptance testing. Because the tester doesn't need to understand the code behind the systems, their testing promotes user perspectives for verification (Patel & Sharma, 2020). For instance, an eligible loan application should yield an outcome of "Eligible" if it has a high-income category and good credit history; any outcomes that differ from the expectation indicate some level of error in the systems' model or logic.



Loan Application(s) Count: 10

First Name	Last Name	Loan Amount	Loan Status	Action
Isaac	Rongoti	500.00	Legible	View
Juliet	Doka	1000.00	Illegible	View
Shingi	Nyamandu	400.00	Legible	View

Figure 4.4: showing list of loan applications approved and declined

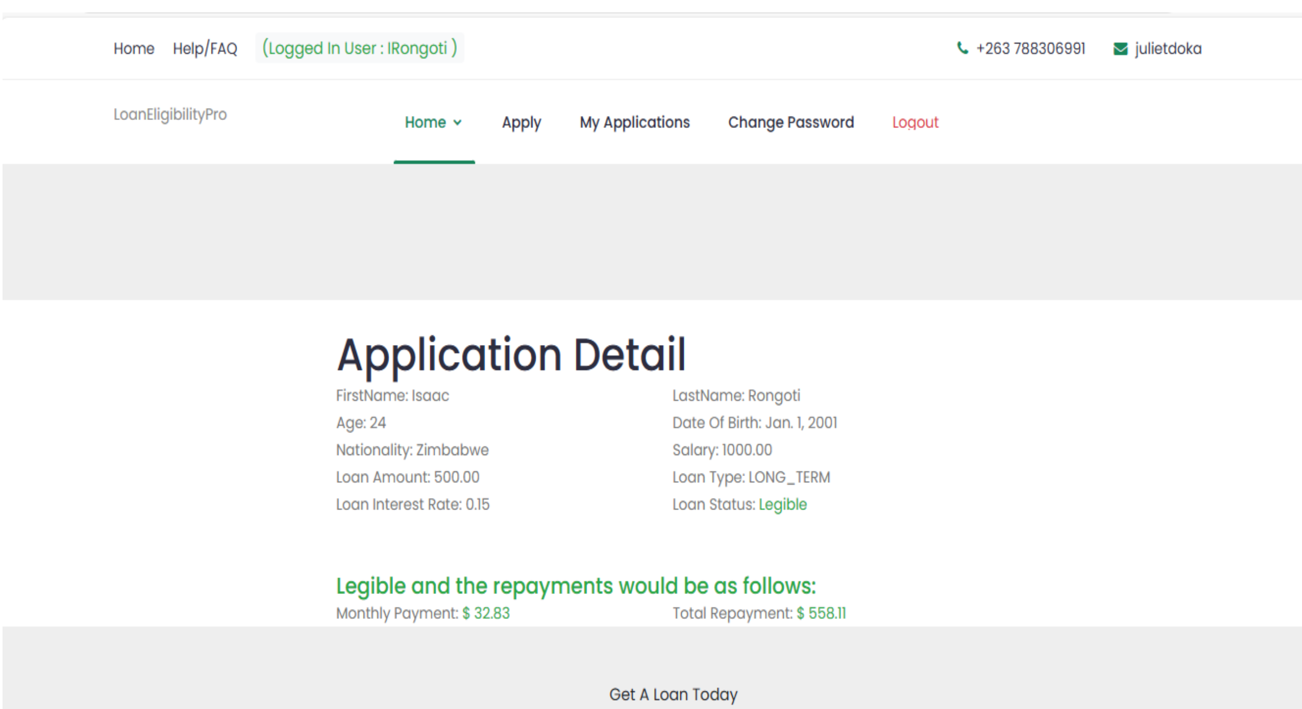


Figure 4.5: legible loan application

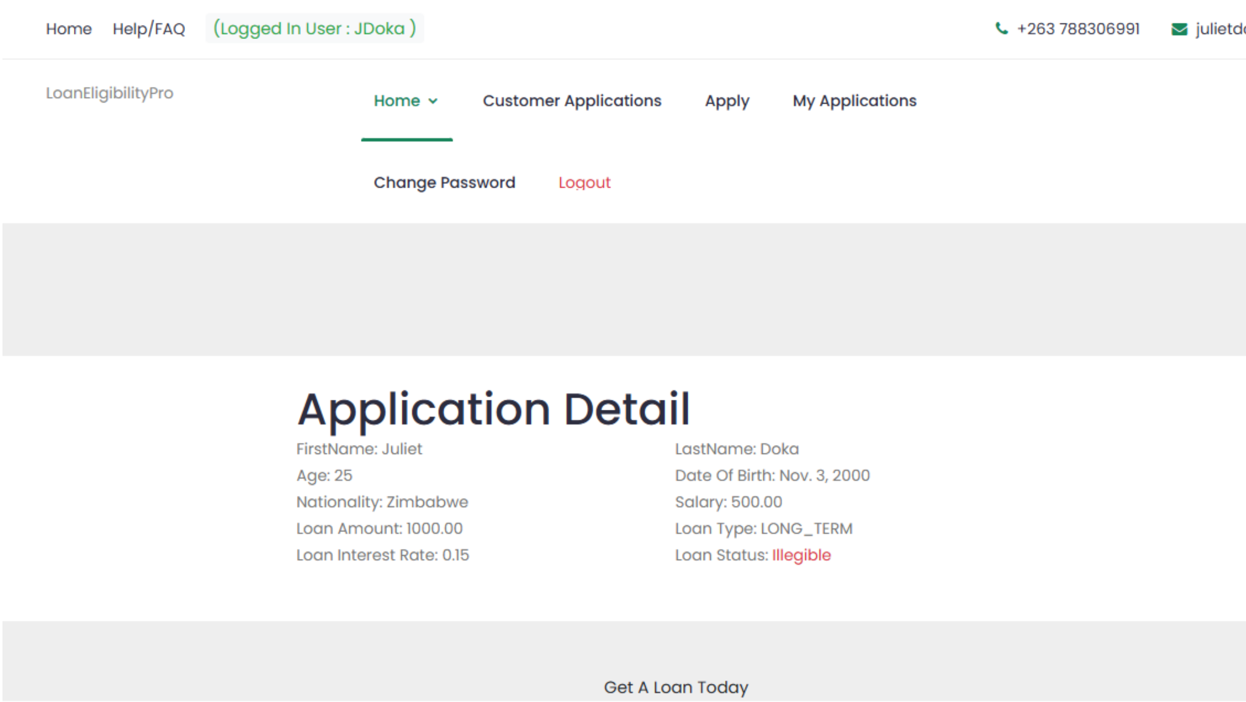


Figure 4.6: declined loan application

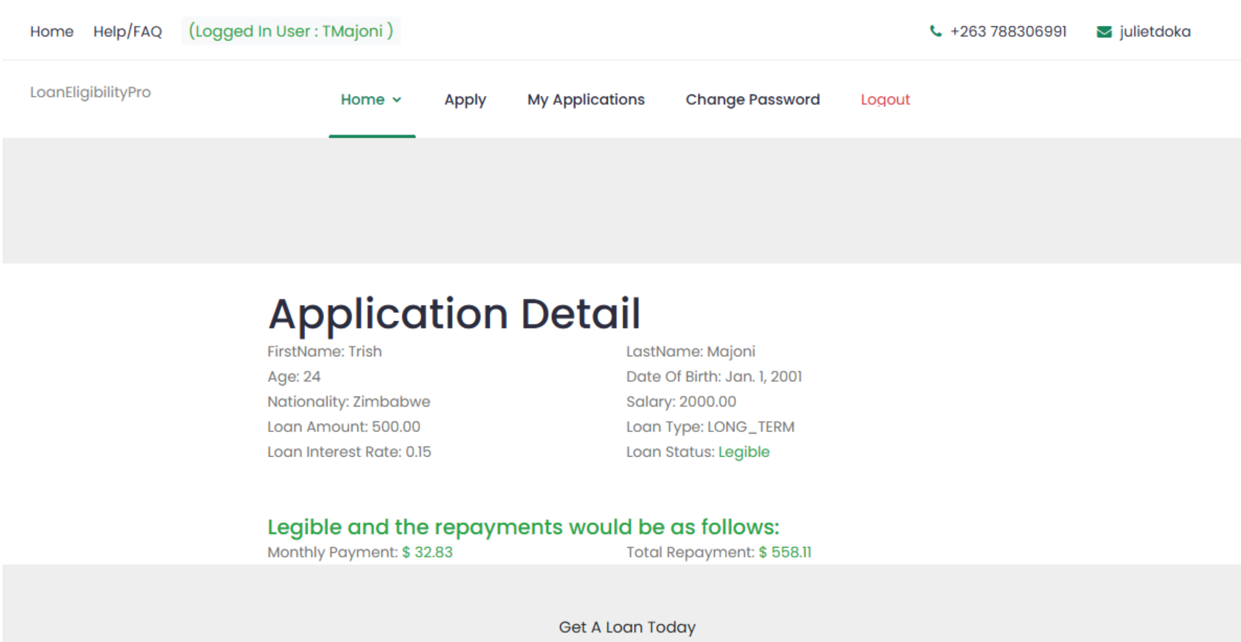


Figure:7 Legible loan application after changing loan default number and outstanding balance

The result below reflects that the user already had an existing debt of $1500 that is 75% of the salary, number of defaults is also high and she adds $500 which pushed the total debt to $2000, that is 100% of her salary. Again, the user has high debt-to-income-ratio left no room for cash flow flexibility.

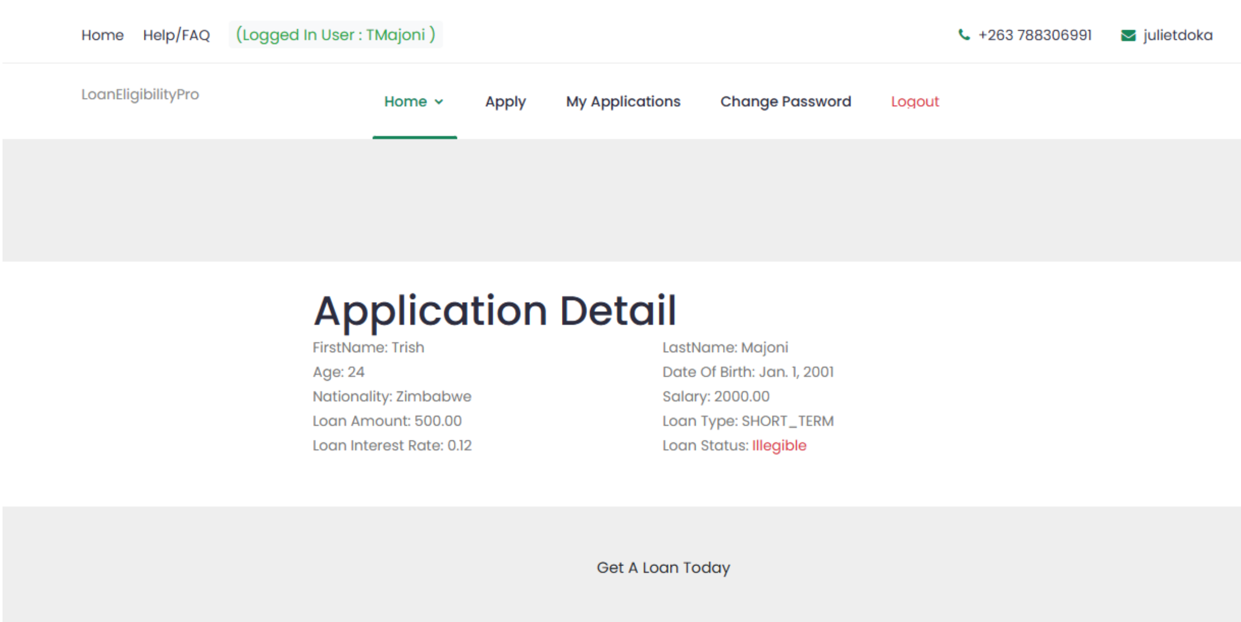


Figure 8: *Illegible loan application after changing loan default number and outstanding balance*

4.2.4 Whitebox testing

White-box testing, also known as structural or transparent-box testing, differs from black-box testing in that it looks at a software application's core logic and components rather to only its external behaviour (Wikipedia, 2024). To make sure that the logic functions as intended, it entails testing the actual code, including branches, loops, and paths (Sharma & Kumar, 2021). In systems like a Loan Eligibility Prediction System, which might use Linear and Logistic Regression models, white-box testing helps verify that the model integration, data preprocessing, and decision rules within the code function correctly. For example, developers might check whether all conditions in a series of `if/else` statements are covered, or whether a loan eligibility threshold is correctly applied (e.g., applicants with a score below 600 should not qualify). White-box testing is typically done by developers because it requires knowledge of the codebase. It often includes techniques such as code coverage analysis, path testing, and unit-level validation to ensure that every possible scenario in the logic is tested (Ali et al., 2022).

```python
class TestViewRendering(TestCase):
    def setUp(self):
        self.client = Client()

    def test_application_view_rendering(self):
        response = self.client.get(reverse('loan_applications:apply'))
        self.assertEqual(response.status_code, second: 200)
        self.assertTemplateUsed(response, template_name: 'loan_application/application.html')
        print(response.content)

    def test_application_list_view_template_rendering(self):
        response = self.client.get(reverse('loan_applications:application_list'))
        self.assertEqual(response.status_code, second: 200)
        self.assertTemplateUsed(response, template_name: 'loan_application/application_list.html')
        print(response.content)

        def test_submit_loan_application_form(self):
            self.client.login(username='tester', password='testuserpassword')
            response = self.client.post(reverse('loan_applications:apply'), {
                'loan_amount': 500,
                'loan_rate': self.loan_rate.id,
                'number_of_defaults': 0,
                'salary': 5000,
                'marital_status': 1,
                'outstanding_balance': 0,
                'monthly_payment': 50
```

Figure 4.15.0: code snippet on user creation and template rendering

```python
class LoanApplicationListView(generic.ListView):
    template_name = 'loan_application/application_list.html'
    context_object_name = 'applications'

    def get_queryset(self):
        applications = LoanApplication.objects.all()
        return applications


class LoanApplicationDetailView(generic.DetailView):
    template_name = 'loan_application/application_detail.html'
    context_object_name = 'application'

    def get_queryset(self):
        return LoanApplication.objects.all()


def calculate_repayment(application:LoanApplication):
    P = application.loan_amount
    annual_rate = application.loan_rate.interest_rate
    r = (annual_rate / 12 )
    n = application.loan_rate.loan_term_in_months

    M = (P * r *(1+r)**n ) / ((1+r)**n -1)
    return M, M*n
```

Figure 4.15.1: code snippet for EMI calculation

4.3 Evaluation measure and results

To assess the performance the logistic regression model used in this project, several standard evaluation metrics were applied. These metrics help determine how well the model can distinguish between eligible and ineligible loan applicants. The following confusion matrix summarizes the result:

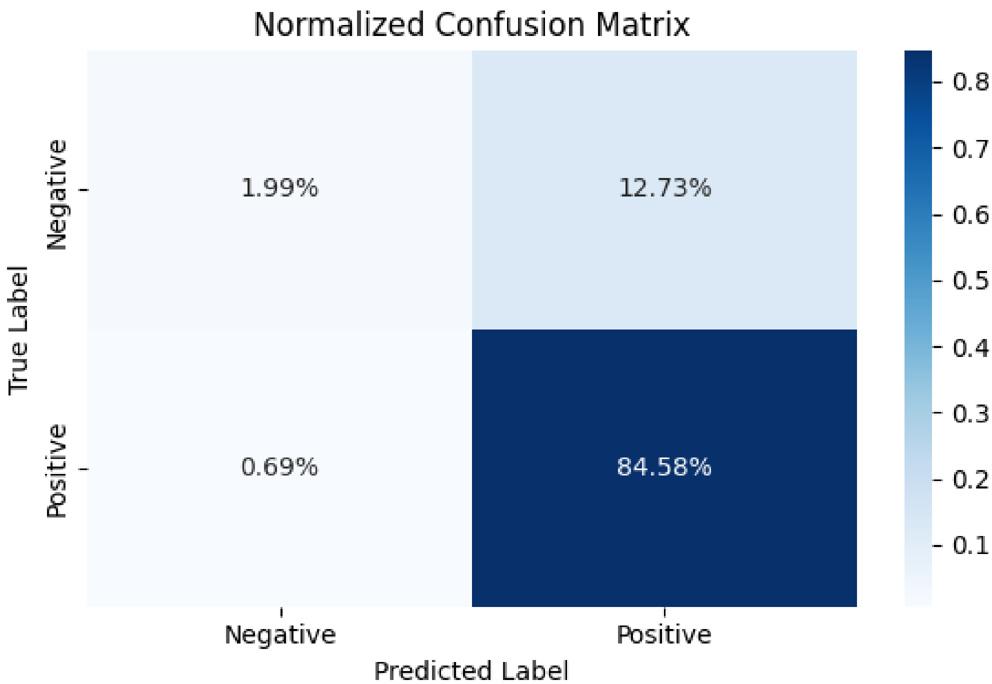


Figure 4.16: Confusion matrix before hyper tuning

```
Classification Report:
              precision    recall  f1-score   support

           0       0.74      0.14      0.23      2945
           1       0.87      0.99      0.93     17055

    accuracy                           0.87     20000
   macro avg       0.81      0.56      0.58     20000
weighted avg       0.85      0.87      0.82     20000

-----------------------------------------------------------
```

Figure4.17: Classification Report before hyper tuning

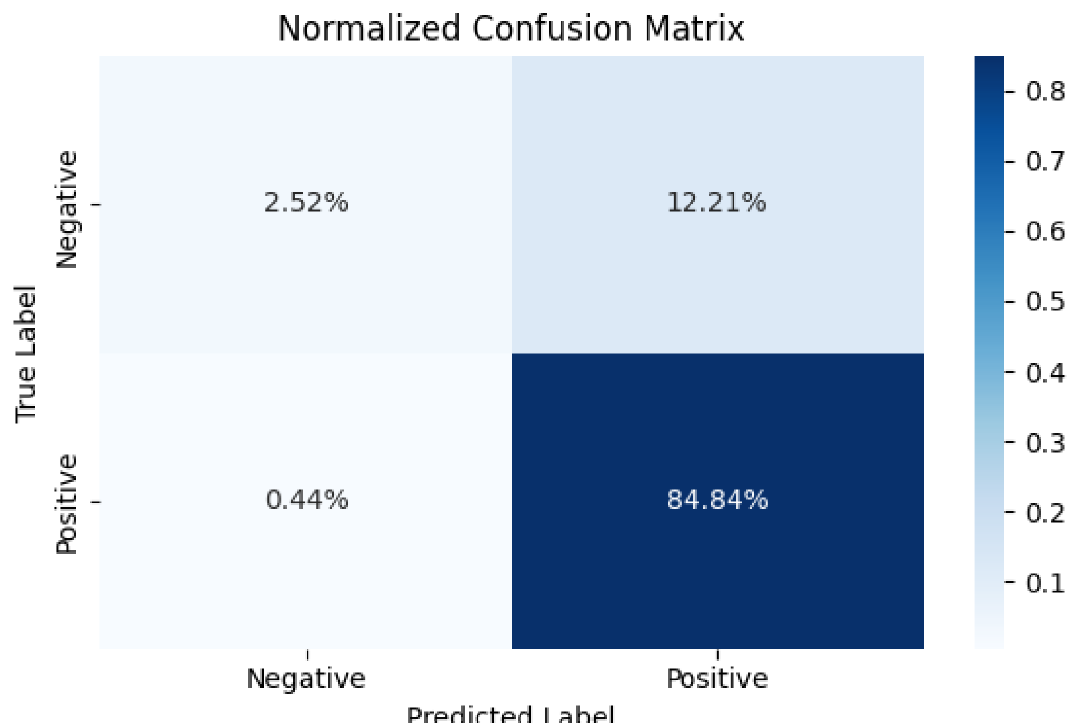


Figure4.18 Confusion matrix after hyper tuning

```
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.17      0.28      2945
           1       0.87      0.99      0.93     17055

    accuracy                           0.87     20000
   macro avg       0.86      0.58      0.61     20000
weighted avg       0.87      0.87      0.84     20000

-----------------------------------------------------------

Model Improvement : 1.30%
```

Figure4.19 Classification Report after hyper tuning

Metric	Class	Before Tuning	After Tuning
Precision	0	0.74	0.85
Precision	1	0.87	0.87
Recall	0	0.14	0.17
Recall	1	0.99	0.99
F1-score	0	0.23	0.28
F1-score	1	0.93	0.93
Accuracy	-	0.87	0.87
Macro avg F1	-	0.58	0.61
Weighted avg F1	-	0.82	0.84

Figure: Table showing before and after hyper tuning

Before Hyperparameter Tuning:

On class 1 (the majority class), the model exhibits excellent performance, with high F1-score (0.93), precision (0.87), and recall (0.99). Class 1 (the majority class) is where the model excels, with excellent precision (0.87), recall (0.99), and F1-score (0.93). A lack of class balance is shown by low macro averages (macro F1: 0.58).

After Hyperparameter Tuning:

Class 0 precision substantially increases (from 0.74 to 0.85). Although it still stays low, class 0 recall somewhat improves (from 0.14 to 0.17). F1-score for class 0 increases (from 0.23 to 0.28), showing better, but still limited, performance on this class. Class 1 metrics don't change, suggesting that majority class performance hasn't decreased. Macro and weighted averages improve (macro F1: 0.61, weighted F1: 0.84), indicating a more balanced model. Overall accuracy remains the same (0.87), as reflected in the 1.3% model improvement noted.

4.4 Summary of research findings

After completing the hyperparameter tuning of the Logistic Regression model, the system attained an overall accuracy of 87% with strong predictive performance for loans approved (class 1) based on its precision of 0.87, recall of 0.99, and F1-score of 0.93. The model did not perform nearly as well predicting loans rejected (class 0), with recall (0.17) and F1-score (0.28) showing high rates of False Negatives. The macro average F1-score (0.61) and weighted-average (0.84) overall F1-scores suggest the model was proficient in handling the class imbalance that exists in its prediction performance. Nevertheless, the model accuracy showed a noticeable 1.30% improvement after the hyperparameter tuning phase, indicating a meaningful improvement in the model's predictive performance. The results imply that while the LR model is competent in predicting loans approved, pathways such as resampling and/or cost-sensitive learning, could be taken to improve loans rejected detection.

4.5 Conclusion

This chapter presented, analyzed, and interpreted the collected loan application data in order to discover meaning and value, as well as to understand the predictive model's effectiveness. Specifically, we took the dataset and preprocessed it by removing null values, encoding categorical variables, selecting features and ensuring it was a solid set of data quality for modelling phase. The model was trained by using logistic regression, and was evaluated using an 80:20 train-test split, in this case, we redeployed a new predictive model to understand eligibility based on the loan data collected. We evaluated the performance of the model on the predicted outcomes and on the test set, which the model predicted loan eligibility reality compared to the results of applicants estimated eligibility, the findings confirmed the model performed acceptably. The features that were indicated as relevant in this dataset were income level, loan amount, and employment status, which indicated to be of significant importance relative importance in eligibility, decision processing, to outcomes passed or not passed. The overall findings confirmed the use of the logistic model was an acceptable fit for binary classification outcomes, eligibility pertaining to the data collected in this thesis. This illustrated an overall sense as we move into chapter 6 to make key conclusions, recommended action, and proposed improvements to either the model or the data collected in the future.

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

Towards the end of the chapter, the writer wants to reflect on all that we have achieved, from initial aspirations to developing and evaluating a loan eligibility prediction system. The discussion will focus on some of the key goals set out at the beginning, and reflect on how they were achieved. The author will discuss the main findings of the project, what those findings mean beyond the research, and if/when they can be useful for financial institutions that seek to make fast and fair loan decisions. With the lessons learned in the course of the project, the researcher will offer some recommendations, both to improve the existing system, or forward for people who may wish to further build on this research when they have the opportunity. The discussion would be an appropriate end for the research journey, reflecting on what has been achieved, and what will come next.

5.2 Aims and objectives realization

The overall objective of the study was to create and validate a predictive model using machine learning algorithms to help identify if a loan applicant is eligible or not. With an analysis of applicant data, history of finance and creditworthiness, the system aimed to make approving a loan more efficient and effective. While the model went well in achieving the goal, the target was accomplished only partially. The researcher was also able to accomplish the pre-defined goals. The Django application was designed, in which the users were able to input loan application details and get real-time eligibility feedback. The data processing, model prediction and the usability of the interface were validated.

5.3 Conclusion

For binary classification problems like loan approval decisions, logistic regression is a powerful and comprehensible machine learning algorithm, according to the development of the loan eligibility prediction system. The model performed effectively after thorough data preprocessing that was critical to its performance. Integration of the system into a simple-to-use Django web application enabled real time, automated prediction and rendered it a valuable tool for financial institutions. Overall, the project proved that machine learning could accurately aid and automate the loan approval process.

5.4 Recommendations

- Make use of more models for example random forest and Gradient Boost to enhance performance.
- Add visual explanations like SHAP values to help users understand the reasoning behind predictions.
- Implement strong encryption and data privacy mechanisms, especially since the system deals with sensitive financial data.
- Collaborate with financial institutions to test the system on real application data for further validation.

5.5 Future work

Although this project succeeded in developing a machine learning driven loan eligibility prediction system using logistic regression, there are many areas to contribute and improve upon in the system for later versions. One of these areas is incorporating an external credit check module that allows the system to check automatically whether or not a user has more than three active or pending loans from other financial institutions. This would be done by connecting the system to national or regional databases or credit bureaus to gather up-to-date credit information about applicants. In this way, the system would not rely on user-provided data but also validate against the applicant's overall debt exposure providing a better risk assessment. Secondly, adding a loan repayment module where authenticated users can pay back directly within the system using several payment options. Expanding the system to include loan repayment history tracking, risk score over a period of time and user feedback build-up would help support broader financial services. It would add convenience to users and ease loan recovery processes to financial institutions.

5.6 Challenges faced

One of the primary obstacles faced by the researcher was obtaining relevant data. Gaining access to real-world loan application records from financial institutions proved to be quite difficult, largely because of stringent confidentiality and privacy regulations. Consequently, the researcher was limited to using publicly available or anonymized datasets. Unfortunately, these do not always include the full complexity and sophistication involved in actual loan applications.

Time constraints also posed a significant challenge. The researcher had to contend with less-than-adequate availability of live APIs such as those provided by payment gateways that restricted the scope for incorporating advanced features, such as online repayment of loans, for example. Despite these limitations, though, the project was able to achieve its core objectives and established a comprehensive framework for potential future enhancements.

REFERENCES

Boukherouaa, E. B., Shabsigh, G., AlAjmi, K., Deodoro, J., Farias, A., Iskender, E. S., Mirestean, A. T., & Ravikumar, R. (2021). *Powering the digital economy: Opportunities and risks of artificial intelligence in finance*. International Monetary Fund.

IntroBooks. (2020). *Artificial intelligence in banking*.

ZIMSTAT. (2021). *Financial sector report: Trends and transformations*. Zimbabwe National Statistics Agency.

Ndayisenga, T. (2021). *Bank loan approval prediction using machine learning techniques* (Doctoral dissertation).

Orji, U. E., Ugwuishiwu, C. H., Nguemaleu, J. C., & Ugwuanyi, P. N. (2022, April). Machine learning models for predicting bank loan eligibility. In *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)* (pp. 1–5). IEEE.

Gonen, B., & Sawant, D. (2020). Significance of agile software development and SQA powered by automation. In *2020 3rd International Conference on Information and Computer Technologies (ICICT)* (pp. 7–11).

Adams, R. (2022). Exploring factors influencing loan approval using regression analysis. *European Journal of Operational Research, 290*(1), 231–250.

Ali, N., Yasmin, T., & Iqbal, S. (2022). White-box testing techniques for AI-enabled software systems: A review and analysis. *International Journal of Advanced Trends in Computer Science and Engineering, 11*(4), 327–336. https://doi.org/10.30534/ijatcse/2022/011142022

Aker, J. C., et al. (2019). Machine learning in credit scoring. *Journal of Financial Innovation, 5*(1), 15–30.

Anyanwu, J. C. (2021). The applicability of linear regression in predicting loan eligibility. *Zimbabwean Journal of Banking, 4*(2), 56–68.

Bai, Z., et al. (2018). Data mining and predictive modeling techniques in credit risk. *Journal of Risk Finance, 19*(3), 211–225.

Belhadi, A., Kamble, S.S., Mani, V., Benkhati, I. and Touriki, F.E. (2021) 'An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance', *Annals of Operations Research* [Online], 1–29. Available at: https://doi.org/10.1007/s10479-021-04210-0

Chen, J., Xie, Z., & Wang, Y. (2019). Predicting credit risk with machine learning: A comparative study. *Financial Innovation*.

Chen, R., & Zhang, W. (2021). Machine learning for credit risk: A comparative study. *Finance Research Letters, 29*(1), 101–113.

Chiri, K., et al. (2021). Credit risk in Zimbabwe: A review of current practices. *Zimbabwe Financial Review, 3*(3), 22–40.

Fernandez, S., & Lee, M. (2020). Code-based testing for intelligent systems. *IEEE Software, 37*(5), 65–71. https://doi.org/10.1109/MS.2020.2993130

Hernandez, M., & Santos, T. (2021). Machine learning vs traditional methods in credit scoring. *International Journal of Credit Risk Management, 13*(2), 121–139.

Jazayeri, A., & Alfayez, R. (2022). Modern techniques in software testing: Trends, challenges, and future directions. *Journal of Software Engineering and Applications, 15*(3), 123–139. https://doi.org/10.4236/jsea.2022.153009

Kim, Y., & Lee, S. (2019). Predicting loan default: A logistic regression approach. *International Journal of Financial Studies, 7*(4), 102–118.

Kumari, S., Swapnesh, D., Nayak, D., & Swarnkar, T. (2023). Loan eligibility prediction using machine learning: A comparative approach. *International Journal of Computer Applications, 3*, 48–54.

Liu, X., et al. (2018). Evaluation of credit risk models using statistical techniques. *Journal of Banking & Finance, 87*, 243–255.

Mahmood, S., Ahmad, M., & Khurshid, K. (2021). A review of black box testing techniques for machine learning-based software systems. *International Journal of Advanced Computer Science and Applications, 12*(5), 456–464. https://doi.org/10.14569/IJACSA.2021.0120560

Martinez, F. (2019). Advancements in credit risk modelling: Challenges and solutions. *Journal of Credit Risk Management, 15*(4), 56–78.

Maliyamkono, T., & Zimba, F. (2020). Logistic regression in financial decision making in Zimbabwe. *African Development Review, 32*(4), 581–596.

Menzies, T. (2014). An analysis of linear models in financial credit evaluation.

Mudavanhu, T. (2022). Alternative data and its impact on credit scoring in Zimbabwe. *Journal of Financial Inclusion*.

Ndlovu, F. (2021). Evaluating logistic regression for credit risk in Zimbabwean banks. *Zimbabwean Economic Journal, 13*(3), 44–62.

Nyoni, T. (2020). The importance of credit risk management. *Zimbabwe Banking Review, 8*(1), 10–15.

Patel, A., & Gupta, R. (2023). Understanding credit risk in the age of machine learning: An empirical study. *Journal of Finance and Accountancy, 22*(3), 45–68.

Patel, N., & Sharma, V. (2020). Software testing methodologies and tools for quality assurance. *International Journal of Scientific Research in Computer Science, 8*(2), 78–85. https://doi.org/10.32628/CSEIT206236

Ravi, S., & Ramesh, C. (2021). Addressing human bias in loan approvals through machine learning. *International Journal of Banking, Accounting, and Finance*.

Sharma, A., & Kumar, R. (2021). Structural testing approaches for high-quality software development. *Journal of Software Engineering and Intelligent Systems, 6*(1), 45–54. https://doi.org/10.5281/zenodo.4681239

Smith, J., & Johnson, L. (2020). The application of machine learning in credit risk assessment: A review. *Journal of Finance and Risk Perspectives, 9*(2), 35–50.

Baesens, B., Van Gestel, T., Stepanova, M., Suykens, J., & Vanthienen, J. (2016). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 67*(6), 900–918. https://doi.org/10.1057/jors.2015.65

Chikodzi, D. (2020). Credit risk assessment challenges in Zimbabwe's financial sector. *African Journal of Finance and Management, 12*(2), 87–98.

Khandani, A. E., Kim, A. J., & Lo, A. W. (2020). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*(11), 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 66*(6), 902–920. https://doi.org/10.1057/jors.2014.22

Mawere, M., & Dube, N. (2021). Manual loan approval systems and credit default risk: Evidence from microfinance institutions in Zimbabwe. *International Journal of Finance and Banking Research, 7*(1), 12–21.

Sengodan, T., Misra, S. and M, M. (eds.) (2026) *Advances in Electrical and Computer Technologies: Proceedings of the 6th International Conference on Advances in Electrical and*

Computer Technologies (ICAECT 2024), Tiruchengode, India, September 26th–27th, 2024, 1st edn. Boca Raton: CRC Press. Available at: https://doi.org/10.1201/9781003515470

Obi, I., Adeyemi, S., & Ncube, N. (2022). Adopting machine learning for credit risk scoring in African financial institutions: A contextual perspective. *African Journal of Information Systems, 14*(3), 45–59.

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2017). *Credit scoring and its applications* (2nd ed.). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718589

Wang, F., & Zhang, J. (2023). Fairness in machine learning-based credit scoring: Mitigation strategies and regulatory perspectives. *Journal of Financial Regulation and Compliance, 31*(1), 32–49. https://doi.org/10.1108/JFRC-11-2022-0112

Torres, R., & Dias, C. (2020). Application of logistic regression in predicting borrower default. *Applied Economics Letters, 27*(12), 926–930.

Deakin, E. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research, 10*(1), 167–179.

Kendall, K.E. and Kendall, J.E., 2010. *Systems Analysis and Design*. 8th ed. Boston: Pearson Education.

Martin, M., 2021. *Phases of Prototyping Model*. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/prototyping-model-in-software-engineering/ [Accessed 26 Jun. 2025].

Pressman, R.S., 2014. *Software Engineering: A Practitioner's Approach*. 8th ed. New York: McGraw-Hill Education.

Sommerville, I., 2011. *Software Engineering*. 9th ed. Boston: Addison-Wesley.