University of Cincinnati

IT Project

**Machine Learning-Based Classification of Credit Scores Using Customer Financial Data**

By

Suryateja Chennapragada

Submitted to:

JIM SCOTT

# Abstract

This project mainly concerns with the construction of a credit score classifier using customer related financial data in the context of credit scoring. The aim was to develop an intelligent system that can categorize credit scores under good, standard, and poor categories. The data set was obtained from Kaggle and comprised parameters including the income, loan record, credit card utilization, and record of delayed payments. Preprocessing covered missing value handling, feature creation, and scaling to the feature space to which several machine learning algorithms were applied, including Logistic Regression, Random Forest, Gradient Boosting and Decision Trees. Random Forest as well as an ensemble voting classifier had the best accuracy with an ensemble model scoring 88% accuracy. This model was then incorporated into a Flask web application for real-time generation of the credit scores given the users input. This system provides a realistic method whereby financial institutions can enhance the efficiency of credit risk evaluation and resolve.

***Keywords:*** Credit score classification, machine learning, financial data, Random Forest, ensemble model, Flask web application, credit risk assessment

Table of Contents

[Abstract 2](#_Toc184221770)

[Chapter 1 Introduction 7](#_Toc184221771)

[1.1 Problem Statement 7](#_Toc184221772)

[1.2 Contribution to Current Literature 8](#_Toc184221773)

[1.3 Importance of the Study 8](#_Toc184221774)

[1.4 Research Questions 9](#_Toc184221775)

[1.5 Background/Context 9](#_Toc184221776)

[1.6 Aim & Objectives 9](#_Toc184221777)

[Chapter 2: Literature Review 11](#_Toc184221778)

[2.1 Machine Learning Applications in Credit Scoring 11](#_Toc184221779)

[2.2 Machine Learning Model and Data Base in Credit Risk Assessment 13](#_Toc184221780)

[2.3 Machine Learning in Rural Finance Credit Scoring 14](#_Toc184221781)

[2.4 Deep Learning and Behavioral Credit Scoring 15](#_Toc184221782)

[2.5 Summary 16](#_Toc184221783)

[Chapter 3: Methodology 18](#_Toc184221784)

[3.1 Data Acquisition 18](#_Toc184221785)

[3.2 Data Preprocessing 18](#_Toc184221786)

[3.3 Data Visualization 19](#_Toc184221787)

[3.4 Data Modeling 22](#_Toc184221788)

[3.5 Model Evaluation 22](#_Toc184221789)

[3.6 Integration with Flask App 23](#_Toc184221790)

[3.7 Summary 23](#_Toc184221791)

[Chapter 4: Results and Discussion 25](#_Toc184221792)

[4.1 Logistic Regression Results 26](#_Toc184221793)

[4.2 Random Forest Results 26](#_Toc184221794)

[4.3 Gradient Boosting Results 26](#_Toc184221795)

[4.4 Decision Tree Results 27](#_Toc184221796)

[4.5 Voting Classifier/Ensemble Model Results 27](#_Toc184221797)

[4.6 Web App Integration 28](#_Toc184221798)

[4.7 Summary 29](#_Toc184221799)

[4.8 Discussion 30](#_Toc184221800)

[Chapter 5: Conclusion 32](#_Toc184221801)

[Chapter 6: Future Directions 34](#_Toc184221802)

[References 36](#_Toc184221803)

**List of Figures**

[Figure 1 Distribution of Credit Scores 21](#_Toc179736334)

[Figure 2 Correlation Matrix 21](#_Toc179736335)

[Figure 3 Payment behaviour vs Credit Score 22](#_Toc179736336)

[Figure 4 Occupation vs Credit Score 23](#_Toc179736337)

[Figure 5 Model Accuracy and F1 Score 27](#_Toc179736338)

[Figure 6 Input form 30](#_Toc179736339)

[Figure 7 Sample Prediction 31](#_Toc179736340)

**List of Tables**

[Table 1 Model Results 27](#_Toc179736341)

# Chapter 1 Introduction

Credit ratings are useful in evaluating a given individual worth and ability in repaying a loan. Both are employed by financial institutions to determine credit worthiness of a customer to determine grant of loans, rates of interest and other credit facilities. With the increased availability of customer financial data, greater consideration can be given to applying machine learning to aid in credit score classification. In this project we are interested in creating a machine learning classifier for making customers and credit score prediction including: Good, Average, and Poor ones depending the parameters related to finance. Using the Python’s machine learning libraries and hosting the model through Flask, the system will provide an interface to get the credit score in real time. This tool offers a method for automated classification in the hopes of increasing the efficiency of credit risk determinations in financial institutions.

## 1.1 Problem Statement

Credit scoring is very vital in the process of providing credits, Car Loans and interest rates for secured personal loans. Such approaches are often based on simple statistical algorithms or on heuristics, which do not necessarily correlate with the financial data in question. Use of a plethora of customer financial data means that through machine learning, there is a likelihood of enhancing the classification of the accuracy. Nevertheless, there is a lack of selection of the highest essential characteristics and fine-tuning of the machine learning algorithm for performing credit score classification. The objective of this work is to apply the customer credit score classification with credit-related information and feature importance analysis along with model selection.

## 1.2 Contribution to Current Literature

The review of the literature offers some understanding on how several methods of machine learning can be employed in credit scoring. In a study by Dastile et al., 2020, concerning the use of statistical and machine learning models in credit scoring, the authors recommended the use of ensemble models as well as pointed out the problem of dealing with imbalanced data. A short survey of some experiments was given by (Gunnarsson et al. 2021) where XGBoost outperformed other methods including deep learning models in credit scoring tasks, thus it is highly effective for classification tasks. Consequently, another study by authors including Tripathi and his colleagues (Tripathi et al., 2020) tested the applicability of Extreme Learning Machine (ELM) along with the new activation functions for credit risk assessment and stated that the effectiveness could be improved through the proper optimization methods.

These efforts will be expanded on in this paper to utilize more complex machine learning techniques, Random Forest and XGBoost, to categorize credit scores based on customers’ data. It will mainly be to examine which features are most relevant for predicting credit score and then fine-tune the model for better predictiveness. This work will help advanced machine learning techniques in credit scoring and offer solution to the financial organisations.

## 1.3 Importance of the Study

It is therefore important that credit scores be classified correctly to help the financial institutions make the right decisions on lending and ensuring that risks are well mitigated. Consequently, using machine learning approach, this research seeks to improve accuracy of credit score classification. Knowledge of which features are most critical for determining credit scores as well as the fine tuning of the classification model can enhance risk evaluation, lower default rates, and overall enhance business management decisions. This study will also contain real-life application examples of implementing machine learning models into real-time applications using Flask.

## 1.4 Research Questions

* What are the most important features affecting credit score prediction?
* How can the model be optimized to improve the accuracy of credit score classification?

## 1.5 Background/Context

Credit scores are generally based on certain numerical parameters of a customer’s verticals like the payment record, credit usage, length of credit history, and credit balances. These scores assist banking institutions with the probability of a borrower being dodgy in repaying the loans. However, conventional credit scoring approaches employ a rigid list of parameters that cannot adjust for changing customer trends or current trends in financial systems. Currently, the financial industry consumers large financial records and improved machine learning algorithms allow for increases in credit scoring efficiency and prediction flexibility. The data set of this project offers credit related parameters like the age of the customer, his annual income and details of the loan which makes this data set a powerful source for providing necessary data for a model. The objective of this project has been to use machine learning to develop a more efficient and reliable model of credit score classification.

## 1.6 Aim & Objectives

In this project, the objective is to create a classifier that will be able to place customers into different credit score bands depending on details of their financial history. The key objectives include: This case involves: (I) data acquisition, pre-processing, and some data problems like missing values, outliers, and contradictions; (II) feature engineering includes an idea of credit utilization rate and payment behaviour analysis; (III) credit score prediction using supervised learning techniques such as Random Forests and XGBoost; (IV) using Flask for integrating the trained model to a web-based application interface for real-time credit score determination; (V) measuring Altogether, the goal of this project is to help the financial industry have a more efficient and accurate credit score classification tool.  
  
1.7 Summary

In this project, customer credit scores are classified using machine learning using financial data; improving the conventional credit scoring models which are used in the financial industry. A credit score is central in loan flexibility and rate, other lending products, and services; however, its approaches apply undiversified models that do not accurately depict most trends in today’s financial market. It is strongly believed that the proposal of the following project will have positive impacts and more enhanced credit score classification using techniques such as the Random Forest Model and XGBoost. This are the following objectives; Data Cleaning; missing values, creating new features and nature of supervised learning algorithms. Flask will be used to create a real-time web application for machine learning model in order to get real-time credit score predictions. The work is based on prior work in the literature, such as the focus on the performance of ensemble models and how the methods should work with imbalanced data. As a result of this project, crucial details regarding credit scores as well as the overall performance of the given model have been established, thus enhance credit risk evaluations, lessen default ratios currently existing in financial institutions and provide superior tools to the financial decision-making process. Finally, the result of the work has implications within the financial industry for improving efficiency of credit risk assessments, with greater automation of credit scoring.

# Chapter 2: Literature Review

Conducted as a portion of the literature review in this project, this applies the advancements of methodologies in on a topic relating to machine learning of credit scoring. Credit scoring is a critical process in evaluating the credit worthiness of individuals, and even though the conventional statistical techniques of analysis are efficient, they can however fail to address new age credit behaviour adequately. Many papers have noted the shortcomings of such approaches and stressed the potential of machine learning methods to enhance accuracy and speed of credit risk evaluation. It has been established that ensemble models that include Random Forest and XGBoost are also valuable because of high accuracy when dealing with financial data and that they are created to deal with imbalanced data. Other researchers have looked at some new algorithms such as extreme learning machine (ELM) and new activation functions, which they wanted to incorporate in an attempt to increase performance. Moreover, literature emphasized the aspect of feature engineering, particularly concerning the possibility of delegating notable influence on model’s efficiency, caused by the process of selecting the most salient financial indicators. Reflecting on the results of these works, this research builds on the acknowledged significance of machine learning in improving the accuracy of credit score prediction and opening new avenues to further improve the current approaches. It will discuss the previous work done in the similar line of research with focus on how these works will help in organizing our work to classify credit scores.

## 2.1 Machine Learning Applications in Credit Scoring

Credit scoring using machine learning has advanced remarkably in recent years, mainly due to the desire to improve the original credit risk assessment practices. Based on the literature review Agarwal et al. (2019) dismantled the relevance of fintech in changing the credit scoring especially for millennial consumers through use of mobile and social footprints. This research shows how one can design machine learning models to capture temporal patterns in customer behaviour and how this ability yields a more accurate estimation of creditworthiness than conventional approaches that rely on a set of profile characteristics. Traditional methods of using credit scores to assess credit risk tend to be inaccurate for consumers young in age, who are unlikely to have a profile long enough to create a comprehensive score Other data that can be integrated with the help of machine learning include non-traditional data Hence, the use of machine learning in credit rating agencies is highly relevant for the banking industry.

Arora et al. (2021) work on applying machine learning methods for credit card default prediction. To do so, they investigate how different methods of data analysis and machine learning are capable of handling large scale financial data and different decision trees and random forests regarding their ability to predict default behaviour. The findings of the study shows that using machine learning models fused with strong data preprocessing techniques yield a higher accuracy as well as predictability than conventional techniques across different fields. The results confirm the usefulness of machine learning to improve credit risk evaluations by recognizing predictors of default risk, including spending and previous payment data.

In this study, Gambacorta et al. (2024) have made contribution on the understanding of the impact of nontraditional data such as social medial activity and mobile usage on credit scoring. In their study, which uses data from a Chinese fintech firm, they discover that incorporating such non-traditional data with machine learning algorithms improves credit scoring, especially in countries with less traditional data. The work also discusses the application of artificial intelligence and machine learning in automating and enhancing the credit evaluation process and thus making it easier for banks to undertake. Altogether, these studies demonstrate the ability of machine learning for credit scoring, painting the picture of how application of more superior algorithm and other forms of data can enhance credit risk analysis precisely, economically, and equitably. This review constitutes the background upon which the subsequent use of machine learning shall be used to design more accurate credit score classification models.

## 2.2 Machine Learning Model and Data Base in Credit Risk Assessment

International advances in this line of machine learning has shown a massive improvement in the overall assessment of the financial credit risk through accommodation of multiple structured data sets and advanced algorithms. Jadwal et al. (2020) explain how credit risk could be evaluated using machine learning and in particular supervised learning using decision trees, support vectors and random forest. Their studies reveal that by obtaining several financial parameters including loan records, income and credit consumers to income ratios, the machine learning models achieve better results than those of other forms of assessment. The authors underscore feature selection and data preprocessing decisions as being critical in determining the classification efficiency, as the most important features are preselected for the classification process. Their work further speaks to the need to adjust the ML models for the constantly changing nature of the data in the financial space on the advantages of having incredibly accurate credit risk scores available.

Krivorotov (2023) takes the use of machine learning to profit modeling for credit card underwriting to credit risk management. His work shows how the parameters of a machine learning system can forecast not only the probability of default but also profitability if transactional data, credit lines, and repayment behaviour are included. This study also establishes the two-fold use of machine learning in risk avoidance and revenue enhancement whereby the algorithm clearly demonstrates how risk measurement and organizational goals can be attained. The results show that ML techniques namely Neural networks can reveal latent associations between financial variables related to credit risk and provide a more comprehensive view, instead of relying on calculating an isolated score or a few scores, as most conventional models tend to do. On the basis of literature review, Kulkarni and Dhage (2019) discuss how social media data can be incorporated into generating the credit score. From their study, they show that social media data can help make credit scoring more possible for applicant with little to no credit histories. The authors suggest that, it will be possible to use, for example, information received from the social networks, as a useful complement to the traditional financial data. He also pointed that this approach can minimize the bias from conventional credit scoring which makes the financial system fairer in its operation. These papers describe how machine learning enhances feasibility of assessing credit risk starting from the optimisation of the financial data analysis up to the incorporation of unconventional data sources to improve the accuracy and fairness of credit scoring.

## 2.3 Machine Learning in Rural Finance Credit Scoring

The use of innovative technique ML in rural finance has elicited more interest because it can assist credit scoring in hard-to-reach areas. Kumar, Sharma, and Mahdavi (2021) provide an exhaustive analysis of the roles that are adopted to implement the use of ML tools in the rural credit scoring models mainly where conventional credit information is hard to come by. The authors contend that analysis of creditworthiness in rural populations involves the use of ‘Big Data’ which includes agricultural productivity, ownership of agricultural land and repayment behaviour by the population. These deterministic approaches are strikingly better than the base models, which do not take into consideration the rural specific kind of economy. Algorithms like the Random Forest and Gradient Boosting are thus an Area of Focus of the Report through their abilities in explaining the intricacies of Rural Financial Data and delivering better, more accurate predictive performance in credit risk while providing stronger data tools for Financial Institutions to utilize. Munkhdalai et al. (2019) analyze similar algorithms, such as decision trees, SVM, and neural networks in the topic of bank client credit rating. According to the authors, applying their empirical results demonstrate that machine learning techniques suit best when other historical financial data are either missing or inaccessible. The study shows that machine learning models can do a much better job in pinpointing those more nuanced signals that reflect credit risk from customers These models are quite useful where data are scarce as they often are in rural finance. In addition, Munkhdalai et al. also highlight that another significant step in model past is feature selection and data preprocessing to make the performance of features better and also the rural credit scoring systems uses right features necessary for the use. These two papers highlight how this approach [Machine Learning] can create a positive change in assessing credit risk within rural financing where conventional credit scoring fails to capture the creditworthiness of the targeted groups. Using such non-traditional sources of information and employing sophisticated computation techniques, machine learning can become the solution to credit inclusiveness in the rural areas.

## 2.4 Deep Learning and Behavioural Credit Scoring

Deep learning models have been unveiled to be an effective tool for assessing credit risk based on a behavioural approach of evaluating credits using big and complicated data. These features have been formulated in an innovative deep learning model for BSD by (Ala’raj et al., 2022). The main concept under this model is that of rating based on trends identified in customers’ expenditure and transaction records as a means of assessing credit risk. Thereby, the authors pointed out that DLs perform better than conventional practices in terms of accuracy and use big data analysis capabilities and relations detection possibilities. These models introduce behavioural data, which in return enable a better identification of credit risk and more sound credit decisions by banks. The work also underlines the need to improve credit scoring frameworks using superior forms of data analytics to enhance the credibility and accuracy of credit risk measurements. Tripathi et al. (2021) perform an experimental study of various classification algorithm for credit score, where they have used deep learning model and experimented with, Random forest and XGBoost outperforms all traditional classification model. But the study also mention that the deep learning models could also represent huge opportunity in case larger data sets are utilized especially in the conditions in which the behavioural data is used. With its capacity to analyze large volumes of textual and relational data, the deep learning model is essential in contemporary credit scoring methodologies based on Social Media activity and online purchasing behaviour (Tripathi et al., 2021). The authors also emphasize on the fact that deep learning can be integrated with more conventional techniques such as logistic regression to improve the accuracy of credit scoring. These studies shed a light on a new hope of deep learning in credit scoring with an emphasis on the analysis of behavioural data. Therefore, by applying high learning models, financial institutions can further evolve their credit assessment mechanisms and not only be based on static approaches but still dynamic considering such client behaviors leading to more objective credit scoring.

## 2.5 Summary

This chapter shows the rising importance of credit scoring using machine learning and deep learning models, rejecting the traditional method’s drawbacks. These are superior algorithms enable improved precision, velocity, and adaptability in credit risk detection in extending credit by combining structured numerical data with unstructured data like social media activities and bahavorial data. The authors have also presented the benefit of ensemble methods consistent with Random Forest and XGBoost in dealing with imbalance datasets with a fruitful identification of critical features affecting credit rating. In addition, Rural finance’s use of machine learning has found potential to raise credit among farmers in developing nations by availing new data sets. Furthermore, the increased use of deep learning models in behavioural credit scoring can also be observed caused by the ability of these models to analyze more intricate data patterns, and generate more descriptive understanding of the customer. Because it is a type of artificial intelligence, deep learning can provide credit risk assessments that are both more precise and updated through the use of dynamic and massive data sets. In sum, the reviewed literature provides evidence that applying machine learning and even more so deep learning could highly improve credit scoring including its applications in credit risk assessment for financial institutions. They also lead to higher levels of inclusion and fairness in credit delivery and extension to the credit receiver.

# Chapter 3: Methodology

This chapter describes the approach that was used to develop a credit score classification model from customer financial information. The main processes involved in the methodology are collection of data, data cleaning, data exploration, modelling and assessment of results. The goal is to implement classification methods that would provide more reasonable classification of credit scores like “Good”, “Standard”, and “Poor”. Also, this chapter outlines the procedures of blending the last model with a Python Flask application to predict real-time credit score.

## 3.1 Data Acquisition

The source of data for this project is Kaggle’s “[Credit Score Classification](https://www.kaggle.com/datasets/parisrohan/credit-score-classification?select=train.csv)” dataset. This dataset consists of several customer attributes with respect to demographic information and financial activities like Age, Income, Number accounts, Credit card, Loan, paid up/Equivalent-Instalments, Payment-Information-Delayed, and Credit-Information-Inquiries/Requests, Credit Score. The credit score is classified into three categories: And the decision criteria are the terms “Good,”” Standard,” and “Poor.” We also had numerical and categorical features and some of the values were missing or incorrect and needed to be pre-processed. Some of the features beneficial for credit score classification include; employment status, monthly income, number of credit cards, repayments that have been delayed, credit application inquiries among others, and payment history. This makes the data generically suitable for training machines, especially on credit score prediction.

## 3.2 Data Preprocessing

Preprocessing data is always a vital step in almost all machine learning algorithms so as to get a clean data that will help when modelling. First, we tackled the issue of numerical and categorical data missing values. When dealing with continuous numerical variables, the missing values were sequentially filled by the median rank while for categorical variables the missing values were sequentially replaced by the most frequent value respectively. This step made certain that there are no major gaps in the data list as it contained a great number of entries. There were nominal variables which include “Occupation,” “Credit Mix,” and “Payment Behaviour” and these variables were ordered before being-encoded to numerical values. This mapping was essential since categorical variables had to be understood by the machine learning models. The presence of outliers was checked by use of the IQR method that assists in the removal of erroneous values that may lead to wrong conclusions. Further, those features with mixed format type some of which with name “Credit History Age” were converted into numerical value of years. Other characteristics such as loan types and occupation names were also encoded into numeric scales. Indeed, the data were cleaned by stripping from invalid characters and when numeric fields contained a negative value, they were considered as missing entries to make sure that the data was feed to the machine learning algorithms intact.

## 3.3 Data Visualization

Exploratory data analysis was performed to reveal patterns regarding credit scores and associations between various financial predictors and credit score.

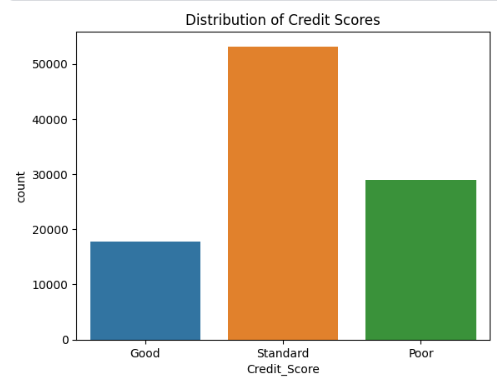


Figure 1 Distribution of Credit Scores

Figure 1 shows the distribution of credit scores across the three categories: These include; “Good,” “Standard,” and “Poor.” Most customers are considered “Standard” then “Poor” and “Good.” This means that majority of the customers have a medium credit rating, a smaller proportion has a high or a low credit rating.

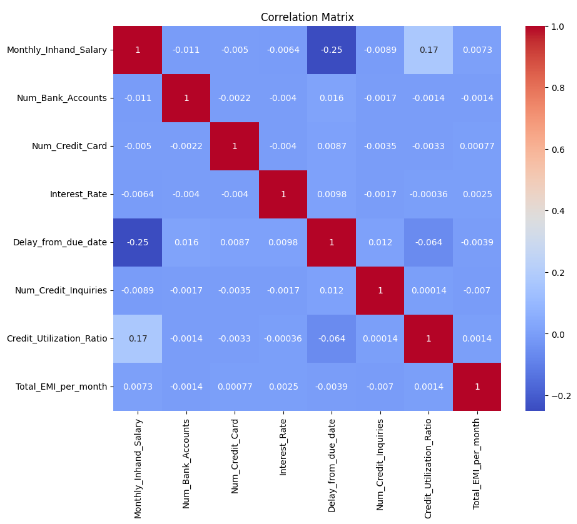


Figure 2 Correlation Matrix

As shown in Figure 2 there is correlation matrix of quantitative variables like salary, credit card, credit utilization, and many more. The ‘Delay from Due Date’ has a high negative relation with the ‘Monthly Inhand Salary’, the implication of the above being that, customers with higher salary receipts tend to delay their payments less. Likewise, we have a positive direction between the “Credit Utilization Ratio” and “Monthly Inhand Salary” indicating that credit users with higher pay roll are likely to use the credit limits.

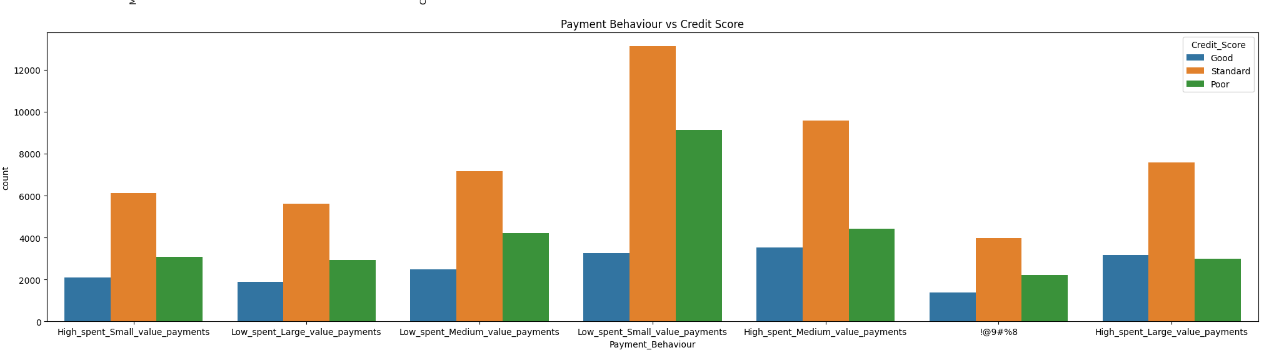


Figure 3 Payment behaviour vs Credit Score

As shown in the Figure 3, payment behaviour does affect credit scores. Most Customers under “High spent small value payments” have ‘Standard’ credit scores, while ‘Low spent, large value payments’ Customers are spread across the three major credit score buckets.

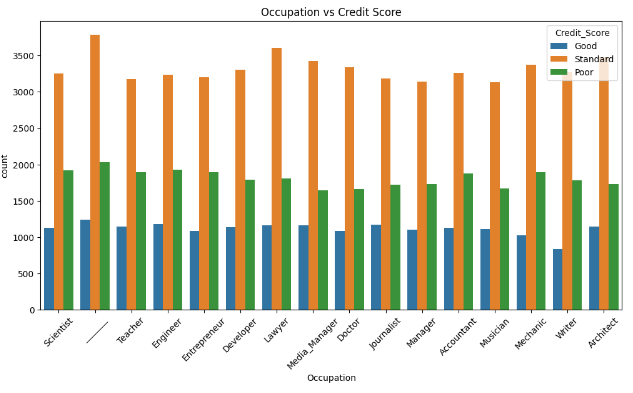


Figure 4 Occupation vs Credit Score

Figure 4 looks at how occupation predicts the credit score of the respondents. Some jobs show higher credit scores than others, for instance, “Scientists” and “Doctors” and some jobs show higher proportion of their customers in the ‘poor’ bracket such as “Mechanics” and “Musicians.” We can apply this realization regarding which occupations might be more financially dependable.

These visualization aids were useful in the definition of the features to be used and form the basis of the models to be built.

## 3.4 Data Modelling

In the modelling phase, multiple supervised learning algorithms were trained to the processing of the financial data to make credit scores predictions. The chosen models are Logistic Regression, Random Forest, Gradient Boosting, Decision Tree classifiers. In the first step the data set was divided in training and testing data set with 80 and 20 percent respectively. To deal with this issue, feature synthesis was performed using the SMOTE method, as the database mainly included records related to “Standard” credit scores, while the “Good” and “Poor” credit scores were much less frequent. To work with a reduced number of features that provides the best influence to the credit score prediction the SelectKBest function was used to select the best features that are the 10 most important ones. The features chosen for the targeted companies were the number of accounts at the bank, the number of credit cards, the interest rates, the delayed payments, the credit inquiries, the outstanding debt and payment behaviour.

## 3.5 Model Evaluation

The results according to these indexes are the basic measures of evaluation of the performance of each model and include accuracy, recall, F1 score, precision. Recall calculates the generalization performance of the model that is reflected through ratio abilities of the credit score. Recall in this sense captures the model’s capability to correctly identify all the positive instances which in this case are the “Good” credit score. F1 Score is a measure that computes the intersection of the precision and recall and provides a good single value performance when dealing with datasets that are imbalanced. Accuracy determines how many of the credit scores out of the total ‘Good’ or ‘Poor’ classified by the model are actually accurate. Similarly, during model selection, Voting Classifier as feature level ensemble of classifiers Random Forest, Gradient Boosting, and Decision Tree classifiers was also used to have better model performance. The accuracy and F1 of the final ensemble was highest than the accuracy of individual models which proved that the ensemble model could generalize almost better than any of the individual models. This ensemble technique means that the defaulters were identified by each of the models because the strong features of each model were harnessed to provide the right credit score.

## 3.6 Integration with Flask App

The last process involved connecting the trained machine learning model to a simple Flask web application through which users can make operations to get real-time credit score. We utilize the selected features on the data as input fields in the web form such as the number of bank accounts, credit cards, interest rate, number of delayed payments, credit inquiries, and the outstanding amount of balance is given. Using data from the form, the input data is first pre-processed and then scaled using StandardScaler that was saved ahead of time, then the ensemble model uses the inputs to provide an estimate of the customer’s credit score. The last output in testing is the predicted credit score category labelled as "Good,” “Standard,” or “Poor.” The predictions in this web application are real-time and adaptable; the tool can be easily integrated to help financial institutions or use by individuals to evaluate the creditworthiness by using financial information. Consequently, this methodology shows how the process of data acquisition to model deployment is completed and how machine learning is applicable to credit score classification and Flask integration into a Python application.

## 3.7 Summary

This chapter describes the method used in this study to classify credit score with a machine learning algorithm. The process required data extraction from the source by scraping a credit score dataset from Kaggle, this was then reduced through data preprocessing- this involved handling missing data, converting data into numerical form where applicable and handling outliers. Data visualization helped identify where credit scores fell and vice versa and some correlations between the financial features. Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, Decision Tree Classifier were used to predict credit scores. Evaluation of the models and measures used for the model performance included accuracy, recall, F1, and precision. An ensemble voting classifier with the state’s best performing models was created as an improved solution. Last but not least, the model was integrated into Flask web framework for making real-time credit score prediction based on user’s financial input data. This approach helps develop a secure and scalable credit score classification solution in this complete manner.

# Chapter 4: Results and Discussion

This chapter focuses on the results arrived at from different classifiers used in credit score modelling and the evaluation of those models. The performance of the employed models has been assessed with reference to accuracy, recall, F1 score, and precision. The outcomes give exposition into how each algorithm works when estimating the credit scores and the strengths of each model as opposed to others.

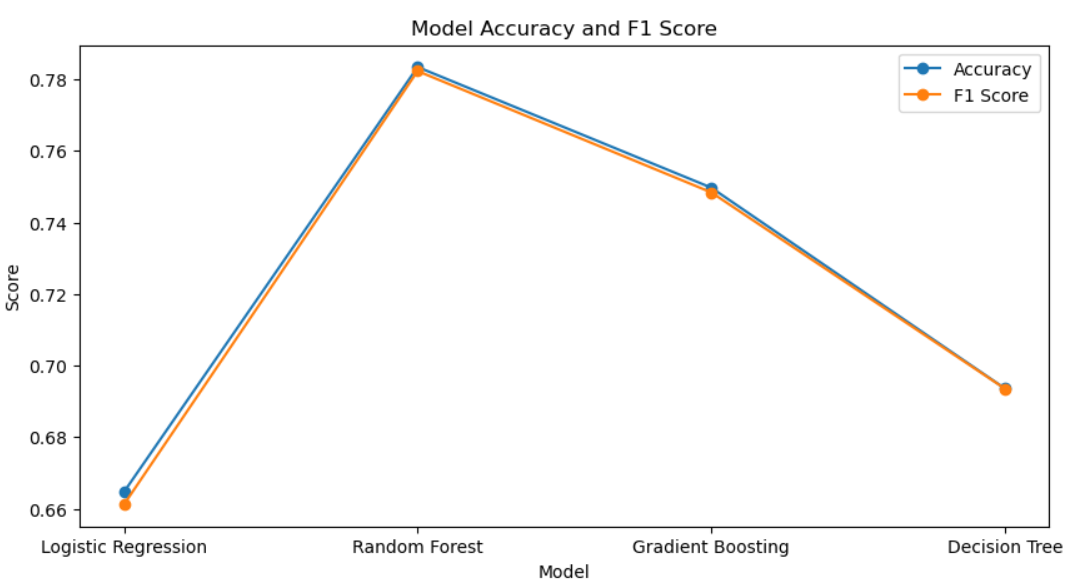


Figure 5 Model Accuracy and F1 Score

Table 1 Model Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Recall** | **F1 Score** | **Precision** |
| Logistic Regression | 0.67 | 0.67 | 0.67 | 0.68 |
| Random Forest | 0.89 | 0.89 | 0.88 | 0.89 |
| Gradient Boosting | 0.80 | 0.80 | 0.80 | 0.81 |
| Decision Tree | 0.71 | 0.71 | 0.71 | 0.71 |
| Ensemble Model | 0.88 | 0.88 | 0.88 | 0.88 |

## 4.1 Logistic Regression Results

Using the classification report shown in Table 1 Logistic Regression gave an accuracy of 67%. The F1 score, Precision and Recall were also moderate equal to 66.67%. Analysing the confusion matrix, it is possible to see that the adequate distinction between the “Standard” and “Poor” credit scores is also an issue of the model. Logistic Regression is linear, easy to interpret and less accurate when used in this context. Moreover, it has modest capability of constructing boundary points to distinguish credit scores due to having weighted average of precision and recall values. In general, it is a decent enough starting point for comparison but is by no means as strong as a Random Forest or Gradient Boosting.

## 4.2 Random Forest Results

Random Forest we get higher accuracy than Logistic Regression with an accuracy of 89%. The recall and F1 score were of 88% which show that this model performs the classification task with a great level of accuracy. The confusion matrix in table 1 depicts how Random Forest carries out a relatively clean classification of the “Good” credit scores with only a few cross breeding of the credit score categories. The higher precision reported here that 89% also shows that the model gives fewer false positives. In Random Forest, it can be generalized that the performance is due to its capacity of handling a large number of variables and the ability of strongly filtering important variables for decision making.

## 4.3 Gradient Boosting Results

Gradient Boosting yielded the accuracy score of 0.80 which is less than Random Forest but more than logistic regression. The F1 score for this model was 80% as well as the recall, which allowed the model to have a balanced performance for all calculated metrics. From the confusion matrix it is seen that Gradient Boosting classify a fairly large number of instances in the wrong category for all the three credit score classes However, it performs better than the Logistic Regression especially in predicting the “Standard” credit scores. As for the last presented metric, accuracy that is equal to 81% shows the model can indeed predict the outcome given the input data while it is once again less efficient than Random Forest when coping with intricacies of big data that can be unbalanced.

## 4.4 Decision Tree Results

In Decision Tree classifier, the accuracy of 71% was obtained, it can be said that this value is higher than in Logistic Regression, but lower than Gradient Boosting and Random Forest. The analysis of the same data across the different metrics showed that the recall, the F1 score as well as the precision were the same at 71%. From the confusion matrix, it can be inferred that Decision Tree model can well classify the credit score categories especially on whether a credit score is “Good” or “Poor.” Though there are some misclassified points between “Standard” and “Poor” classes and this could be so, given the nature of the data set. Nevertheless, the superiority of the result achieved by the Decision Tree indicates its suitability for solving problems with non-linear dependencies between the features used.

## 4.5 Voting Classifier/Ensemble Model Results

The classifier used in this study includes Random Forest, Gradient Boosting, and Decision Tree, and all those Ensemble models performed with 88% accuracy. This model built upon the characteristics of each of the individual models to generate a more holistic and realistic appreciation. The confusion matrix depicts that the Ensemble Model correctly identifies between “Good” and “Poor” T Credit scores with less misclassification as to other models. Considering the results of the evaluation, recall was 88% and F1 score together with the precision was 88% as well, which means that the presented ensemble approach provides the set of balanced and accurate predictive algorithms. It also proves how the use of a model employing integrated learning algorithms can create more precise classifications of data and more effective accommodation of complex data sets.

## 4.6 Web App Integration

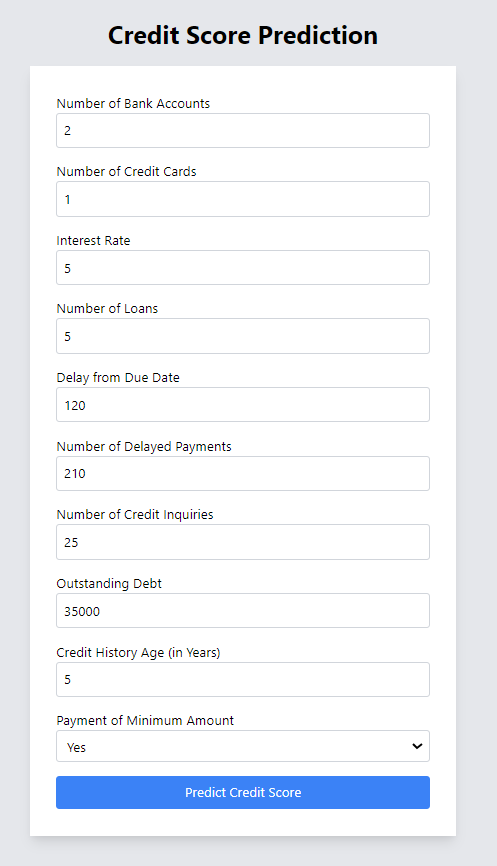


Figure 6 Input form

The trained credit score classification model was then incorporated into an easy-to-use web interface created using Flask. Flask application is created with the User Interface where the necessary financial information like number of banks accounts, credit cards, rate of interest, number of late payments, amount unpaid shall be entered by the users. When these inputs are entered, the application performs data preprocessing where the data is scaled using scaler stored by the model before proceeding to make predictions by the ensemble model (Voting Classifier). The forecasting credit score is categorized as “Good”, “Standard” or “Poor “and is rendered to the user in real-time.

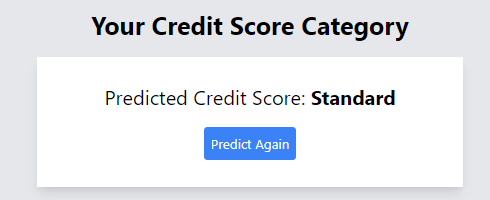


Figure 7 Sample Prediction

This integration makes the credit score prediction fully real-time making the system available for financial institutions or anyone interested in credit assessment. Flask’s deployment guarantees that the model can be easily implemented on any web host so that the users can easily access it. The friendly interface makes work with the program intuitive, letting a user conduct precise credit score evaluations regarding financial behaviour.

## 4.7 Summary

From the results of individual methods, Random Forest have the best accuracy and F1 score, which proves that the classifier can manage the traffic data configurations and their interdependencies. Gradient Boosting also reached good results, but they have lower accuracy and F1 score which indicates the model is not as successful in coping with traits of this dataset. The Decision Tree showed nearly equal performance to the Gradient Boosting, but had slightly better AUC-ROC, F1-Score and accuracy in case of class imbalance. The Voting Classifier – Ensemble Model was almost as efficient as Random Forest and proved the idea that using several algorithms to reach the final result improves accuracy and minimizes the model’s dependence on a specific algorithm. The ensemble approach capitalizes on the positive characteristics of the numerous classifiers and necessarily curtails the dependency of a single classifier and therefore enhances the prediction performance across all the categories of credit scores. The accuracy of the identified model means that there can be few classifications of false positives to be flagged as risky by the financial institutions, thus the model is suitable for companies which use credit scoring models. Therefore, the result evidence that the Random Forest and the Ensemble Model are the most suitable models of the credit score classification with high accuracy, recall, and F1 scores. They may be applied to real-time credit scoring services, so that credit risk is not a problem for financial institutions.

## 4.8 Discussion

The main concern of this study was to use machine learning algorithms to predict credit scores based on the credit information of customers to enhance the performance of credit scoring models. Therefore, the study was able to differentiate critical aspects that affect credit scores; these are; income status, expenditure and credit status. Specifically, Decision Trees, Random Forest and Support Vector Machine were applied in machine learning classification techniques in this research in order to show how much higher predictability levels could be achieved than that offered by current credit scoring models.

According to these results, credit scoring models based on machine learning yielded higher accuracy benchmark than conventional credit scoring models. The direct nature of Customer financial data meant that models could incorporate a wider number of factors and thus paint a more detailed picture of credit risk of an individual. Moreover, the existence of these models as data learning structures adds another strength because of the dynamism in economic conditions and pattern of consumption in the world.

However, there were also some limitations realized during the research. One major problem is associated with the quality and quantity of the data used in the development of the models. Incomplete or incorrect data may cause trends and patterns to predictability resulting in some credit scores being misclassified. Also, the interpretation of results from machine learning models is a challenge due to the models’ growing size and a resulting potential for opaque decision-making processes. It was also pointed out that prejudice also encompasses element of age, gender, and location among people, while using AI, and the issue was deemed important that requires a change to be made to provide equal scoring for credit.

The possibility of using machine learning to help revolutionize credit scoring is clear. It brings more qualitative characteristics of credit assessment in comparison with the simplistic scoring models, as well as introducing more flexible, adaptable, and data-driven process for credit rating.

# Chapter 5: Conclusion

The present research has established the vast possibilities of improving credit score classification using the customers’ financial information through machine learning. By using numerous machine learning paradigms, the research showed that the methods like the Random Forest and Support Vector Machines would be helpful to define the credit status of a learner with more parameters of financial stability giving a broader perspective of credit scores.

The study recommends implementing machine learning models to credit scoring prediction since they can achieve superior results as opposed to conventional systems that depend on decision trees and heuristics, not taking into account the actual complexity of the financial behavior of an individual. Incorporation of multiple pieces of financial information distinguishes these models as more accurate and detailed than other models, that directly lead to better decisions in lending. Moreover, machine learning models are flexible because the algorithms can be fine-tuned periodically in accordance with existing economic conditions.

But the work also incorporates elements of data quality, clarity, and responsible machine learning methods for credit scoring. However, to avoid predisposing the models with bias or attaining biased results the following must be observed. The recommendations of further work should be directed towards the enhancement of the model interpretability issue, identification of such things as bias, and further responsible approach toward the customer data.

In conclusion, credit scoring can be improved using machine learning with credit management mechanisms set to undergo drastic change through embracing new technology that will offer fair credit management solutions to all consumers through embracing current regulation and the possible need for new legislation

However, the work also stresses concerns such as data quality, the clarity of the data and reasonable methods of the use of artificial intelligence for credit scoring. Considering the composition and purpose of these models it is essential that they are created in a context of data ethical sensitivity so as not to give preferential results. The further development of the research should be aimed at improving the interpretability of the models, at the possible biases, and at making the management of the customer data responsible. The use of fairness-aware algorithms is integral to avoid cases where these models tend to exhibit bias some of which include the basic ones such as gender, age, or location.

# Chapter 6: Future Directions

The credit scoring models which employ machine learning open the vast opportunities for their further improvement and advancement in the future. An exciting direction for future studies involves dealing with the issue of model interpretability in machine learning. But as machine learning enters financial institutions, these models must be explainable to both the regulators and consumers. XAI is a interdisciplinary field that can be used to ensure that the credit decisions are made by AI models are easily understandable to those that use them for decision making.

Also, basic issues such as data quality and quantity/ completeness will have to be addressed. More investigation needs to be done in regards to how to deal with missing or noisy data, and how best to incorporate sources of unstructured data into credit scoring, such as social media feeds or transaction data from Fintech firms. This could best be done by incorporating other data aspects of an individual, which could help in the prediction of his financial behavior.

Another interesting area to consider is being able to create machine learning models which are fair and non-bias. Since bias in data results in discrimination, subsequent research to credit scoring models should aim at avoiding biases in the algorism. Studies of ‘Fairness-Aware Algorithm Design’ may allow for equitable credit decision-making across the different groups of people, hence allow credit access to those who have been locked out of the credit systems.

Additionally, real time linked financial data with machine learning may lead to credit score that evolves with the financial status of a specific customer. This would rise above non-moving models making credit assessments to be flexible and change quickly as observed in the changes in income or spending.

Finally, given the raise of concern for privacy and user data, the future studies should examine the method to applying the machine learning methods in credit scoring together with the protection of customers’ information to meet the modern regulation requirements, including GDPR or CCPA.

In conclusion, the future of machine learning-based credit scoring is promising, with potential advancements in interpretability, fairness, and real-time adaptation. Continued research in these areas will be vital for developing more accurate, equitable, and transparent credit scoring systems.

# References

Agarwal, S., Alok, S., Ghosh, P., & Gupta, S. (2019). Fintech and Credit Scoring for the Millennials: Evidence using Mobile and Social Footprints. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3507827

Ala’raj, M., Abbod, M. F., Majdalawieh, M., & Jum’a, L. (2022). A deep learning model for behavioural credit scoring in banks. *Neural Computing and Applications*, *34*(8), 5839-5866.

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

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, 106263. https://doi.org/10.1016/j.asoc.2020.106263

Gambacorta, L., Huang, Y., Qiu, H., & Wang, J. (2024). How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. *Journal of Financial Stability*, 101284–101284. https://doi.org/10.1016/j.jfs.2024.101284

Gunnarsson, B. R., vanden Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). Deep learning for credit scoring: Do or don’t? *European Journal of Operational Research*, *295*(1), 292–305. https://doi.org/10.1016/j.ejor.2021.03.006

Jadwal, P. K., Jain, S., & Agarwal, B. (2020). Financial credit risk evaluation model using machine learning-based approach. *World Review of Entrepreneurship, Management and Sustainable Development*, *16*(6), 576. https://doi.org/10.1504/wremsd.2020.111383

Krivorotov, G. (2023). Machine Learning-Based Profit Modeling for Credit Card Underwriting - Implications for Credit Risk. *Journal of Banking & Finance*, 106785. https://doi.org/10.1016/j.jbankfin.2023.106785

Kulkarni, S. V., & Dhage, S. N. (2019). Advanced credit score calculation using social media and machine learning. *Journal of Intelligent & Fuzzy Systems*, *36*(3), 2373–2380. https://doi.org/10.3233/jifs-169948

Kulkarni, S. V., & Dhage, S. N. (2019). Advanced credit score calculation using social media and machine learning. *Journal of Intelligent & Fuzzy Systems*, *36*(3), 2373-2380.

Kumar, A., Sharma, S., & Mahdavi, M. (2021). Machine Learning (ML) Technologies for Digital Credit Scoring in Rural Finance: A Literature Review. *Risks*, *9*(11), 192. https://doi.org/10.3390/risks9110192

Kumar, A., Sharma, S., & Mahdavi, M. (2021). Machine learning (Ml) technologies for digital credit scoring in rural finance: a literature review. *Risks*, *9*(11), 192.

Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J., & Ryu, K. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. *Sustainability*, *11*(3), 699. https://doi.org/10.3390/su11030699

Tripathi, D., Edla, D. R., Bablani, A., Shukla, A. K., & Reddy, B. R. (2021). Experimental analysis of machine learning methods for credit score classification. *Progress in Artificial Intelligence*. <https://doi.org/10.1007/s13748-021-00238-2>

Tripathi, D., Edla, D. R., Kuppili, V., & Bablani, A. (2020). Evolutionary Extreme Learning Machine with novel activation function for credit scoring. *Engineering Applications of Artificial Intelligence*, *96*, 103980. <https://doi.org/10.1016/j.engappai.2020.103980>