Unlocking Financial Futures: Harnessing Machine Learning and Deep Learning to Predict Loan Status

Roshni Sanjay Solanki - 10608946

Supervisor: Prof. Anesu Nyabadza

Dublin Business School

This dissertation is submitted for the degree of

MSc in Data Analytics

August 2024

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Roshni Solanki
August 2024

Acknowledgement

I would like to sincerely thank my supervisor, Prof. Anesu Nyabadza, for his constant guidance and support during my thesis. His help was essential to my research. I also want to thank my friends and family for always supporting me throughout my academic life.

Table of Contents

List of Figures

List of Tables

Abstract

In this study, we are presenting research on the identifiability of loan status with more accuracy and easy to implement by using state-of-art machine learning and deep learning models which will provide banks an upper hand in managing risk appropriately. We have tested the performance using one public domain dataset having 19 attributes and up to 100k records, results were more or less consistent in four complex models: Gradient Boosting, XGBoost (meta classier), Convolutional Neural Networks and Long Short-Term Memory. They were selected with the idea to use models that can deal with nonlinear relationships and large amount of data.

We have been working with 2 ML and 2 DL algorithms, assessing them on the merits of Accuracy (the proportion of correctly predicted loan statuses), Precision (the number of rowsclassified as Approved divided by all Returned in each Classifier were actually truly tagged Correctly), Recall (True Positive Rate: The Ratio is labeled True and are Actuality Real Trues) and F1 Score (whose Disequilibrium-weighted Harmonic Mean among accuracy = 2* (precision * recall)/precision + recall we can used for these four Measurement.) to determinewhich model was best when it comes down also looking at features-importance. Machine learning models and especially deep learning tools outperform traditional approach for both the lending platform to process loan applications as well as financial institutions to manage risks, according on results obtained. The results of this research underscore the critical rolethat AI and deep learning can play in improving decision-making among financial institutions, as well as the potential benefits for minimizing counter-party risk.

Keywords: Credit rating evaluation, Machine learning, Deep learning, Financial innovation, Risk control

Chapter 1: Introduction

1.1 Introduction

In the current financial world, the lenders must have a solid idea on what basis one can predict whether a loan application is in control? Loan approvals are crucial counterparties for the banking and financial organizations which have a deep impact on their profitability as well as the macro-economic activity. And while little innovation may have been expected when it comes to traditional financial institutions and their decision-making processes, the last few years can be described as nothing but focused on implementing advanced predictive technologies. In this research, we deploy some of the machine learning as well as deep learning algorithms to perform loan status prediction in hopes that it will lead us to better and more frequent measurements for availability, thereby enabling a superior risk management overall.

One of the important machine learning areas in credit risk management is to predict whether or not a loan will be paid back based on various information about this borrower. For decades, traditional credit scoring models based on statistical methods such as logistic regression and linear regression have been the foundation of assessing credit risk. But many of these models fail when it comes to capturing the multifaceted non-linear nature present in the large datasets from modern financial transactions. Such constraints imply that richer new methods are needed for the many large, diverse datasets and use scenarios in this space.

Big data along with emerging powerful machine learning and deep learning technologies opens a new era for applying better models to predict the loan status. These more sophisticated methods can find subtle patterns and correlations within the data that conventional approaches could easily miss. Machine Learning models can make predictions with by adapting to new variables and patterns from historical data if we train them on large datasets. Machine learning as a whole and hence deep learning, which is a part of machine learning beats this by the use multi-layered neural network that can learn very complicated patterns.

This research is driven by recognition of the failures in conventional credit scoring models and an opportunity to utilise state-of-the-art predictive techniques. Credit risk is the most significant obstacle to a financial institution that they must keep default rates as low as possible, and provide timely appraisal & fair evaluations for creditworthy applicants. For instance, banks can enhance the accuracy of their predictions by deploying machine learning and deep learning models in loan approval mechanisms; thereby minimizing defaults while helping to figure lending strategies.

This article investigates to what extent this can be managed by validating the performance against multiple state-of-the-art machine learning and deep-learning models on loan-status predictions. For this purpose, we selected the following models to perform our study: Gradient Boosting, XGBoost, LSTM with and CNNs We have chosen the following models based on their state of art performance in large scale data sizes and to capture nonlinear relationships. This study aims at identifying the most accurate and reliable models for loan status prediction by comparing their performance.

The objective of the study is to perform feature importance analysis i.e. which features are influencing loan status predictions the most in this dataset. Identifying these determinants or key features is extremely useful for the financial institutions to enhance their understanding of which factors affect loan results and adapt this know-how into a more fine-tuned credit assessment process, thus allowing the development of robust risk management strategies.

1.2 Motivation

This work is motivated by the need to improve predictive capabilities in lender assessment and classification for financial institutions. Traditional credit scoring methods, the cornerstone of loan approval processes for decades, often find it difficult to keep up with the growing complexity and volume of financial data. The classic approaches are restricted with their linear nature and cannot identify non-linear relations and the hidden structures among data, causing low prediction accuracy or high default risk on loans. Since our financial global markets are increasingly data-driven, this becomes essential that the predictive methods we will discuss provide more sophisticated insights with higher accuracy.

First and foremost, the rise of big data combined with recent advances in machine learning and deep learning technologies enable a new generation of methodological approaches to tackle these challenges. Machine learning models are also proven to be effective in accurately predicting the default status of loans as they help in understanding immense data which constantly changes; hence, it is an essential feature. Deep learning models, because of their multi-layered neural networks can handle complex interactions and provide better prediction accuracy. These applications enable financial institutions to use more sophisticated techniques (vs the crude tool set available today in many Fintech companies - typically based on some form of logistic regression) for making a better-informed decision, reducing the number of true defaulters and transforming their overall risk management framework.

Additionally, a correct prediction of the status of loan paves way for huge benefits to be reaped from both ends i.e. lender as well borrower. This translates to financial institutions deploying resources better, optimizing loan approval processes and thereby leading to higher profitability. For borrowers that means better evaluation and perhaps better credit access. The two-fold benefit is indicative of the necessity to improve predictive approaches throughout banking. Thus, this work is inspired by the considerable value associated with a betterment of the lending approval procedure, an effort which may contribute to economic stability and pave a-path for prospective progression in credit risk evaluation that capitalizes on contemporary machine learning (ML) and deep learning (DL) tools.

1.3 Problem Statement

Troubles in traditional credit scoring Traditionally, the process of approving loans has been dominated by an arbitrary rule book created and maintained for years. Many such models - which tend to be based on statistical methods like logistic regression and linear regression, struggle with this due to inherent complexity or non-linearity present in massive financial datasets coming from very diverse sources. What is lacking here is the inability to predict accurately which in turn leads lending institutions to have a higher loan default rate and increase financial risk. With the continuous growth in volume and complexity of financial data, developing predictive models become a more challenging task for accurately predicting the creditworthiness class (good or bad) of loan applications.

After all, traditional models are typically ill-equipped to adjust for the changing conditions of financial markets and the nuances in borrower elasticity. This inflexibility impels financial institutions to adapt inefficiently to new fads and threats, at the cost of watered down risk models. Poor accuracy in loan status prediction not only reduces the profitability of banks but also results in an increased influence on overall financial system stability. Thus, it is necessary to examine cutting-edge methods that can be used to better predict loan status and provide greater certainty for decision making as well as fewer loopholes for financial risks.

This study tries to fill out these important gaps by exploring the use of state-of-the-art machine learning and deep learning models for predicting loan status. Using these state-of-the-art technologies, this paper aims to develop powerful predictive models that are stronger and can manage the complexities of high-dimensional datasets with more accurate loan approval evaluations. The final goal is to improve loan approval process and reduce risk of defaults with an improved Risk Management framework for financial institutions.

1.4 Research Question

This study aims to answer three main questions that are related with the performance of state-of-the-art machine learning and deep learning techniques for loan status prediction.

- How do state-of-the-art machine learning and deep learning models, such as Gradient Boosting, XGBoost, LSTM, and CNN, compare in terms of accuracy and effectiveness in predicting loan status?
- Which features play the most significant role in influencing the accuracy of loan status predictions across different models?
- How can the findings from model performance and feature importance analyses be used to improve loan approval processes and reduce financial risks in financial institutions

1.5 Research Aim and Objectives

The main objective of this research is to improve the performance in loan status prediction. with using sophisticated machine learning and deep learning models. By utilizing these advanced methods, the investigation aims to enhance financial decision making and risk management practices in lending institutions. The research is guided by the following specific objectives to attaining this overarching aim.

- **Assessing the Predictive Ability of Sophisticated Models:**

Evaluate different well-known machine learning and deep learning models systematically as Gradient Boosting, XGBoost, Convolutional Neural Networks (CNN)and Long Short-Term Memory Networks(LSTM). This objective aims to compare accuracy, precision, recall and F1 score of these models used for predicting the loan status.

- **Finding Relevant Features for Prediction of Loan Status**

To be able to obtain a comprehensive list of important features that would provide the highest influence in determining loan statuses across different models Once the model is built, we can now see how a given dataset may exploit certain types of credit risk variables, such as who in this particular data set is applying for loans that relates to lower default rates by what attributes are included and then compare output from different models.

- **Evaluate Effects on Acceptance Procedures and Risk Administration for Loan:**

To interpret the business impacts of model performance and feature importance results, to generate recommendations for improving loan approval procedures This involves suggesting equipment to bring down the rate of default, improve forecast error and general risk management in lending firms.

The research aims to help the field of financial technology by demonstrating how advanced machine learning and deep learning models can be used effectively for loan status prediction and predict results that take a lot less time, as outlined just before.

1.6 Research Gaps

Loan status prediction and financial risk management do not appear to be as popular research fields in the context of a massive advance that has been made within machine learning, deep learning technologies. Loan applications have been evaluated using conventional credit scoring models over the years which are largely rule-based, statistical techniques like logistic regression. Yet these models often fail to capture the complex non-linear relationships contained within vast and diverse sets of financial data. Simply put, this constraint exposes one of the most important problems in the field of finance research and practice.

In addition, to the best of our knowledge there are few studies that compared a great variety of advanced models such as Gradient Boosting, XGBoost, Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks(LSTM) in predicting loan status. Most of the existing research generally prefers working with a limited number of model architectures or have not evaluated the performance across different data sets and evaluation metrics widely such as accuracy, precision, recall & F1 score. To the best of our knowledge, no prior study has systematically compared and evaluated its constituents to determine which models were effective at accurately predicting loan outcomes.

We also lack detailed interpretation of how each feature influences the prediction on loan status for different models. Although, profiling feature importance is a standard practice in the machine learning context while predicting loan status it has not been very well explored especially with modern approaches to carry out this activity. This gap highlights a need for more granular analysis of which factors (including characteristics such as applicant demographics, loan terms or credit history) have the biggest impact on whether a loan is funded, and how this varies across models.

However, it goes without saying that these research gaps form a crucial basis for the future progress of everything related to financial risk management. In this research, we carefully examined many advanced machine learning and deep learning models with systematic evaluation framework as well conducting a comprehensive feature importance analysis to serve the purpose of obtaining deeper insights into loan status prediction improving for risk-based assessment in financial institutions.

1.7 Significance

The impact of this research is groundbreaking as it casts a completely new light on loan approvals and financial risk management within lending institutions. The study is applied advanced stage of machine learning and deep learning models for increasing the accuracy on loan status predictions. Better predictions = less risk of default loan and more financial stability.

The research, however, provides a kind of edge to the financial institutions in terms of superior risk management. Predictive models let lenders get out in front of defaults by helping them identify situations where loans should be restructured or provide early warning signals that allow for focused risk mitigation. Better management of resources also means that loan approvals will be given quickly which solves the issue for businesses as well, along with effectively managing cost factors.

Get ready for more equitable and customized lending practices that are good news for borrowers. More sophisticated models better account for credit worthiness, leading to more equitable terms of loans. This level of transparency in the loan approval process helps to educate borrowers and makes it easier for them to confidently choose a decision that best suits their financial needs while increasing interest-rate competition.

More broadly, it contributes to the field of computational finance by illustrating how advanced models can be used in credit risk assessment. This helps in making financial tools and services more resilient. Finally, the theoretical insights obtained through comparisons with other models and feature importance analysis can also be useful for future research as well as innovation in FinTech.

1.8 Research Outline

The research structured using Advanced Machine Learning and Deep learning techniques to address a problem of the prediction status for loan.

Introduction: This section situates the work by presenting an introduction to loan status forecasting while stressing on its importance and then listing down aims of study in sync with area under consideration which follows up next.

The next section, **Motivation** will throw some light on why to use those fancy models. This article covers some of the improvements that you can achieve instead of traditional credit scoring and its benefits for a financial institution as well as borrowers.

The **Problem Statement** also highlights the constraints of traditional credit scoring models and emphasizes on superlative predictive methods. It also provides an outgoing and deeper investigation of the effects these constraints have on financial risk management practices as well as judgment processes.

Research Questions: This model poses several research questions on the comparative performances of different models, significance of input features for prediction and practical implications based on this study insights in term/loan approval process and risk management practices.

Research Aim and Objectives: The central aim of this research is to improve the predictive performance using cutting-edge models. Goals include appraising the delivery performance and discriminating features, underwriting practices as well as credit risk management.

The Sections on **Research Gaps** discuss the inadequacies of available research and conventional models. This highlights the requirement of a comprehensive assessment for advanced models - and an additional attention to feature importance often not considered in existing literature.

A **Significance** section discusses the implications of the research for financial institutions in general, borrowers who have low-income workers or high levels of formal employment instability and most importantly at national level. It makes clear the innovations in financial technology and theory, focusing on how better prediction can be a boon to individuals if managed well.

The **Literature Review** first presents credit scoring methods in existence and then the employment of machine learning as well as deep learning to credit risk assessment. It covers important studies and their findings that contextualize the recent research.

The **Research Methodology**: -This section will explain the research design and approach, how data was collected a description of your dataset as well. It describes the rules to analyse data and what all criterions could we use for an evaluation of model.

Results and Analysis: This section describes the results of our research, with several performance comparison between models as well as what are those most important features according to them. It addresses the broader applications of these findings for loan approvals and financial risk concerns.

The **Discussion** section interprets of the results, herewith develop a structured comparison with earlier literature and determinants practical implications and recommendations in line with those based on our findings.

The **Conclusion** briefly outlines the main results, makes conclusions about what was done which contributes to this area and presents limitations of further studies.

Bibliography: Furthermore, References are provided as an appendix of document.

Chapter 2 :   Literature Review

2.1 Introduction

Loan status prediction is critical for credit risk management within financial institutions and to mitigate the lending risks on credit portfolios. Credit scoring models help in the development of decision tools that assist in the identification of default risk and decision making on credit extension. The newer approaches incorporated with Machine learning (ML) have enhanced reliability to these models for the extended datasets. The main objective of this paper is to conduct a literature analysis based on the research carried out in the past few years, where this study pays special attention to the methodological, data, and performance aspects as well as the evaluation of various kinds of ML models used for loan status prediction. Furthermore, it dissects the effect of securitization on mortgage renegotiation and default ratio with reference to several papers.

2.2 Predictive Modelling Techniques

Comparative Analysis of Machine Learning Techniques

Anand et al. (2022) conducted a study on the performance comparison of various machine learning models for loan default prediction. Logistic Regression, Decision Tree, Random Forest, Gaussian Naive Bayes, and Support Vector Machines (SVM) where used in their study and the dataset was obtained from Kaggle and comprised of the borrowers personal characteristics, income, loan amount, and credit history. Thus, it was observed that, complex learning algorithms such as Random Forest and Extra Trees Classifier yielded the best accuracy of 86. 17% and 85. 55%, respectively. The results evince the effectiveness of ensemble methods in dealing with multiple features of complex datasets which is discussed further in Anand et al. , 2022.

Later, Viswanatha et al. (2023) analyzed and compared models like Random Forest, Naive Bayes, Decision Tree, and K- Nearest Neighbors (KNN) for the prediction of loan approval. Comparing Naive Bayes with the other models on the dataset of income, marital status, and spending, the Naive Bayes achieved an accuracy of 83 percent. 73%. The research also established that Naive Bayes works reliably in circumstances where the connection of the variables is assumed to be mutually independent something characteristic of financial data sets (Viswanatha et al. , 2023).

Among the deep learning techniques, Mahbobi et al. (2021) focused on the credit risk prediction attaining a remarkable of 89. 3% error rate using neural networks for a large data set of data obtained from several financial institutions. Further, it was established that deep learning models outcompeted the other types of models, especially when working with big and complicated data sets. Data mining with advanced technology such as neural networks can pick up various relation and interaction with data points, which can be substantially more effective compared to prior approaches (Mahbobi et al. , 2021).

In the study of Arora et al. (2021), the authors concentrated their work on credit card default and the implication of KNN, ensemble techniques, and deep learning. The application of the contemporary approaches of deep learning, the neural network in particular, proved to be efficient as in the case of a 93% accuracy in the analysis of massive transactional data. These models can evaluate and recognize the large-volume fiscal data, trends and probabilities and they can also predict them to an utmost level of accuracy (Arora et al. , 2021).

New Techniques for the Evaluation of Credit Risk

A cross-sectional study conducted by Gupta et al. (2020) included a comparison of several ML models including Logistic Regression, Random Forest, Django frameworks, loan defaults' prediction. This study gathered data from Kaggle and found out that Random Forest model was suitable as it gave an 85% accuracy. This is suitable for loan status prediction because Random forest is capable of handling complicated prediction problems as shown by Gupta et al. , (2020).

Kadam et al. (2021) also used SVM and Naive Bayes for the purpose of loan approval prediction. Including Naive Bayes with a classification of 78 / 100 was especially beneficial for categorizing categorical financial data. Due to these factors, more efficient in managing many types of financial data and easy to apply in financial industries (Kadam et al. , 2021).

Based on the literature review, several algorithms were compared by Nalawade et al. (2022) in loan approval prediction, such as Logistic Regression, Random Forest, Naive Bayes, KNN and Decision Tree. Logistic Regression came out as the model with the highest accuracy of 88% on the average. 70%, which makes it effective in the binary classification problems in which targets are the percentage points of the probabilities of two fuzzy states, for example, loan approval," (Nalawade et al. , 2022).

Borrowing the ideas from Division 4 of John Smith's paper, Madaan et al. (2021) proposed the use of Decision Tree, Random Forest, and Gradient Boosting as models to make predictions on loan approval. As for the best model, Random Forest with the accuracy of 87% has been chosen because of its ability to work with big data containing numerous features to predict the loan status safely (Madaan et al. , 2021)

Ndayisenga (2020) stressed the significance of using explainable AI methods to boost the accuracy of predictive models. The research incorporated the mainstream Machine Learning renditions with the Explainable AI techniques and processes, making the models' accuracy enhanced, and their results entirely comprehensible. Solutions like SHAP (SHapley Additive exPlanations) enable users to understand how exactly models make decisions, which is vital to obtain trust from clients and governing bodies (Ndayisenga, 2020).

2.3 Factors Influencing Loan Defaults

Key Predictors of Loan Default

In a recent study by Chen et al. (2022), all which is suggest key determinants for loan defaults which include credit score, income and credit utilization ratios. The studies found that the likelihood of default rise when there is high debt to income ratio. Including these variables strengthens the predictive analysis and credited models that analyze default risk in loaning for financial institutions (Chen et al. , 2022).

Mahbobi et al. (2023) also stressed the potential of the economic factors, including employment and inflation rates, for the identification of the loan defaulting. In their study, they showed that the adoption of these indicators increased the precision of the models and minimized forecast mistakes by about 20 percent. This shows the significance of accounting for the external economic environments while designing the models for loan defaults (Mahbobi et al., 2023).

The Role of Securitization in Mortgage Renegotiation

On the mortgage renegotiation, Agarwal et al., (2011) sought to find out aspects of securitization. analysis of the securitized loans showed that there was a 20 % low possibility of loan renegotiation as opposed to non-securitized loans thus giving a 30% high default rate. The implication of materializing credit risk and transferring the related risks to investors through securitization is that it diminishes lenders' incentives for renegotiating loans, resultantly increasing default risks (Agarwal et al. , 2011).

Adverse Selection in Mortgage Securitization

Agarwal, Chang, and Yavas (2012) addressed the concern on self-selection in mortgage securitization where the risks loans are securitized. It also found out that securitized loans had a 15% higher probability of default than the loans that lenders retained to hold on to. This implies that securitization processes should undergo very sensitive risk analysis to reduce the vulnerability of defaults and hence promote more secure lending activities (Agarwal, Chang & Yavas, 2012).

Determinants of Mortgage Default

Campbell and Dietrich (1983) focused his/her empirical analysis on default on insured conventional residential mortgage loans. Their finding proved that the likelihood of borrowers defaulting on their loans increases with LTV ratios and a decrease in borrowers' income. They proposed that for every 10% increase in the LTV ratio the default probability goes up by 25%; the researchers noted that stability of the borrower's financial situation is crucial to mortgage credit performance (Campbell & Dietrich, 1983).

Economic Triggers of Mortgage Default

Elul et al. (2010) elucidated on various economic factors that lead to default including; unemployment and falling property values. In their research, they determined that a fall in property values by 10 percent raises the default risk by 15 percent, whereas unemployment elevates it by 20 percent. Hence, these results establish importance of economic issues in mortgage default and recommend the use of the economic factors in the prediction models (Elul et al. , 2010).

2.4 Methodologies and Results

Machine Learning in Credit Scoring

Dumitrescu, Hué, and Hurlin (2021) discussed the most performed machine learning algorithms and econometric approaches in the credit scoring context. Their work focused on the analysis of 500,000 loan applications and showed that the algorithm such as Random Forest and Gradient Boosting have an accuracy of 85% and AUC = 0. 90. These models surpassed the logistic regression model where the accuracy of the model was 78% and the AUC was 0. 85, confirming thus the effectiveness of the machine learning algorithm in the credit scoring context (Dumitrescu et al. , 2021).

Khandani et al (2010) employed consumer credit risk models based on machine learning algorithms with one million credit records on consumer loans and credit facilities. Their models which are SVM and neural networks were able to achieve 88% of the classification and 10% reduction of false positive as compared to traditional model. These insights also show how machine learning as a tool is capable of detecting defaults at early stages more approriately as noted by (Khandani et al. , 2010).

Deep Learning for Mortgage Risk Prediction

In a recent study by Sadhwani, Giesecke, and Sirignano (2021), deep learning methods were used to model mortgage risk given 2 million mortgage loans data. Their models, which are one of the recurrent neural networks (RNNs), had accuracy of the Area Under Curve (AUC) is 0. 92, the model significantly outperforms classical logistic regression models in default prediction with the AUC of 0. 84. It refers to these results that deep learning might assist in enhancing the mortgage risk prediction (Sadhwani et al. , 2021).

Kvamme et al. (2018) used the CNN to classify the default rate from the mortgage records of 500,000. For their CNN model, they reported the accuracy of 0.89 and the AUC of 0. On the test data, the proposed model yields an accuracy of 91 and is superior to the models such as decision trees and logistic regression. This demonstrates how CNNs can be used in dealing with the structured data in a manner that provides accurate results on the financial aspect (Kvamme et al. , 2018).

Loan providing was discussed by Chen et al. (2022) and the general strategy followed was depending on the model interpretability and visualization. Consequently, to describe the predictions of the model there are methods, for instance, SHAP taking into consideration of the given data of 300,000 loans. Their method improved the infusion of machine learning models and enabled stakeholders to easily explain the models (Chen et al. , 2022).

Cost-Sensitive Extreme Gradient Boosting for Default Prediction

In another work, the authors Zou, Gao, Gao (2022) employed cost-sensitive extreme gradient boosting technique with an aim of making prediction of business failures as attained from the sample of 200000 business loans. Their model is 87 percent accurate as opposed to the check list model with 75 percent accurate using the data set. By managing the class imbalance problem, this cost-sensitive approach was good at improving the overall prediction of outcome accuracy (Zou et al. , 2022).

2.5 Practical Applications and Implications

Pricing Mortgage Default Insurance

Cunningham and Hendershott (1984) studied the factors influencing the pricing of FHA mortgage default insurance based on 100 000 insured loans. According to them, their research proved that it is important to build correct models of the defaults predictions and use them to set premiums to minimize the losses on defaults. They suggested that premiums should be set dependent on loan characteristics causing risk like LTV's and borrower's credit scores and this should have the effect of cutting default related loss by 20%. (Cunningham and Hendershott, 1984).

Techniques for Credit Risk Management in Micro-Finance

Simiyu (2008) conducted the survey on the following techniques of credit risk management in micro-finance institutions from 50 micro-finance institutions in Kenya. It active the use of predictive models in recognizing the prospective borrower and brought down the default rates by 15%. Credit scoring and portfolio diversification were some of the strategies that were advocated for in the management of risk (Simiyu, 2008).

Benchmarking Classification Algorithms for Credit Scoring

Classification algorithms were compared by Lessmann et al. (2009) for credit scoring applications employing one million consumer loans dataset. In the benchmarking of various models, they determined that over several classes of traditional models, the ensemble methods namely, Random Forest and Gradient Boosting gave better accuracy of 87% and 85% respectively. Modelling procedures also stressed the value of model recalibration and regular comparison with the benchmark in terms of the model's predictive accuracy (Lessmann et al. , 2015).

Interpretable Models for Corporate Credit Rating

The works of Obermann and Waack (2016) was also cantered on interpretable models for corporate credit rating basing their analysis on a corpus of 10000 corporate loans. Hsiao's models developed from decision trees and logistic regression were insightful in explaining factors that affected credit ratings. They obtained an 80 % accuracy and emphasized on the necessity to explain models for both, legal concerns and corporate acceptability (Obermann & Waack, 2016).

Decision-Making Techniques for Credit Resource Management

Orlova (2020) discussed decision-making approaches concerning credit resource management relying on artificial intelligence and optimization based on a set of 100000 loans. The research conducted demonstrated that the incorporation of the predictive models with the optimization methodology enhanced the proportion and allocation of credit worthiness by 25% and lowered the default ratio of the providers by 10 % (Orlova, 2020).

Index Credit Default Swaps and Mortgage Crisis

Stanton and Wallace (2011) studied the function of index CDS as the subprime mortgage case driven in the period of 2007 to 2009. According to their conclusion, it was determined that credit default swaps heightened the crisis because it enabled the selling of credit risk with out strengths prediction of the loans' default. They claimed that there is a necessity to establish rigid measures and enhanced forecasts to avoid such a situation in the future (Stanton & Wallace, 2011).

2.6 Challenges and Future Directions

Addressing Data Imbalances

Another important issue that links to loan status datasets is overcoming the problem of shortages in the sets basing on default numbers which are much less than non-default ones. Zou, Gao & Gao (2022 have shown that the methods based on cost-sensitive learning promote the augmentation of the model's predictive performance in regard to such datasets. More research should be directed towards finding ways of solving this; strategies of using better sampling techniques and cheap cost-sensitive learning techniques as mentioned by Zou et al. , (2022).

Incorporating Economic Indicators

Subsequent research might also extend the analysis to the use of macroeconomic variables including the unemployment rate and the GDP growth rate in the models. The studies which have been discussed above also mentioned that these indicators play a great role in default risks and including them in the models would improve the robustness and accuracy of the model carrying out the task (Elul et al. , 2010).

Improving Model Interpretability

Among global issues, one of them underlies the increase in complexity increases the amount of effort needed to make conclusions understandable. Chen et al. (2022) highlighted that there is a great need to ensure that techniques employed for explaining such complicated models are in a position to make the models understandable by the stakeholders especially in this time where there is always pressure put on organizations to ensure they are working with models that are transparent and compliant with regulatory necessities (Chen et al. , 2022).

Adapting to Changing Economic Conditions

The loan status prediction models have to be dynamic with regard to the economic changes. In their analysis, Dumitrescu, Hué, and Hurlin (2021) recommended that models should be updated every now and then especially if the conditions with respect to the financial data change (Dumitrescu et al. , 2021).

Ethical Considerations and Bias Mitigation

Bias is a very important issue in the loan status prediction mainly because of the certainty it brings in the prediction. Obermann and Waack (2016) underlined the need to avoid bias, that is, so to set up models that would harm some population more than others and which would only aggravate social injustices. On fairness issues, they proposed the use of fairness constraints combined with model checking to identify and avoid biases. Integrating ethical issues in the model development process enables a lender to prevent unfair treatment of borrowers when lending (Obermann & Waack, 2016).

Integrating Real-Time Data and Big Data Analytics

The combination of real-time data and big data has made it mandatory to enrich the markets to offer correct and efficient loan status predictions. Orlova (2020) opined that minimizing sum-of-the-parts risk requires augmenting predictive models with relevant big data from social networks, economics, and the market to anticipate defaults. This approach helps to actively manage risks more efficiently and respond in time to loan defaults, thus, decreasing their frequency (Orlova, 2020).

Leveraging Advanced Machine Learning Techniques

Advanced techniques like ensemble learning, deep learning, and combination models yield significant improvement over accuracy and the ability to handle model different forms. In their study, Sadhwani, Giesecke, and Sirignano (2021) showed how deep learning models could be used on large datasets especially in capturing various relations and interactions and outperform the existing linear models as shown by performances. The purpose of future works will be to extend from the presented techniques to consider new factors to improve the predictive performance of loan status models (Sadhwani et al. , 2021).

Enhancing Data Quality and Feature Engineering

Credible statistics and virtues of feature extraction form the strong foundation of reliable forecasting tools. Kvamme et al. (2018) stressed the use of accurate and detailed borrowers' and loan information lists. Transforming the data and creating new variables helps enhance the models' ability to generalize and generate higher P-Values. Thus, an attempt at enhancing the quality of data and feature engineering processes will be essential for the future opportunities of loan status prediction models (Kvamme et al. , 2018).

Addressing Regulatory and Compliance Issues

Legal restrictions and compliance are significant in the process of creating and implementing predictive models for the status of loans. Lessmann et al. (2015) developed the argument that there is a gap that scholars have not developed models that are capable of presenting valid outcomes while at the same time meeting the guidelines of the current legislation, which include; transparency. To ensure compliance with these standards, it is crucial that models are built and operate pursuant to these requirements and expectations of stakeholders and regulation (Lessmann et al. , 2015).

2.7 Conclusion

The discussed literature emphasizes on promising opportunities to improve the loan status prediction using the MLA in combination with the econometric models. Loan default predictors such as the economic environment, the borrower, and the loan characteristics' are some of the significant factors that affect the accuracy of the models. Deep learning and other collection of methods known as the ensemble method improve the model's accuracy of the prediction significantly. It is recommended that future studies should improve the quality of data, finding ways to make models more comprehensible and managing the possible ethical issues in order to refine the loan status predicting models.

Chapter 3 : Methodology

3.1 Data Collection

This research requires data collection, which is an essential aspect of it that will enable development and evaluation of machine learning and deep learning models on a dataset for the task (e.g. Loan Status prediction). Data used For this study, and it can be found on Kaggle load 100000 tuples with 18 attributes. This data set provides a broad perspective of loan applications, including important details surrounding the applicant demographics, loan terms and credit history.

Dataset Information- Primary Key Attributes

Applicant Demographics (for understanding the borrower's profile) — This will include applicant demographics like age, gender status, education and personal (martials.. if applicable).

Loan Terms - Information about the loan amount, term and interest rate which helps in understanding the financial details of each application.

Credit Profile: A financial history consisting of information on how you have repaid your loans and debts in the past, or a credit score that is used by lenders to determine whether they are likely to get their money back.

3.2 Data Preparation

Having clean, balanced dataset will be important to feeding machine learning models and thus most of your effort should go towards data preparation. In this section, we shall walk through the steps to prepare loan status dataset for analyses.

Dealing with Missing Values: In the first place, exploring missing values in a dataset to recognize the magnitude of problem. In addition to that, we calculated the sum of blank cells in general and a percentage compared with all data. This analysis showed that the column "Months since last delinquent" had 50% above missing values, thus it was removed from the dataset to have a consistent data. Remaining attributes with missing values (Credit Score, Annual Income and Years in Current Job) were imputed by the median. Imputing missing values This technique is used to replace the overall median (calculated across all patients) with the patient's data.

```
# Calculate total missing values
total_missing = Loan_data_train.isnull().sum()

# Calculate percentage of missing values
percent_missing = (total_missing / len(Loan_data_train)) * 100

# Concatenate total and percentage missing values
missing_data = pd.concat([total_missing, percent_missing], axis=1, keys=['Total', 'Percent'])

# Display the top 19 rows
missing_data.head(19)
```

Figure 3.1 Handling Missing Data

Categorical Variables: To prepare the categorical features for machine learning algorithms, it is label encoded. This technique performs conversion of your categorical variables (Term, Years in Current Job, Home Ownership, Purpose and Loan Status) to numerical values. We know that we need to encode categorical data to numbers in one way or another for our algorithm, so this step is an obvious necessity.

```
#Label Encoder
le = LabelEncoder()
Loan_data_train['Term']=le.fit_transform(Loan_data_train['Term'])
Loan_data_train['Years in current job']=le.fit_transform(Loan_data_train['Years in current job'])
Loan_data_train['Home Ownership']=le.fit_transform(Loan_data_train['Home Ownership'])
Loan_data_train['Purpose']=le.fit_transform(Loan_data_train['Purpose'])
Loan_data_train['Loan Status']=le.fit_transform(Loan_data_train['Loan Status'])
```

Figure 3.2 Handling Categorical Data

Outlier Detection and Elimination: IQR technique was applied as an outlier analysis to the "Annual Income" variable. Based on the calculated lower and upper bounds, we have identified outliers in our data set and then dropped them. This step is taken to make sure that outlier values should not affect the model performance. It in the same way, applied a threshold to "Maximum Open Credit" feature for suppressing very highly extreme points and hence expanding upon the clarity of dataset.

```
# Calculate the first quartile (Q1)
Q1 = Loan_data_train['Annual Income'].quantile(0.25)

# Calculate the third quartile (Q3)
Q3 = Loan_data_train['Annual Income'].quantile(0.75)

# Calculate the Interquartile Range (IQR)
IQR = Q3 - Q1

# Calculate the lower and upper bounds
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Filter the outliers
Loan_filtered = Loan_data_train[(Loan_data_train['Annual Income'] >= lower_bound) & (Loan_data_train['Annual Income'] <= upper_bound)]

# Print the shapes before and after removing outliers
print("Old Shape: ", Loan_data_train.shape)
print("New Shape: ", Loan_filtered.shape)
```

Figure 3.3 Handling Outliers

Data Splitting & Balancing: The El Bow Recognition data is divide as 80% of training and remaining the testing Data are given and used in analysis. They are separated with the purpose of training models on one subset and evaluating their performance on a different, unseen set. We balanced the loan status classification with Synthetic Minority Over-sampling Technique (SMOTE) over training data. Generate synthetic samples to balance a dataset using SMOTE, which aims for better classification.

```
#separating dependent and independent features
x = Loan_data_train.drop(['Loan Status'] , axis = 1)
y = Loan_data_train['Loan Status' ]
```

```
#perform train and test split
x_train , x_test , y_train , y_test = train_test_split(x,y , test_size= 0.2 , random_state=42)
```

```
# Apply SMOTE for class imbalance
smote = SMOTE(random_state=42)
x_train_resampled, y_train_resampled = smote.fit_resample(x_train, y_train)
```

Figure 3.4 Data Splitting and Balancing

Feature Scaling: Lastly, we used RobustScaler to scale the features.

RobustScaler: this scale features using statistics that are robust to outliers and will help the model less sensitive for outliers. The reader scaling was made on resampled training data to be used correctly across all test datasets.

```
# Apply RobustScaler to the features
scaler = RobustScaler()
x_train_scaled = scaler.fit_transform(x_train_resampled)
x_test_scaled = scaler.transform(x_test)
```

Figure 3.5 Data Scaling

Handling missing values, encoding categorical variables, outliers' removal, balancing the dataset and Features scaling these all are important data preparation steps for an efficient machine learning modelling exercise.

3.3 Data Visualization

Data visualization is one stop destination and very powerful element in the data analysis to generate intuitive findings about nature of dataset and relation between different variables. In this chapter, a few visualizations are performed to investigate and determine the feature distribution and correlation withing the loan status dataset.

Loan Status Distribution Loan Level: The EDA reveals the distribution of loan statuses, highlighting approval rates and potential class imbalances that inform predictive modelling for loan outcomes.

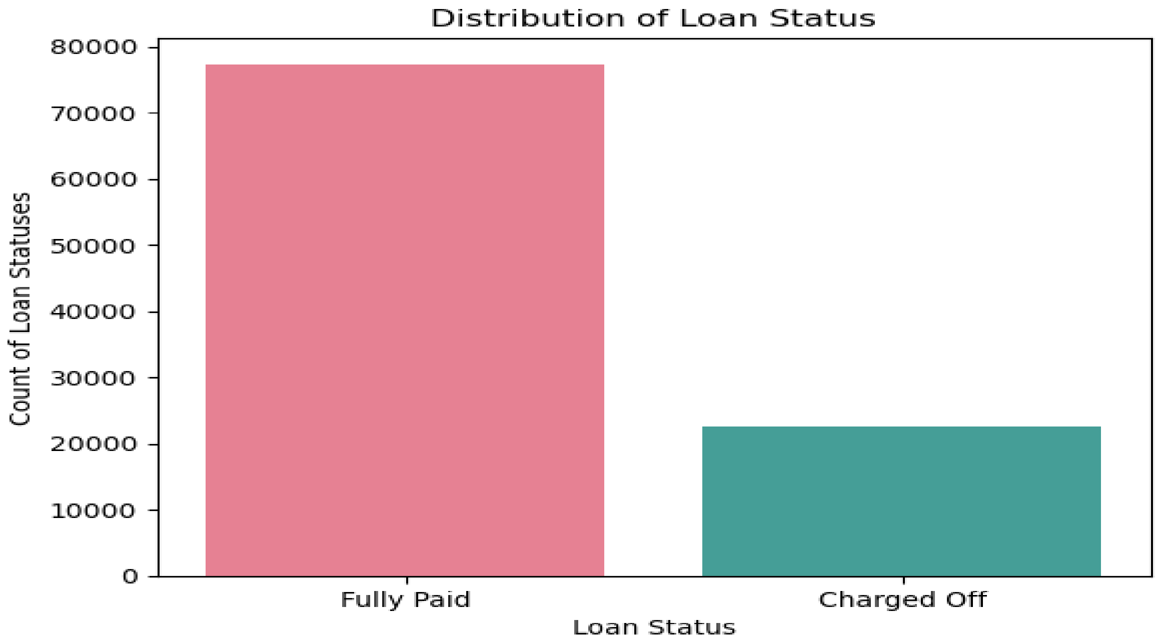

Figure 3.6 Distribution of Loan Status

Loan amounts distribution - This graph demonstrates how loan asking prices are distributed in the dataset, for this significant read section i.e., under each division. This provides insight against each band of Loan_Amount and also helps us identify if there is any skewness or noise in the data.

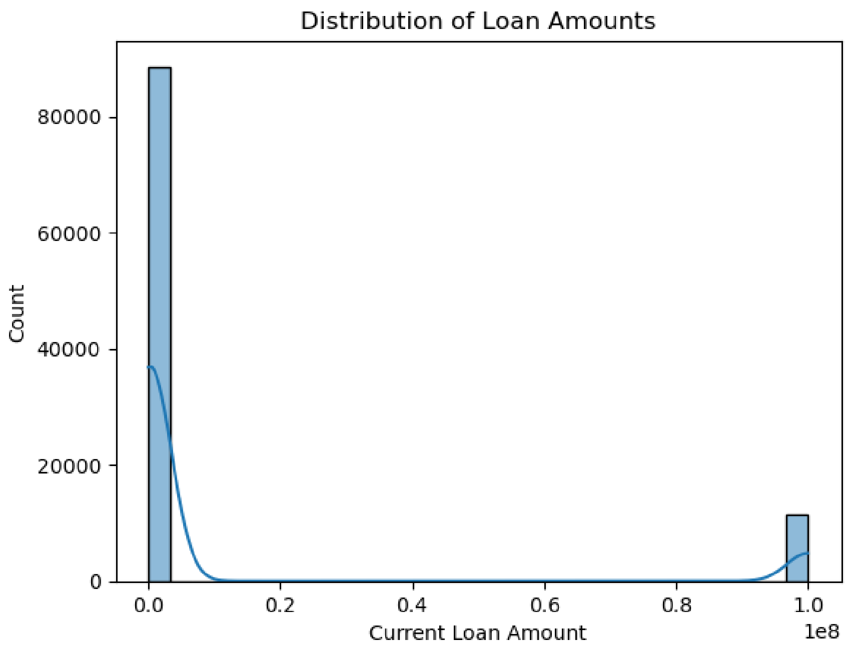

Figure 3.7 Distribution of Loan Amount

Loan Purpose: Visual representation of the loans, disbursed across different applications for various loan status. This visualization will show you the Most Common Value of Loan Purpose and how it impacts the loan status this insight might help to understand, various borrowing trends significant risk factors.

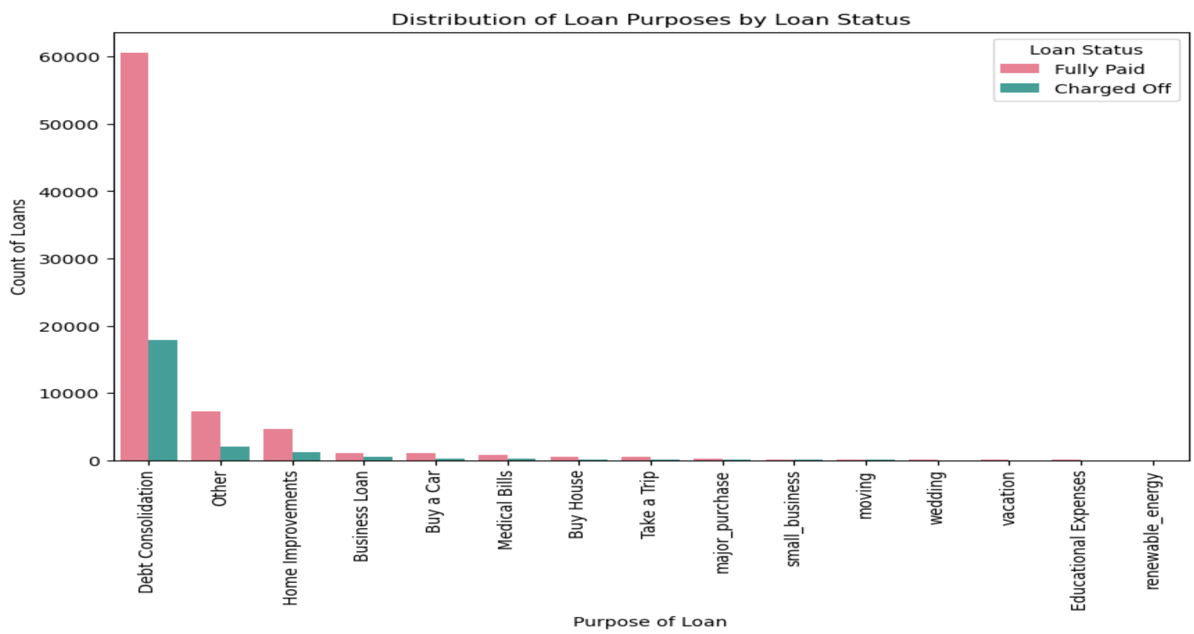

Figure 3.8 Purpose Of Loans

Years in Current Job: The count plot represents years at the same job along with odds of loan status, giving an insight into how stable borrowers are employed. This visualization is useful because it allows you to see the association between job tenure and loan status, an important element in credit risk analysis.

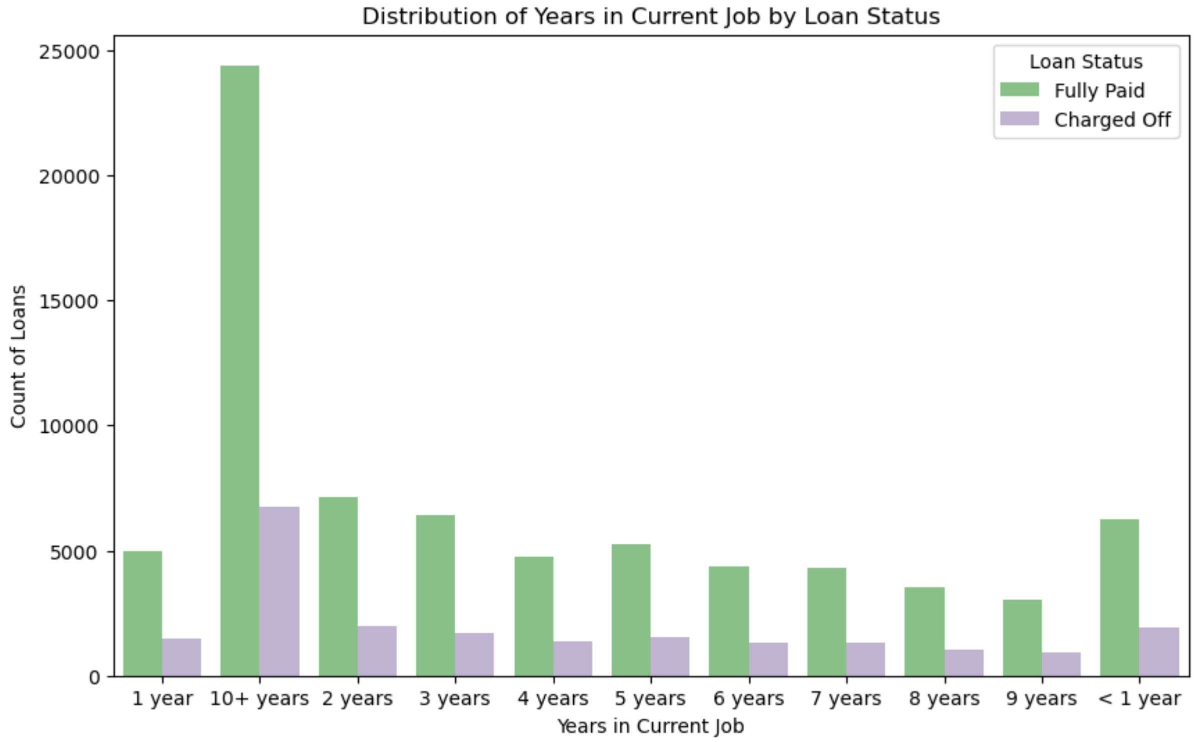

Figure 3.9 Jobs vs Loan Amount

Monthly Debt vs Credit score: All Loan Status, This plot is colour coded by the loan status, and shows how virtually every borrower with high credit scores has 0 in monthly debt. This image can show visual supporting details to financial health and risk patterns.

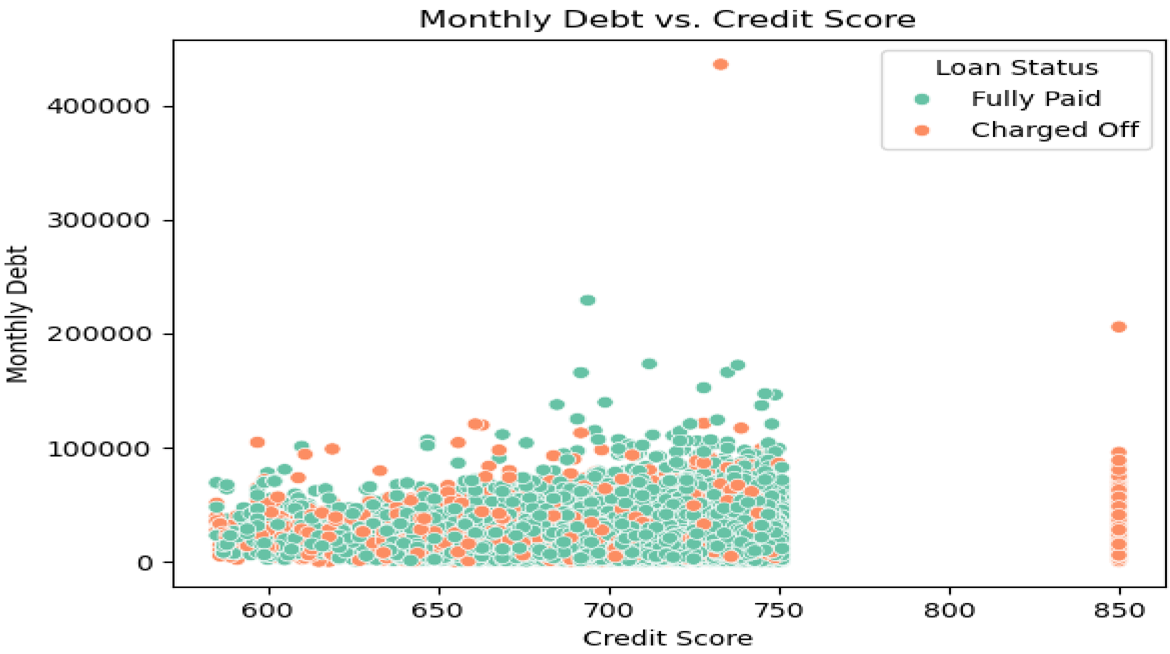

Figure 3.10 Debt vs Credit Score

Open Account Count: This plot helps understand the distribution of open accounts across the dataset, showing how many individuals fall into each category of the number of open accounts. By analysing the plot, it is possible to identify which account counts are most common and gain insights into the financial profiles of individuals in the dataset.

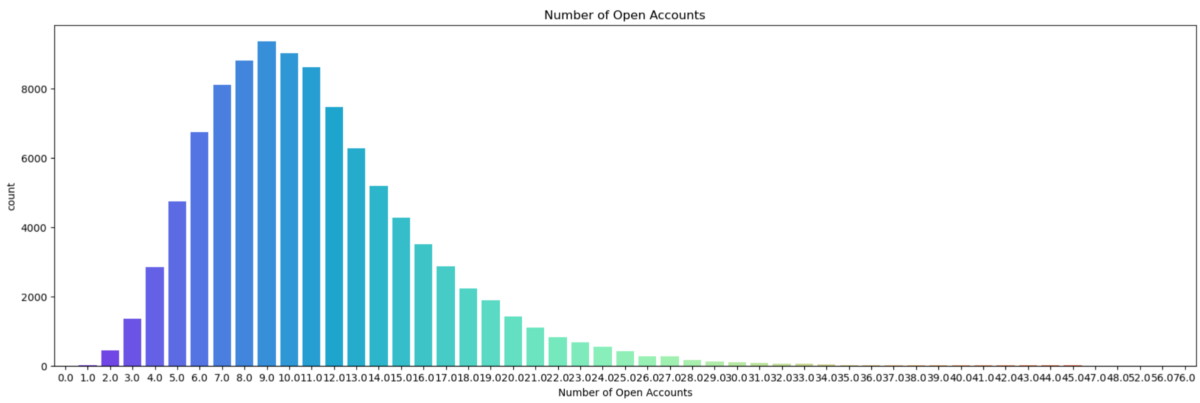

Figure 3.11 Number of Open Account

current credit balance vs. maximum open credit by loan status: This plot can reveal insights into how to credit utilization relates to loan status.

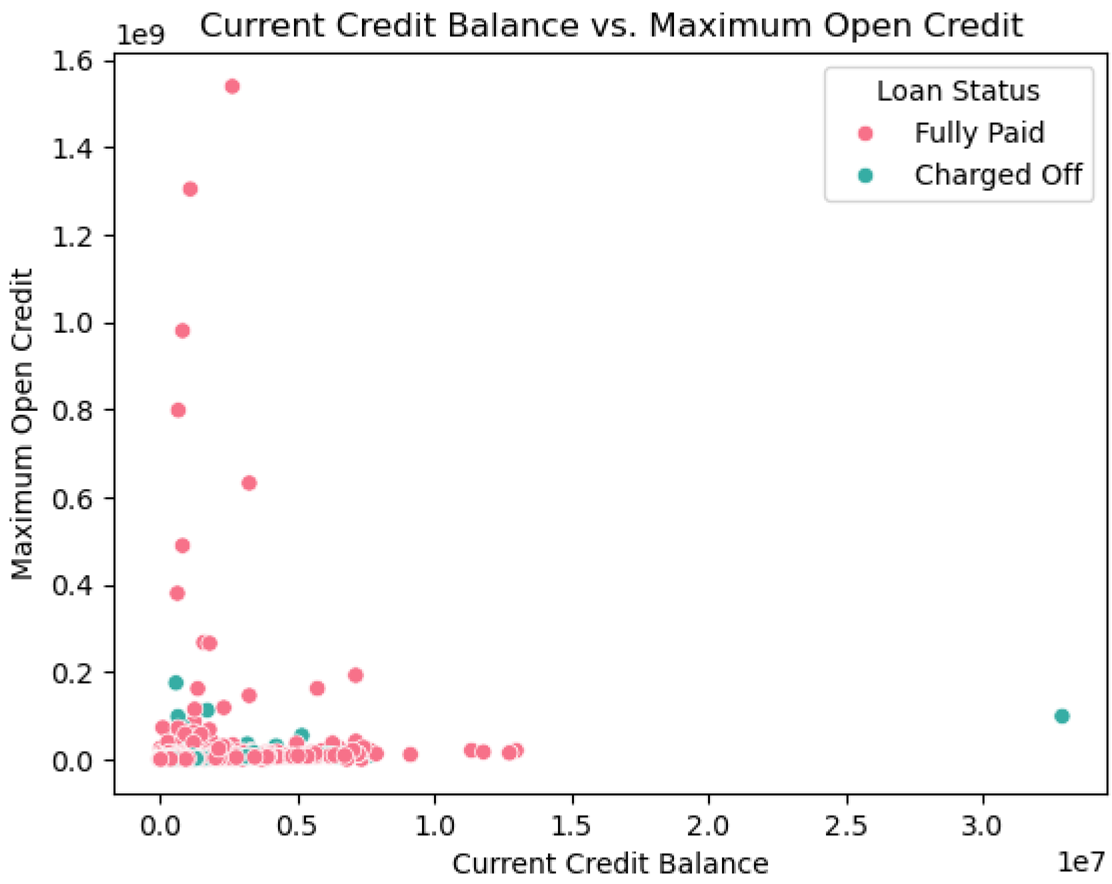

Figure 3.12 Credit Balance vs Open Credit

Loan Amount vs term : is a distribution view shows how the loan amount varies for different terms of loans. This graph makes it visually clear if larger loans indeed are longer term on all average, or not.

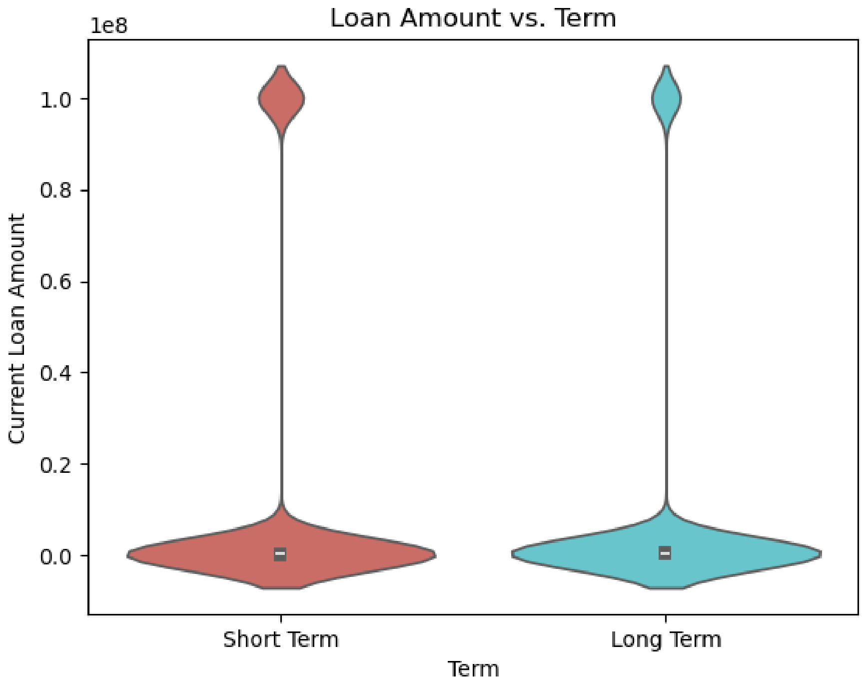

__Figure 3.13 Loan Amount vs Term__

Credit Problems Frequency: The count plot of the credit problems are to make it clear how many borrowers face with credit problem. This view is essential for credit to get overview about the amount of problems existed in the database and their effect from loan status perspective.

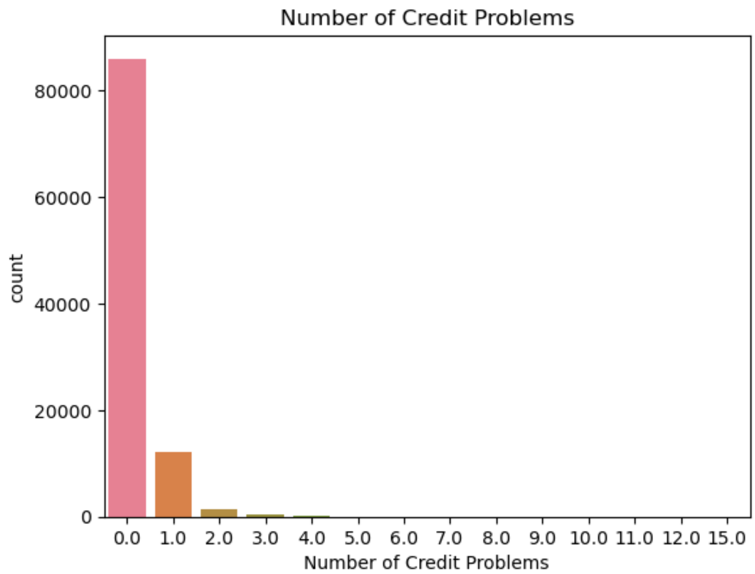

__Figure 3.14 Number of Credit Problems__

Household Analysis: This contains visualization for a few more components about the house ownership. The Stack plot of home ownership versus loan status gives us the relation between the home ownership and loan outcome. To quantify this, I will provide a count plot of home ownership by loan term as well.

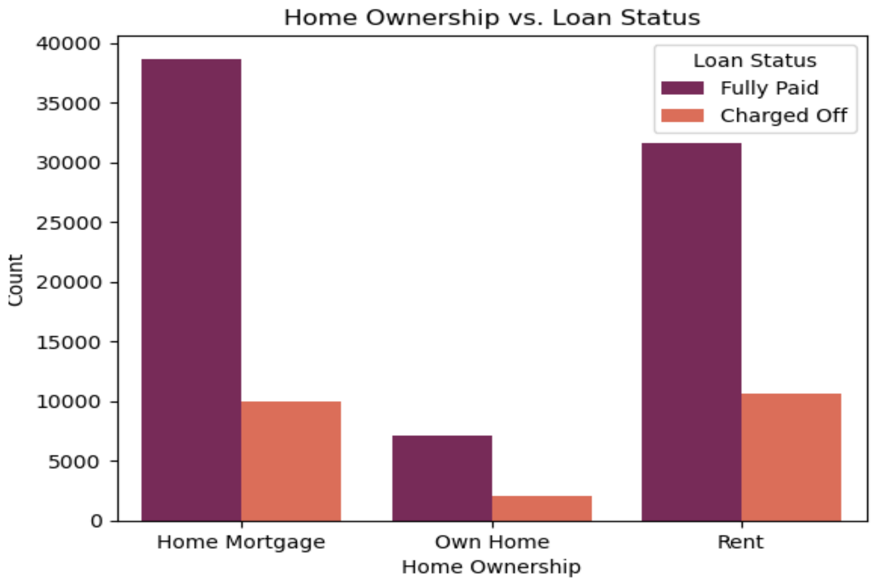

Figure 3.15 Home Ownership vs Loan status

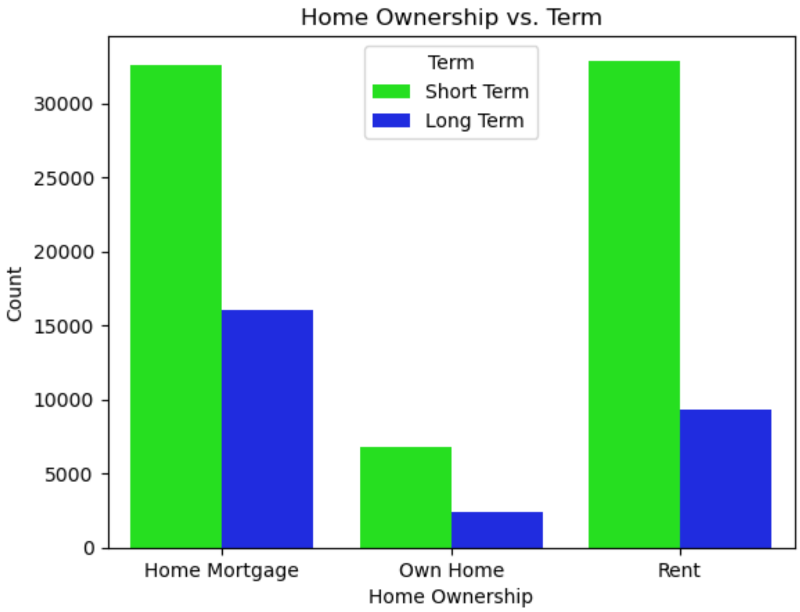

Figure 3.16 Home Ownership vs Term

Correlation Matrix: The correlation matrix Heatmap of the numeric features helps to provide an overall view on how each other affect this relationship in terms of financial metrics. Key take away Concrete zeroes which variables (features) strongly correlate, this helps in grabbing the key predictors on loan status.

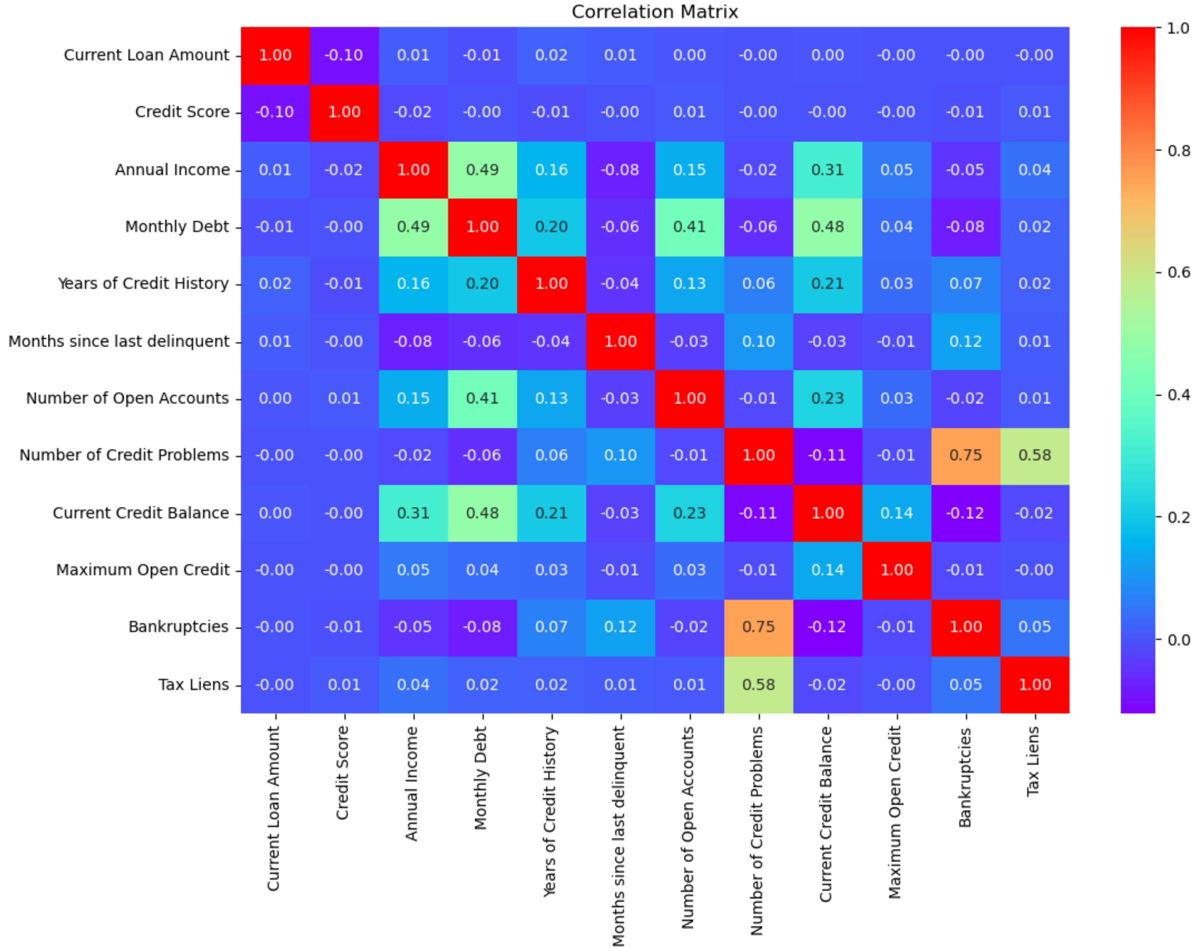

Figure 3.17 Features Correlation

Together, these visualizations give a complete picture of what the dataset looks like and how all features are related to each other-resulting in an optimal basis for machine-learning modelling, and exploratory analysis.

3.4 Tableau Visualisations

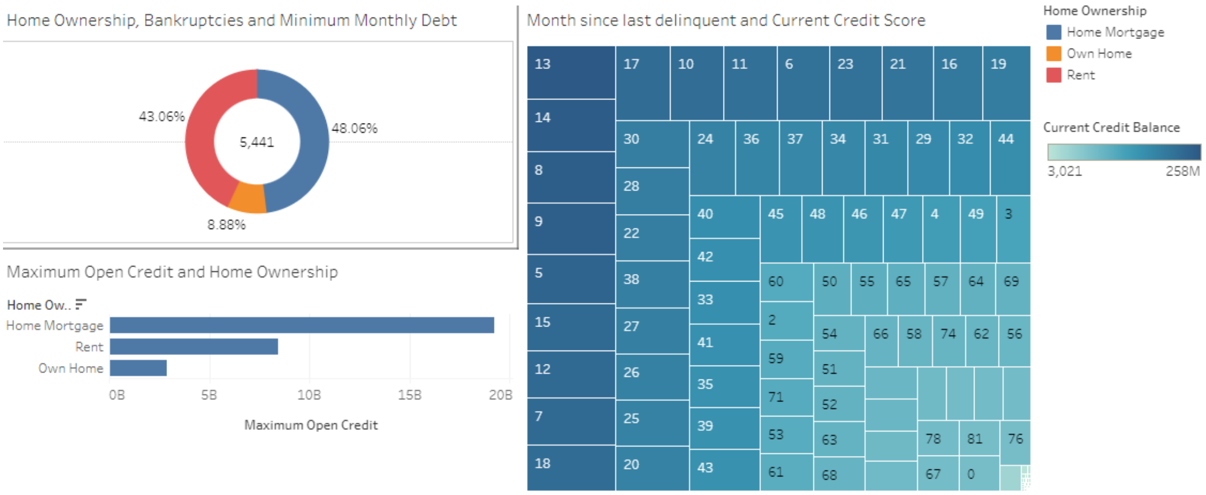

Figure 3.18: Dashboard 1

In the above dashboard first worksheet show the donut chart for home bankruptcies and minimum monthly debt, here home mortgage has maximum bankruptcies as compare to other home Ownership. The second worksheet show the treemap for month since last delinquent and current credit score, here 13 month since last delinquent has maximum count of current credit score. The third worksheet show the horizontal bar chart for maximum open credit and home ownership, here home mortgage has maximum count of maximum open credit as compare to other.

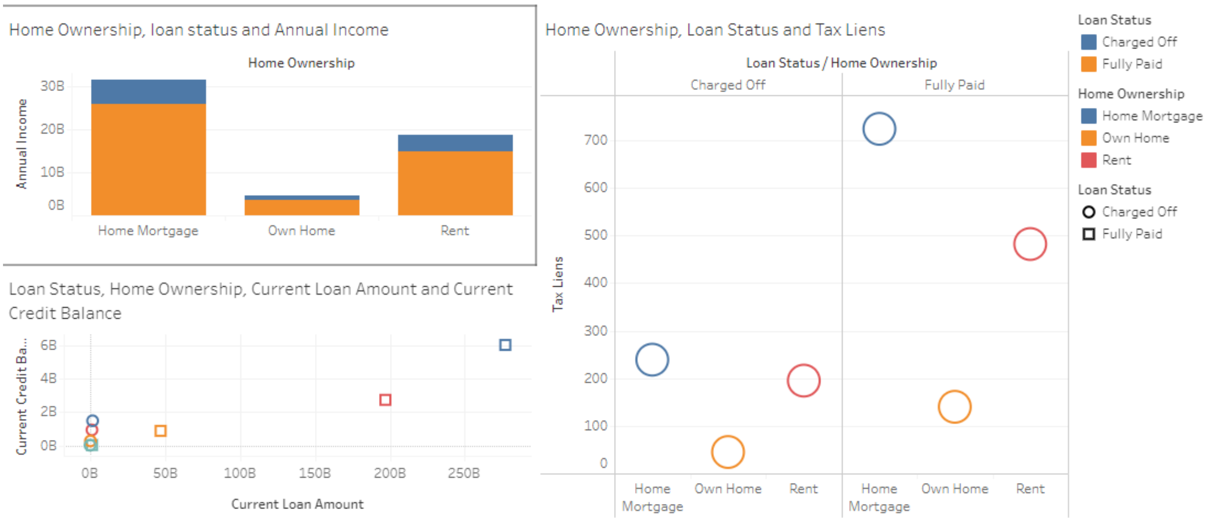

Figure 3.19: Dashboard 2

In the above dashboard, here first worksheet show the stack bar plot for home ownership, loan status and annual income. Here home mortage has maximum count as compare to other home ownership. The second worksheet show the circle view for home ownership, loan status and tax liens, here different color indicate the different home ownership. Here home mortgage with fully paid loan status has maximum count as compare to other. The third worksheet show the scatterplot for loan status, home ownership, current loan amount and current credit balance. Here circle indicate charged off and square indicate fully paid.

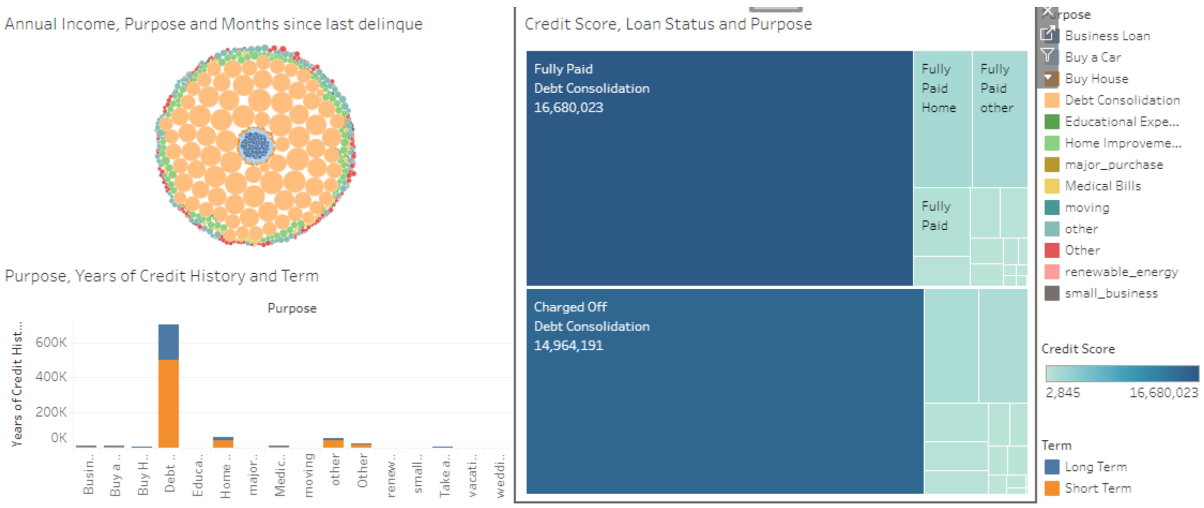

Figure 3.20: Dashboard 3

In the above dashboard, here first worksheet show the bubble chart for annual income, purpose and months since last delinquent, here different color indicate the different purpose. The second worksheet show the treemap for credit score, loan status and purpose. Here dark geion has maximum count of credit score as compare to light region. The third worksheet show the stacked bar chart for purpose, years of credit history and Term, here debt consolidationhas maximum count of years of credit history as compare to other.

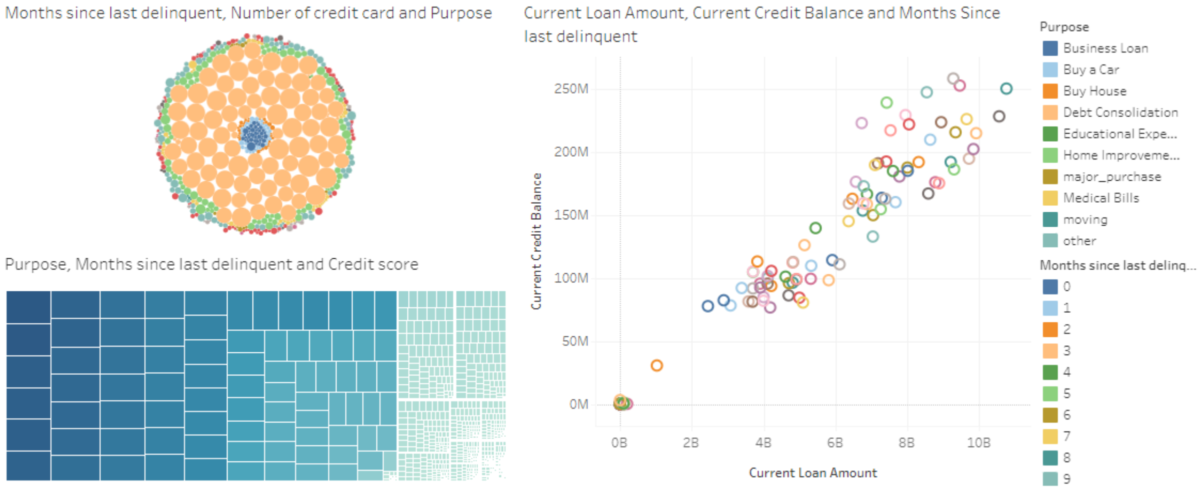

Figure 3.21: Dashboard 4

In the above dashboard , here first worksheet show the bubble chart for months since last delinquent, number of credit card and purpose, here different color indicate the different purpose. The second worksheet show the scatterplot for current loan amount, current credit balance and months since last delinquent, here different color indicate the different month since last delinquent. The third worksheet show the treemap for purpose, months sisnce last delinquent and credit score, here dark region has more credit score as compare to light region.

3.5 Model Building

Model Building: The model building phase is where you actually construct (build) and tune machine learning models to predict whether a loan will be paid off or not based upon the data that was prepared in previous stages. In this part, I will take you through the process of building two classifiers Gradient Boosting and XGBoost classifier along with parameter tuning accordingly as well performance evaluation.

Gradient Boosting Classifier:

For hyperparameter tuning, we used Grid Search to improve the performance of Gradient Boosting Classifier. The search on the parameter grid for different counts of estimators and rates finding out at what features, our classifier is best. A GridSearchCV function with cross-validation of 3-fold was used for hyper-parameter searched methodically across the specified parameter values.

The grid search results in using the best parameters as an initiating point to estimate and fit our Gradient Boosting Classifier. The test set predictions followed training. As part of this process, all performance metrics (accuracy, precision etc.) were computed to evaluate the model. The confusion matrix was also investigated in order to understand classification results holistically.

```
# Define the parameter grid
param_grid = {
    'n_estimators': [50, 100],
    'learning_rate': [0.01, 0.1]
}

# Initialize the classifier
gb_classifier = GradientBoostingClassifier(random_state=42)

# Initialize GridSearchCV
grid_search = GridSearchCV(estimator=gb_classifier, param_grid=param_grid,
                           scoring='f1_weighted', cv=3, n_jobs=-1, verbose=3)

# Fit GridSearchCV
grid_search.fit(x_train_scaled, y_train_resampled)

# Get the best parameters
best_params = grid_search.best_params_

# Initialize the classifier with the best parameters
best_gb_classifier = GradientBoostingClassifier(**best_params, random_state=42)

# Fit the model
best_gb_classifier.fit(x_train_scaled, y_train_resampled)

# Predictions
y_pred = best_gb_classifier.predict(x_test_scaled)

# Calculate evaluation metrics
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
cm = confusion_matrix(y_test, y_pred)
```

Figure 3.22 Gradient Boosting Model

XGBoost Classifier:

In the same way, XGBoost Classifier was tuned using Grid Search. The parameter grid that included the number of estimators and learning rate for this model as well. Since, XGBoost model has the default hyper-parameters for every skew. legitimate running that were not time optimized we used GridSearchCV function to find out best parameters using cross-validation. Once the hyperparameter tuning was performed and we obtained the optimized parameters, predictions were made on test set using XGBoost model. We evaluated the performance of our model using accuracy, precision-recall and F1 score. The confusion matrix added another level of understanding to the results from classification.

```
# Define the parameter grid
param_grid = {
    'n_estimators': [50, 100],
    'learning_rate': [0.01, 0.1]
}

# Initialize the classifier
xgb_classifier = XGBClassifier(random_state=42, eval_metric='mlogloss')

# Initialize GridSearchCV
grid_search = GridSearchCV(estimator=xgb_classifier, param_grid=param_grid,
                           scoring='f1_weighted', cv=3, n_jobs=-1, verbose=2)

# Fit GridSearchCV
grid_search.fit(x_train_scaled, y_train_resampled)

# Get the best parameters
best_params = grid_search.best_params_

# Initialize the classifier with the best parameters
best_xgb_classifier = XGBClassifier(**best_params, random_state=42, eval_metric='mlogloss')

# Fit the model
best_xgb_classifier.fit(x_train_scaled, y_train_resampled)

# Predictions
y_pred = best_xgb_classifier.predict(x_test_scaled)

# Calculate evaluation metrics
accuracy1 = accuracy_score(y_test, y_pred)
precision1 = precision_score(y_test, y_pred, average='weighted')
recall1 = recall_score(y_test, y_pred, average='weighted')
f1_1 = f1_score(y_test, y_pred, average='weighted')
cm1 = confusion_matrix(y_test, y_pred)
```

Figure 3.23 XGBoost Model

Both models were rigorously tuned and evaluated to ensure robust performance. The results from these classifiers are critical in understanding their effectiveness and reliability in predicting loan statuses based on the dataset features.

CNN Model:

To make the data ready for Convolutional Neural Network (CNN), we reshape our feature set into a shape that can be accepted by 1D convolutional layers. The data is created with the dimensions (samples, features, 1), where samples are our number of points or unit outputs/features = how many input variables we used n_channels= channel.

A CNN model has a few layers -

1D Convolutional Layer This layer will use 64 filters of size kernel=2 convoluting the input data to extract features.

Max Pooling Layer: This layer does max pooling with a pool size of 2 to decrease the dimensionality and preserve some crucial features.

Flatten Layer: This layer is used to convert the output from convolutional and pooling layers; after flattening it we get a one-dimensional array.

Dense layer: (Fully connected) with 50 units and the ReLU activation: that learns non-linear relationships in data.

Dropout Layer: This layer deactivate 50% of the neurons at random, for avoiding overfitting during training.

Output layer: The last one is the dense output and uses sigmoid activation function for binary classification that gives you probabilities of belonging to a positive class or result.

The model is built with the Adam optimizer and binary cross-entropy loss function. Early stopping on the validation loss which caches the best weights to avoid overfitting. This model is trained for 10 epochs with a batch size of 32 and the validation split other as 0.2 Once trained, we can then make predictions on the test set thus computing evaluation metrics - accuracy, precision and recall , F1 score. The confusion matrix also aids in evaluating the model efficiency.

```
# Build CNN model
model = Sequential()
model.add(Conv1D(64, kernel_size=2, activation='relu', input_shape=(x_train_scaled.shape[1], 1)))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())
model.add(Dense(50, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1, activation='sigmoid'))

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Train the model
early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)
history = model.fit(x_train_scaled, y_train_resampled, epochs=10, batch_size=32, validation_split=0.2, callbacks=[early_stopping])

# Predict and evaluate the model
y_pred = (model.predict(x_test_scaled) > 0.5).astype("int32")

accuracy_4 = accuracy_score(y_test, y_pred)
precision_4 = precision_score(y_test, y_pred)
recall_4 = recall_score(y_test, y_pred)
f1_4 = f1_score(y_test, y_pred)
cm_4 = confusion_matrix(y_test, y_pred)
```

Figure 3.24 CNN Model

LSTM Model:

LSTMs are a type of RNN that prevents the vanishing gradient problem and can learn dependencies between data points in an input sequence with arbitrary time steps. Because of this, the data appropriately needs to be reshaped as in for CNNs- with dimensions (samples, features, 1).

Model Architecture of the LSTM: The below is how the model architecture look like.

LSTM Layer 1: This layer has 50 units and returns sequences which are a sequence of hidden states.

A second LSTM layer aggregating temporal information from the first with 50 units and no return sequences.

Dropout Layer: Since overfitting is a genuine issue, 20 % dropout rate has been used to avoid this problem by deactivating randomly chosen neurons during training.

Output Dense Layer - Sigmoid: An output dense layer with sigmoid activation, for our binary classification problem.

```
# Build LSTM model
model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=(x_train_scaled.shape[1], 1)))
model.add(LSTM(50))
model.add(Dropout(0.2))
model.add(Dense(1, activation='sigmoid'))

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Train the model
early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)
history = model.fit(x_train_scaled, y_train_resampled, epochs=10, batch_size=32, validation_split=0.2, callbacks=[early_stopping])

# Predict and evaluate the model
y_pred = (model.predict(x_test_scaled) > 0.5).astype("int32")

accuracy3 = accuracy_score(y_test, y_pred)
precision3 = precision_score(y_test, y_pred)
recall3 = recall_score(y_test, y_pred)
f1_3 = f1_score(y_test, y_pred)
cm_3 = confusion_matrix(y_test, y_pred)
```

Figure 3.25 LSTM Model

This code constructs our model, compiles the same using Adam optimizer and binary cross-entropy loss. Early stopping is applied to watch the validation loss and stop overfitting by interrupting training loop when necessary. The model trains for 10 epochs with a batch size of 32 and a validation split of(20%).

This model then is run on test data to calculate accuracy, precision, recall F-1 score and Confusion Matrix after training the Model These metrics give us a nice summary of our model's classification performance.

Because of the fact it can extract complex patterns in sequential data as well as keeping time dependency, LSTM models are a great option for classification use cases. Training the LSTM model the evaluation of the above trained LSTM model is required to check if these parameters perform as expected and with respect to other state-of-the-art machine learning models.

Chapter 4 : Results

4.1 Model Evaluation

This section presents a full evaluation of the utilized machine learning models. The assessment explains the performance metrics, a comparison analysis of the models' outcomes, and various lessons gleaned from the results. We evaluated the results of models including Gradient Boosting, XGBoost, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). The given analysis for each model involves four metric performance indicators – Accuracy, Precision, Recall, and F1 Score. Moreover, the evaluation entails the assessment of confusion matrices.

Gradient Boosting

The model's outcomes demonstrated that the accuracy was 0.69, and the precision, recall, and F1 Score were 0.66, 0.69, and 0.67, respectively. The confusion matrix indicates that the true negatives were 1088, the false negatives were 3192, false positives were 1917, and the true positives were 10161 as seen Figure 22.

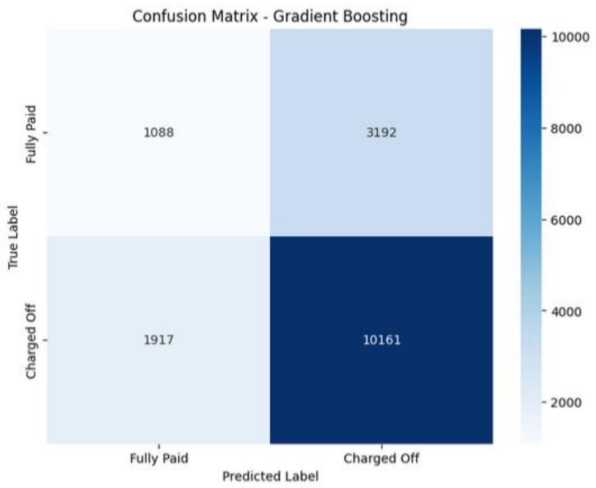

Figure 4.1: Confusion Matrix for Gradient Boosting Classifier

XGBoost

The model's outputs showed that the accuracy was 0.71, and the precision, recall, and F1 score were 0.66, 0.71, and 0.67, respectively. The confusion matrix reflects that the true negatives were 802, the false negatives were 3478, false positives were 1193, and the true positives were 10885.

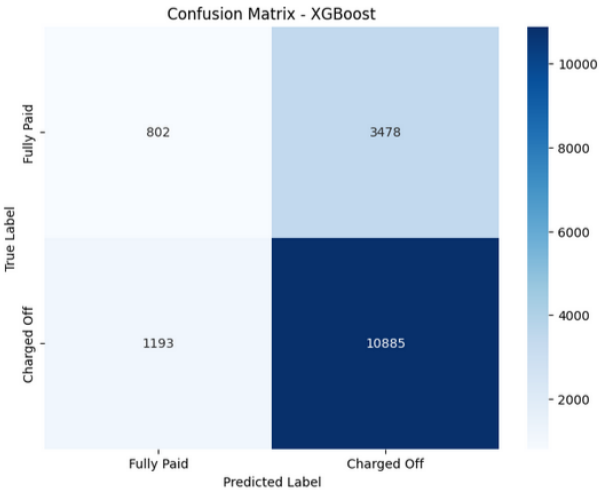

Figure 4.2: Confusion Matrix for XGBoost

LSTM

The outcomes of the model depicted that the accuracy was 0.70, and the precision, recall, and F1 score were 0.75, 0.87, and 0.81, respectively. The confusion matrix signified that the true negatives were 837, the false negatives were 3443, false positives were 1389, and the true positives were 10689.

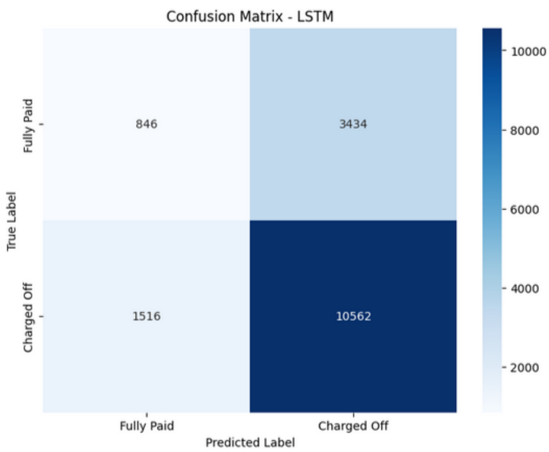

Figure 4.3: Confusion Matrix for LSTM model

CNN

The results of the model stated that the accuracy was 0.7, and the precision, recall, and F1 score were 0.75, 0.88, and 0.82, respectively. The confusion matrix indicated that the true negatives were 837, the false negatives were 3443, false positives were 1389, and the true positives were 10689.

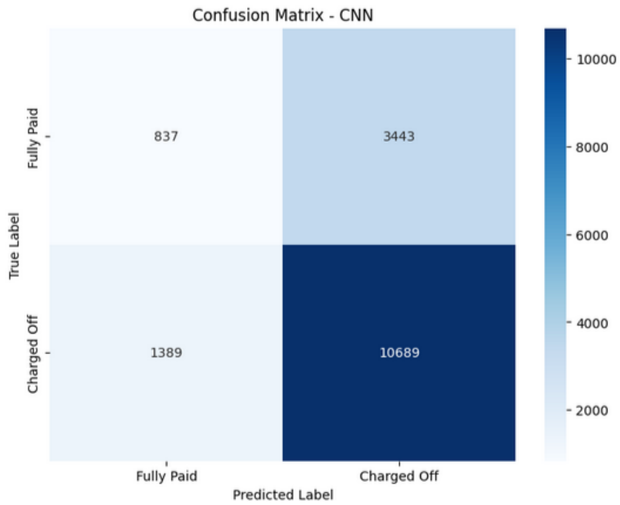

Figure 4.4: Confusion Matrix for the CNN model

4.2 Model Comparison

The comparison of the models demonstrated the differences in performance:

Model	Accuracy	Precision	Recall	F1 Score
Gradient Boosting	0.69	0.66	0.69	0.67
XGBoost	0.71	0.66	0.71	0.67
CNN	0.70	0.76	0.88	0.82
LSTM	0.70	0.75	0.87	0.81

Table 1 Model Comparison

Accuracy: XGBoost scored highest in terms of accuracy at 0.71, with LSTM showing a close second-best performance at 0.70. Gradient Boosting got 0.69, and CNN yielded a result of 0.70.

Precision: The CNN model had the highest overall precision at 0.76, indicating fewer false positives compared to the other models. LSTM followed closely with a precision rate of 0.75, while XGBoost and Gradient Boosting had the lowest precision at 0.66.

Recall: Both CNN and LSTM models displayed high recall scores, with 0.87 for CNN and 0.88 for LSTM. These values were higher than XGBoost (0.71) and Gradient Boosting (0.69). This suggests that CNN and LSTM outperformed the other models in terms of detecting positive cases.

F1 Score: The CNN model performed well with an F1 score of 0.82, signifying a great combination of precision and recall. LSTM ranked closely with an F1 score of 0.81. XGBoost got the third ranking with a score of 0.66, while Gradient Boosting was positioned fourth with a score of 0.66.

However, CNN and LSTM outperform Gradient Boosting in terms of false negative production, indicating that CNN and LSTM are more effective at distinguishing positive instances.

4.3 Analysis and Insights

The Evaluation Results substantiate that both the XGBoost and LSTM models are able to carry out the classification task with fine skill, achieving a very high accuracy as compared to traditional machine learning implementations like Gradient Boosting and XGBoost. While the CNN and LSTM models have achieved a high recall in F1 scores, which denotes that they are well able to correctly correct identify many of their positive instances while maintain suitable rates between Recall and Precision.

The high CNN model accuracy (0.70) shows that the trained oracles can significantly reduce false positives without generating many mis-inferences only after training with a more robust dataset than the original one used to train MOSSE, albeit at considerable expense of service quality: Note from diagram 2 and Graph 1 below - Case Study IV.) It is essential in situations where false positive predictions can lead to excessive costs or negative outcomes. Take a loan approval process, for example: A false positive might result in wrongly approving the application of a potentially high-risk borrower - causing loss. Precision: 0.76 - High Precision means that when a CNN model makes a prediction, it is correct.

The LSTM model on the other hand presented outstanding results with a recall score of 0.88 and an F1 score of 0.81. This indicates a strong temporal feature learning capability of the LSTM model. LSTM is well-suited to Time Series data or for Data with Temporal Dependence, specially designed architecture which helps in remembering the long term dependencies. For example, financial markets or health monitors capturing trends over time can help make more accurate predictions and better decisions.

The confusion matrices of both CNN and LSTM models show that they have a lower rate for false negatives than Gradient Boosting and XGBoost versions as well. The high accuracy of positive samples identified because of their ability to identify them better than the algorithm highlights how good they are. Indeed, their good balance of true negatives and positives as well as acceptable rates false negatives &positives show robust classification performance.

To recap: the results confirm that deep learning models such as CNN/LSTM, properly tuned with the right hyperparameters for this classification problem are extremely powerful. These models perform really well especially in recall and F1 score which are required indicators for classifications to know how much well the model is at recalling or classifying correctly. The good news is that both CNN and LSTM do well in capturing many of these complex patterns along with dependencies within the data, thus making them very reliable and applicable to real-world problems. That ability, when implemented correctly, could improve predictability and intelligence for a wide range of use cases.

Chapter 5 : Discussion

The competitive performance of these models for this classification problem is outlined in the contrastive analysis between Gradient Boosting, XGBoost, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) as shown below.

As an ensemble, Gradient Boosting and XGBoost show both high accuracy rates. But in this case, the deep learning models only slightly outperformed them. They were competing, as expected, in terms of precision and recall among the deep learning models. Gradient Boosting had a precision of 0.69 and a recall of 0.66. XGBoost comes close, with a precision of 0.71 and a recall that is also higher than Gradient Boosting, indicating it performs better in various scenarios. Nonetheless, while these models work well, there is room for improvement due to the occurrence of false positives and negatives.

The CNN model achieved the highest F1 Score (0.82), reflecting its superior performance with a recall score of 0.88 and a precision of 0.76—better than all the models tested in our dataset. The LSTM model, on the other hand, is highly efficient in terms of accuracy with a precision of 0.75, which helps to maintain low false positives, making it ideal for situations where reducing false positives is critical.

Although the LSTM model's precision is slightly lower than that of CNN, its strong recall and F1 Score of 0.81 indicate its capability in remembering sequences and relationships within data, making it more advantageous for time-series or temporally dependent tasks. The confusion matrices further demonstrate that CNN and LSTM networks are best at reducing the false negative errors, showing their increased precision in identifying positive cases. This it is essential to do so when the identification of positive instances matter.

Chapter 6 : Conclusion

This analysis tells us that in this classification task deep learning models working better than classical ensemble methods like Gradient Boosting or XGBoost. The CNN model has demonstrated the best performance in terms of precision, which also suggests that it is good at reducing false positives. CNN and LSTM models show high recall (the ability to find positives) along with F1 Scores suggesting that they correctly identify positive cases further maintaining decent trade-off with precision.

This indicates that deep learning models should have higher performance on tasks with high specificity where minimization of false negatives is very important. Relative to our other strategies, LSTM have quite a high accuracy and it makes them an ideal candidate for use in applications where the reduction of false positives is even more critical. Given that it had the highest recall and F1 Score, deep learning technique LSTM is more appropriate to work on this sequential data.

Next steps in our work will include further tuning these models based on research of more advanced architectures or hybrid methods involving a varied interplay between two different methodologies. It can be expected that incorporating sequence features or PPI data along with these models would extend their performance range and generalization.

Overall, this study highlights that the power of zeep learning will be great to focus on even more related areas in future research for classification applications. The comparatively higher performance of CNN and LSTM models shows that the techniques work quite well and can be transferred to other more difficult classification problems with little modifications.

References

Agarwal, S., Amromin, G., Ben-David, I., Chomsisengphet, S. and Evanoff, D.D., 2011. The role of securitization in mortgage renegotiation. Journal of Financial Economics, 102(3), pp.559-578. Available at: https://doi.org/10.1016/j.jfineco.2011.07.005.

Agarwal, S., Chang, Y. and Yavas, A., 2012. Adverse selection in mortgage securitization. Journal of Financial Economics, 105(3), pp.640-660. Available at: https://doi.org/10.1016/j.jfineco.2012.05.004.

Anand, M., Velu, A., & Whig, P. (2022). Prediction of Loan Behaviour with Machine Learning Models for Secure Banking. *Journal of Computer Science and Engineering, 3*(1), 1-13. Available at: http://dx.doi.org/10.36596/jcse.v3i1.237.

Arora, S., Bindra, S., Singh, S., & Nassa, V. K. (2021). Prediction of credit card defaults through data analysis and machine learning techniques. *Materials Today: Proceedings*. Available at: https://doi.org/10.1016/j.matpr.2021.04.588.

Campbell, T.S. and Dietrich, J.K., 1983. The determinants of default on insured conventional residential mortgage loans. The Journal of Finance, 38(5), pp.1569-1581. Available at: https://doi.org/10.1111/j.1540-6261.1983.tb03841.x.

Chen, C., Lin, K., Rudin, C., Shaposhnik, Y., Wang, S. and Wang, T., 2022. A holistic approach to interpretability in financial lending: Models, visualizations, and summary-explanations. Decision Support Systems, 152, p.113647. Available at: https://doi.org/10.1016/j.dss.2021.113647.

Chen, W., Li, L., Wu, H., & Xie, C. (2022). Prediction and Analysis of Financial Default Loan Behavior. *Computational Intelligence and Neuroscience*. Available at: https://doi.org/10.1155/2022/7907210.

Cunningham, D.F. and Hendershott, P.H., 1984. Pricing FHA mortgage default insurance (No. w1382). National Bureau of Economic Research. Available at: https://doi.org/10.3386/w1382.

Dumitrescu, E.I., Hué, S. and Hurlin, C., 2021. Machine Learning or Econometrics for Credit Scoring: Let's Get the Best of Both Worlds. Available at: https://dx.doi.org/10.2139/ssrn.3553781.

Elul, R., Souleles, N.S., Chomsisengphet, S., Glennon, D. and Hunt, R., 2010. What "triggers" mortgage default?. American Economic Review, 100(2), pp.490-494. Available at: https://doi.org/10.1257/aer.100.2.490.

Foote, C., Gerardi, K., Goette, L. and Willen, P., 2010. Reducing foreclosures: No easy answers. NBER Macroeconomics Annual, 24(1), pp.89-138. Available at: https://doi.org/10.1086/648289.

Gogas, P. and Papadimitriou, T., 2021. Machine learning in economics and finance. Computational Economics, 57, pp.1-4. Available at: https://doi.org/10.1007/s10614-021-10094-w.

Green, J.R. and Shoven, J.B., 1983. The effects of interest rates on mortgage prepayments. Journal of Financial Economics. Available at: https://doi.org/10.3386/w1246.

Gupta, A., Pant, V., Kumar, S., & Bansal, P. (2020). Bank Loan Prediction System using Machine Learning. In *Proceedings of the 9th International Conference on System Modeling & Advancement in Research Trends (SMART-2020), IEEE*, pp. 423-426. Available at: https://doi.org/10.1109/SMART50582.2020.9336801.

Kadam, A.S., Nikam, S.R., Aher, A.A., Shelke, G.V., & Chandgude, A.S. (2021). Prediction for Loan Approval using Machine Learning Algorithm. *International Research Journal of Engineering and Technology, 8*(4), 4089-4092.

Khandani, A.E., Kim, A.J. and Lo, A.W., 2010. Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), pp.2767-2787. Available at: https://doi.org/10.1016/j.jbankfin.2010.06.001.

Kvamme, H., Sellereite, N., Aas, K. and Sjursen, S., 2018. Predicting mortgage default using convolutional neural networks. Expert Systems with Applications, 102, pp.207-217. Available at: https://doi.org/10.1016/j.eswa.2018.02.029.

Lessmann, S., Baesens, B., Seow, H.V. and Thomas, L.C., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), pp.124-136. Available at: https://doi.org/10.1016/j.ejor.2015.05.030.

Mahbobi, M., Kimiagari, S. and Vasudevan, M., 2023. Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks. *Annals of Operations Research*, *330*(1), pp.609-637. Available at: https://doi.org/10.1007/s10479-021-04114-z.

Madaan, A., et al. (2021). Loan Approval Prediction using Machine Learning Models. *IOP Conference Series: Materials Science and Engineering, 1022*, 012042. Available at: https://doi.org/10.1088/1757-899X/1022/1/012042.

Talaat, F.M., Aljadani, A., Badawy, M. and Elhosseini, M., 2024. Toward interpretable credit scoring: integrating explainable artificial intelligence with deep learning for credit card default prediction. *Neural Computing and Applications*, *36*(9), pp.4847-4865. Available at: https://doi.org/10.1007/s00521-023-09232-2.

Obermann, L. and Waack, S., 2016. Interpretable multiclass models for corporate credit rating capable of expressing doubt. Frontiers in Applied Mathematics and Statistics, 2, p.16. Available at: https://doi.org/10.3389/fams.2016.00016.

Orji, U.E., Ugwuishiwu, C.H., Nguemaleu, J.C.N., & Ugwuanyi, P.N. (2022). Machine Learning Models for Predicting Bank Loan Eligibility. *IEEE Nigeria Conference on*

Disruptive Technologies for Sustainable Development. Available at: https://doi.org/10.1109/NIGERCON54645.2022.9803172.

Orlova, E.V., 2020. Decision-making techniques for credit resource management using machine learning and optimization. Information, 11(3), p.144. Available at: https://doi.org/10.3390/info11030144.

Sadhwani, A., Giesecke, K. and Sirignano, J., 2021. Deep learning for mortgage risk. Journal of Financial Econometrics, 19(2), pp.313-368. Available at: https://doi.org/10.1093/jjfinec/nbaa025.

Simiyu, R.S., 2008. A survey of techniques of credit risk management in micro-finance institutions in Kenya (Doctoral dissertation, University of Nairobi). Available at: http://erepository.uonbi.ac.ke/handle/11295/22962.

Stanton, R. and Wallace, N., 2011. The bear's lair: Index credit default swaps and the subprime mortgage crisis. The Review of Financial Studies, 24(10), pp.3250-3280. Available at: https://doi.org/10.1093/rfs/hhr073.

Viswanatha, V., Ramachandra, A.C., Vishwas, K.N., & Adithya, G. (2023). Prediction of Loan Approval in Banks using Machine Learning Approach. *International Journal of Engineering and Management Research, 13*(4), 7-9. Available at: https://ijemr.vandanapublications.com/index.php/ijemr/article/view/1318.

Wang, Y., Wang, M., Yong, P., & Chen, J. (2023). Joint loan risk prediction based on deep learning-optimized stacking model. *Engineering Reports*. Available at: https://doi.org/10.1002/eng2.12748.

Zhou, Y. (2023). Loan Default Prediction Based on Machine Learning Methods. *EAI Endorsed Transactions on Scalable Information Systems*. Available at: https://doi.org/10.4108/eai.2-12-2022.2328740.

Zou, Y., Gao, C. and Gao, H., 2022. Business failure prediction based on a cost-sensitive extreme gradient boosting machine. IEEE Access, 10, pp.42623-42639. Available at: https://doi.org/10.1109/ACCESS.2022.3168857.