Journal of Statistics Applications & Probability
*An International Journal*

http://dx.doi.org/10.18576/jsap/12S105

# Understanding and Forecasting of Credit Defaulters Using R-Programming

*V. Bhatt*[1], *P. Pathak*[2], *S. Rastogi*[3], *Venkata M. Bhimavarapu*[4], *P. Tapas*[1], *S. Kadam*[1], *N. Parashar*[5], *P. Barge*[1], *and B. Agarwal*[3,*]

[1]Symbiosis Institute of Business Management, Symbiosis International (Deemed) University, Pune, India
[2]Symbiosis Institute of Digital and Telecom Management, Symbiosis International (Deemed) University, Pune, India
[3]Symbiosis Institute of Business Management, Symbiosis International (Deemed) University, Nagpur, India
[4]Symbiosis Institute of Business Management, Symbiosis International (Deemed) University, Bengaluru, India
[5]Symbiosis School of Banking & Finance, Symbiosis International (Deemed) University, Pune, India

Received: 7 Mar. 2023, Revised: 10 Oct. 2023, Accepted: 21 Nov. 2023
Published online: 1 Dec. 2023.

**Abstract:** In Technology has provided numerous significant benefits to the financial industry. Financial transactions are now much smoother and faster than they were previously. Creditworthiness is a measure of how likely you are to repay your debt obligations, and it supports lenders decide whether or not to extend new credit to you. The current paper attempts to comprehend credit defaulters and develops a model to aid in understanding the determinants and prediction. A dataset of 376 responses was divided into training and testing data sets in proportions of 70% and 30%, respectively. The authors used traditional Binary Logistic Regression, Deep Learning, and Random Forest to achieve the empirical results. Logistic regression, an extension of linear regression with a categorical dependent variable, will also be used for comparison. IBM SPSS was used to run the binary logistic regression. and creates a model to aid in understanding the determinants and prediction.

**Keywords:** Credit; Defaulters; Binary Logistic Regression; Deep learning; R-programming

## 1 Introduction

Advancements in the financial service industry are an ongoing process that includes financial technology to facilitate new ways of doing business. Financial institutions are now focusing on offerings for lending that are closer to the market. They are also exercising to reduce their loan application turnaround by adopting a transparent and non-discriminating approach to credit scoring. The whole exercise is to improve the methods for deciding the indebtedness of the clients (Kokate & Chetty, 2021). Various researchers have mentioned the importance of machine-learning tactics in developing credit-scoring models (Neto et al., 2017). In past research studies, various models have been used aiming to discriminate between good and bad borrowers.

The creditworthiness of a customer can be represented by a numeric value called a credit score. Credit scoring models can be developed by considering two aspects, i.e., behavioral scoring and application scoring. The application scoring approach can help to foresee default risk at the time of application scrutiny. At the same time, behavioral scoring utilizes accounting transactions and financial information of the existing customers. If a client is categorized as risky, financial institutions can adopt preventive actions to protect themselves against any future loss (Óskarsdóttir & Bravo, 2021). The credit scoring model is based on statistical methods to exhibit the probability of default or delinquency and is widely used for making decisions on whether to grant credit or not. The objective of developing a credit scoring model is to present all the borrower's information into a score that is further compared with the predetermined threshold credit value and decisions made accordingly. A credit score is also beneficial for monitoring the creditworthiness of existing borrowers.

In the past few years, credit scoring methods have evolved from traditional methods which are based on statistical tools to more innovative approaches such as machine learning and artificial intelligence. In machine learning algorithms most popular methods are logistic regression, gradient boosting, and deep neural networks. Machine learning methods improve the discriminative power of the credit scoring model, which helps in identifying various risk drivers. Machine learning models are also considered to provide perfectness in the feature selection process, along with handling data cleaning and data quality checks.

Demaris (1995), logistic regression is considered a special case of linear regression and has become an important analytic

*Corresponding author e-mail: agarwal.bhakti96@gmail.com

technique for multivariate dependent variables. Logistic regression is popular due to its simple probabilistic formula to classify the instances; however, it is unable to solve non-linear problems. Khalilia et al. (2011), random forest is another classifier based on ensemble learning, i.e., generates multiple classifiers and aggregates the results. A random forest classifier can handle high dimensional data and promise for improved results. Deep learning models exhibit dramatic improvement while classifying the large datasets (LeCun et al., 2015). The architecture of deep learning methods is a multilayer stack for non-linear computing input and output mappings. Many classification learning algorithms are used to predict and select the best customer for approving the Credit scoring, and this will result in decreasing the defaulters of the banking system. Specially defaults in credit cards have been increasing drastically in recent years. The customers have become defaulters willingly. This trend negatively affects the banking sector. Managers and providers of credit cards in the banking and finance sector must have the ability to identify credit card defaulters easily. But various Banks use various kinds of credit scoring models and risk analyzing models using both statistical and machine learning approaches. Further, it is important to study whether the different algorithms behave differently.

## 2. Literature Review

Behavioural scoring is popular to use by lenders to assess the likelihood of defaulters for a specific period. However, behavioural scoring has not gained the attention of researchers as in the case of application scoring. Kennedy et al. (2013), focus their research on behavioural scoring on evaluating the impact of changing the performance period and outcome period. The study also quantifies the performance differences of logistic regression due to altering different outcome periods.

While making credit-granting decisions, credit scoring models can help financial institutions identify defaulters and non-defaulters (de Moraes & Costa, 2022). In recent years, some advanced techniques, such as machine learning and artificial intelligence, are proven to be well-performing techniques in credit scoring. Goh & Lee (2019), discussed two techniques, viz., support vector machines and metaheuristic approaches, which can be utilized for credit scoring models. The authors also mentioned the lack of availability of proper datasets in the public domain for behavioural credit scoring due to confidentiality issues. The hybrid model can be utilized to develop a credit scoring model to combine the advantages of various single classifiers (Zhu et al., 2018). Deep learning techniques can be further explored to reshape huge data and optimize the structure.

Credit scoring is important in measuring the risk while making a decision for applicants. Machine learning models are utilized for the classification of credit risk. Tripathi et al. (2020), proposed an evolutionary approach to obtain optimized weights and biases to improve the performance of extreme learning machine, which is inspired by the artificial neural network, for credit risk evaluation. Authors claim an improved performance as compared to the traditional approach and existing evolutionary approach. Yuping et al. (2020), came up with the customer segmentation through personal credit scores and also suggested the aspect of the evaluation system can be improved. Authors advocated logistic regression and neural network model to utilize for creating a scorecard model. Dastile et al. (2020), present a systematic survey of different statistical and machine-learning models used for credit scoring. It is also mentioned that deep learning models have not been extensively used for credit scoring yet. Another conclusion drawn from the survey is that ensemble of classifiers can show more promising results than a single classifier (Nikitin et al., 2018). They also emphasized balancing classes of datasets in credit scoring for future research direction. Gunnarsson et al. (2021), analyzed the relevance of using deep learning methods. The study provides a comparison of different methods used for the credit scoring model. The conclusion claimed by the authors is that the ensemble method can give the best performance. Deep learning can be a better approach but with a computational cost. Sometimes multilayer perceptron and deep belief networks can perform worse than two ensemble methods.

An interpretable credit scoring method is proposed that is claimed to resolve the issue of lack of interpretability of ensemble methods (Dumitrescu et al., 2020). The anticipated method is based on the concept of penalized logistic regression for improving the performance of logistic regression by using the information from the decision tree. Goel & Rastogi (2021), mentioned the impact of psychological factors while constructing a behavioral credit scoring model. The study claims to reveal six major factors to predict the indebtedness such as materialism, integrity, personality, self-control, and locus of control. The scope of the proposed model was to explore the relationship between identified factors as well as to identify some more traits. The use of machine-learning and AI with alternative data sources, which can deal with some issues like information asymmetry, moral hazard, etc., has been advocated in the literature (Mhlanga, 2021). Alternative data sources strongly impact credit risk assessments and verify the repaying capability of the clients. However, the paper does not describe the credibility issue of alternative data sources (Simumba et al., 2022).

In literature, a five-step credit scoring model is proposed, which uses real data provided by a bank (Sum et al., 2022). The authors use various data mining techniques such as support vector machine, multilayer perception, and logistic regression, for credit score modelling (Hsieh, 2004). The model is tested on real data of personal loan customers. However, all data provided by banks may not be useful, and some of the unimportant variables were omitted during the study. The authors presented a comparative study of credit scoring models based on machine learning (Bücker et al., 2022). The study finds improved performance in comparison with advanced techniques such as gradient boosting and support vector machines. However, the comparison was drawn to run different models of different complexity for a specific situation. Remolina (2022), also discussed the role of financial regulators along with challenges while using credit score models in the decision-making process of loan applications.

Due to aggressive loan granting, there is a rapid growth in consumer credit, which needs careful assessment while granting credit to potential customers. Some inefficient policies and judgemental approaches are still in use to evaluate credit scoring. According to authors, Ala'raj et al. (2018), developed economies are using automated credit scoring models with the use of well-known classification methods such as logistic regression, support vector machine, and artificial neural networks. However, there is no model that can be considered optimal since the selection of classification methods depends on the nature of the problem and the availability of a suitable dataset (Tripathi et al., 2022).

Onay & Öztürk (2018), suggested that lending institutions are in the transformation stage. Various credit scoring methods are evolving, and the major thing facing transformation is the data sources, led by the big data. However, data-centric approaches for credit scoring and their regulatory aspects are addressed by very few researchers only. They also mentioned that social media activities have the potential to assess and identify the behavioural pattern of the borrowers. However, there is still a dilemma for the exclusion and inclusion of non-traditional data sources with respect to the opacity of algorithms. Trivedi (2020), stated feature selection is equally as important as selecting a machine learning algorithm while performing a credit scoring assessment. The author advocated including and combining expert knowledge with credit scoring algorithms (Lappas & Yannacopoulos, 2021). The importance of expert opinion is considered very helpful in processing credit applications. The study proposed a combination strategy integrating expert knowledge with soft computing methods. Expert opinion is claimed in the research to have the power to strengthen the predictive power and interpretation of the features of the credit dataset. Genetic methods can be embedded with machine learning methods for feature optimization and classification. However, in the case of big data, it proved to be tedious to take expert opinions for each and every feature, evaluating borrowers' credit qualities (Pang et al., 2021) based on many aspects. These aspects may be basic information, credit consciousness, and credit performance. However, the credit qualities of the borrowers may change over time, which cannot be reflected instantly in the data sources.

Due to the increase in traditional services and social lending platforms for credit scoring has disrupted. However, there is a risk if financial institutes depend completely on these platforms for credit risk assessment. Moscato et al. (2021), discussed setting a benchmark of machine learning methods for credit scoring, which may solve the issue of class imbalance based on different sampling techniques. However, deep learning and ensemble approaches could show better performance while managing unbalanced datasets and treating class imbalance issues. In recent years there has been a rapid increase in using machine learning approaches for credit scoring in the financial sector. However, another issue arises, i.e., the complexity of these methods, and it is very difficult to explain and interpret the impact of their prediction. Bueff et al. (2022), proposed a counterfactual that can help interpret and understand the model in regard to the decision boundaries along with the impact of prediction.

## 3. Methodology

This research paper attempts to understand credit defaulters, i.e., both the wilful defaulters and the credit defaulters and prepares a model that can help better understand the determinants and prediction. For using the technique that gives maximum prediction accuracy, a comparison has been drawn using traditional Binary Logistic Regression, Random Forest, and Deep Learning. Logistic regression, an extension of linear regression having a dependent variable as categorical, will also serve as a basis for comparison. For performing the binary logistic regression, IBM SPSS was used. As the problem under investigation is of a classification type, out of all the available machine learning classifiers, the random forest classifier, however, is towards the top of the classifier hierarchy. A new age tool has also been used: deep learning, which is simply a subset of a neural network with three or more layers. For performing both the random forest and deep learning, R programming was used.

The entire data set, consisting of 376 responses, was divided into training and testing data sets. The training data set was used to generate a model, which was used to predict the outcome for the testing data set.

### *3.1 Deep Learning*

Greater "depth" (complexity) is added to the model, and the input is changed using a variety of functions that permit hierarchical data representation at various levels of abstraction. DL expands on traditional ML. A key benefit of DL is featuring learning, or the automated extraction of features from raw data, where features at higher levels of the hierarchy are composed of features at lower ones.

# 4. Results

As the model is used for prediction, the focus is on the accuracy of prediction, and hence the classification table is very important. The entire data set was randomly divided into training and testing data sets, in the ratio of 70% and 30%, respectively. In the training data set, there were 268 cases, and in the testing data set there were 108 cases. The following section explains the results for both the credit default as well as wilful default, first for the binary logistic regression, second, for the random forest, and third, for the deep learning.

## 4.1 Binary Logistic Regression Results

The typical linear regression is expanded upon by logistic regression. When the dependent variable, Y, is binary, it is applied. When the dependent variable is binary, the link between the predictors (our independent variables) and the predicted variable (the dependent variable) is determined using the statistical method known as logistic regression.

### *4.1.1 Binary Logistic Regression for Credit Default*

**The classification table indicating the predicted vs. the actual credit default cases is shown below for both the training as well as testing data set.**

**Table 1: Classification Table for Credit Default**

<table>
<tr><td colspan="3" rowspan="4">Observed</td><td colspan="6">Predicted</td></tr>
<tr><td colspan="3">Selected Cases</td><td colspan="3">Unselected Cases</td></tr>
<tr><td colspan="2">Credit Default</td><td>Percentage Correct</td><td colspan="2">Credit Default</td><td>Percentage Correct</td></tr>
<tr><td>No</td><td>Yes</td><td></td><td>No</td><td>Yes</td><td></td></tr>
<tr><td rowspan="3">Step 1</td><td rowspan="2">Credit Default</td><td>No</td><td>212</td><td>0</td><td>100.0</td><td>80</td><td>7</td><td>92.0</td></tr>
<tr><td>Yes</td><td>0</td><td>56</td><td>100.0</td><td>5</td><td>16</td><td>76.2</td></tr>
<tr><td colspan="2">Overall Percentage</td><td></td><td></td><td>100.0%</td><td></td><td></td><td>88.9%</td></tr>
</table>

As depicted in Table 1 above, the accuracy for prediction for the training dataset is 100%, but what is more important is to check the accuracy in the test data set. The accuracy in the testing data set comes to be 88.9%, as 96 cases were predicted correctly out of 108 cases. There has been a difference in prediction accuracy between the training data set and testing data set, i.e., 11%. Although the difference is higher, still the accuracy of the prediction for the testing data set is relatively high, i.e., 88.9%.

### *4.1.2 Binary logistic regression for wilful default*

**Table 2: Classification Table for Wilful Default**

<table>
<tr><td colspan="3" rowspan="4">Observed</td><td colspan="6">Predicted</td></tr>
<tr><td colspan="3">Selected Cases</td><td colspan="3">Unselected Cases</td></tr>
<tr><td colspan="2">Wilful Default</td><td rowspan="2">Percentage Correct</td><td colspan="2">Wilful Default</td><td rowspan="2">Percentage Correct</td></tr>
<tr><td>No</td><td>Yes</td><td>No</td><td>Yes</td></tr>
<tr><td>Step</td><td>Wilful</td><td>No</td><td>190</td><td>4</td><td>97.9%</td><td>85</td><td>4</td><td>95.5%</td></tr>
</table>

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | Default | Yes | 3 | 66 | 95.7% | 2 | 22 | 91.7% |
| | Overall Percentage | | | | 97.3% | | | 94.7% |

As depicted in Table 2, for the wilful default, the number of correctly predicted cases is 256 out of 263 cases, resulting in 97.3% predicted accuracy for the training data set. The number of correctly predicted cases for the testing data set is 107 out of 113, resulting in a prediction accuracy of 94.7%. There has been a difference in prediction accuracy between the training data set and the testing data set, i.e., 2.65%, which is a small difference. The small difference indicates that the model trained on the training data set is also effective in predicting the testing data set, indicating the model's power to predict.

### 4.2 Random Forest Results

The very popular machine learning method known as random forest combines the results of various decision trees to get a single result. Its versatility and usefulness, which it uses to address classification and regression problems, are what drive its widespread use. The number of trees and variables that were tested at each split must be specified for Random Forest to run the study. For analyzing and extracting the maximum accuracy, various combinations of a number of trees and variables tried at each split were tested, and the combination resulting in maximum accuracy was selected. Using the same combination, actual and predicted cases for credit default were compared. Using the 'sample' function of the R programming data set was divided into training (263 observations) and testing (113 observations) data sets, using 70% and 30% split, respectively.

*4.2.1 Random Forest for Credit Default*

**Table 3: Classification Table for credit default using Random Forest**

| Observed | | | Predicted | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Selected Cases | | | Unselected Cases | | |
| | | | Credit Default | | Percentage Correct | Credit Default | | Percentage Correct |
| | | | No | Yes | | No | Yes | |
| Step 1 | Credit Default | No | 204 | 5 | 97.6% | 82 | 8 | 99.5% |
| | | Yes | 8 | 46 | 14.8% | 6 | 17 | 92.6% |
| | Overall Percentage | | | | 95.1% | | | 87.6% |

As Table 3 above depicts, for the training data set, correctly predicted cases were 250 out of 263 cases, i.e., 95.1%. Whereas, for the testing data set, correctly predicted cases were 99 out of 113, i.e., 87.6%. There has been a difference in prediction accuracy between the training data set and testing data set, i.e., 7.4%. Although the difference is higher, still the accuracy of the prediction for the testing data set is relatively high, i.e., 87.6%.

*4.2.2 Random Forest for wilful default*

**Table 3: Classification Table for willful default using Random Forest**

| Observed | Predicted | | | |
|---|---|---|---|---|
| | Selected Cases | | Unselected Cases | |
| | Credit default | Percentage | Credit default | Percentage |

| | | | | | Correct | | | Correct |
|---|---|---|---|---|---|---|---|---|
| | | | No | Yes | | No | Yes | |
| Step 1 | Credit default | No | 178 | 16 | 99.5% | 73 | 16 | 99.5% |
| | | Yes | 14 | 55 | 92.6% | 9 | 15 | 92.6% |
| | Overall Percentage | | | | 88.6% | | | 77.9% |

As Table 4 above depicts, for the training data set, correctly predicted cases were 233 out of 263 cases, i.e., 92.6%. Whereas, for the testing data set, correctly predicted cases were 88 out of 113, i.e., 77.9%. There has been a difference in prediction accuracy between the training data set and testing data set, i.e., 10.7%.

### 4.3 Deep Learning Results

The advance of artificial intelligence, predominantly deep learning, has led to stepping up and development in the processing of collected data. Deep ensemble learning models bring together the benefits of both deep learning models and ensemble learning, improving the generalized performance of the resulting model.

The entire dataset, having 376 responses, was divided into training and testing data sets in the 70% and 30% ratio, respectively. The model was trained using the training data set, and then the model was used to predict the response variable for the testing data set.

For improving the model built based on the training data set, values of hidden layer and epoch were utilized. The hidden layer, which is a layer between input and output layers, enables the network to be broken down into specific data transformations. An epoch, which is a hyper-parameter, signifies one complete pass of the training dataset through the algorithm. Every sample in the training dataset, having a chance to revise the internal model parameters, makes one epoch. In this paper, a combination of various values of hidden layers and epoch was used for predicting through deep learning. The number of hidden layers used in this study ranged from 10 to 100, and epochs ranged from 50 to 500.

Although using the same combination of hidden layers and epochs run multiple times resulted in different prediction accuracy. Therefore, the same combination of a number of hidden layers and epoch was run a hundred times to arrive at the average prediction accuracy.

The same procedure was followed for measuring the accuracy of prediction for wilful default as well as credit default. The following section explains the comparison between predicted and actual default**.**

*4.3.1 Deep Learning for Credit Default*

**Table 4: Average Prediction Accuracy of 100 runs for Credit Default**

| | | Epoch | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 50 | 100 | 150 | 200 | 250 | 300 | 350 | 400 | 450 | 500 |
| Hidden Layer | 10 | 98.2% | 98.9% | 98.5% | 98.9% | 98.5% | 98.5% | 98.5% | 97.8% | 98.1% | 97.6% |
| | 20 | 98.4% | 98.8% | 98.8% | 98.6% | 98.6% | 98.5% | 98.3% | 98.0% | 97.7% | 98.0% |
| | 30 | 98.6% | 98.9% | 98.9% | 98.4% | 98.6% | 98.1% | 98.4% | 97.7% | 98.1% | 98.4% |
| | 40 | 98.3% | 98.8% | 98.9% | 98.7% | 98.2% | 98.1% | 97.4% | 98.1% | 98.6% | 98.5% |
| | 50 | 98.8% | 98.7% | 99.1% | 98.7% | 98.5% | 97.9% | 98.5% | 98.2% | 98.2% | 98.5% |
| | 60 | 98.5% | 98.4% | 98.8% | 98.6% | 98.3% | 97.9% | 98.5% | 98.5% | 98.8% | 98.6% |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 70 | 99.0% | 98.9% | 98.6% | 98.6% | 98.6% | 97.6% | 98.2% | 98.5% | 98.9% | 98.5% |
| | 80 | 99.0% | 98.5% | 98.7% | 98.4% | 98.1% | 97.9% | 98.0% | 98.4% | 98.8% | 98.9% |
| | 90 | 98.6% | 98.5% | 98.2% | 98.2% | 98.0% | 98.1% | 98.2% | 98.4% | 98.5% | 98.4% |
| | 100 | 98.4% | 98.4% | 98.7% | 98.8% | 98.2% | 98.0% | 97.9% | 99.0% | 98.5% | 98.5% |

The values in the Table 5 depict the average prediction accuracy of running the model a hundred times using the same number of hidden layers and epoch values on the testing data set. The highest and the lowest average prediction accuracy for the testing dataset are 99.1% and 97.4%, respectively. Considering the average value of averages, 98.4% is the highest prediction accuracy in comparison to other methods used for predicting test dataset, i.e., Logistic Regression (88.9%) and Random Forest (87.6%).

*4.3.2 Deep Learning for Wilful Default*

**Table 5: Average Prediction Accuracy of 100 runs for Wilful Default**

| | | Epoch | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 50 | 100 | 150 | 200 | 250 | 300 | 350 | 400 | 450 | 500 |
| Hidden Layer | 10 | 97.2% | 97.4% | 97.6% | 98.1% | 97.5% | 97.8% | 97.4% | 97.5% | 97.5% | 97.2% |
| | 20 | 96.9% | 97.5% | 97.6% | 97.8% | 98.2% | 98.4% | 97.1% | 97.5% | 96.4% | 97.1% |
| | 30 | 96.7% | 97.6% | 98.2% | 98.1% | 97.8% | 97.6% | 97.6% | 96.7% | 97.2% | 97.7% |
| | 40 | 96.8% | 96.8% | 97.9% | 98.1% | 98.1% | 97.1% | 96.8% | 96.3% | 97.6% | 97.2% |
| | 50 | 96.8% | 96.3% | 97.9% | 98.3% | 98.2% | 97.6% | 96.1% | 97.9% | 98.1% | 97.5% |
| | 60 | 97.4% | 97.0% | 97.3% | 97.9% | 98.6% | 97.1% | 96.7% | 97.7% | 98.1% | 97.4% |
| | 70 | 97.1% | 97.9% | 97.2% | 98.1% | 97.7% | 96.3% | 97.2% | 98.0% | 97.9% | 97.6% |
| | 80 | 97.9% | 97.7% | 96.8% | 97.4% | 98.4% | 97.2% | 97.5% | 97.1% | 97.8% | 98.0% |
| | 90 | 97.3% | 96.8% | 96.9% | 97.8% | 98.4% | 97.7% | 97.3% | 97.7% | 97.9% | 97.5% |
| | 100 | 97.4% | 97.9% | 97.2% | 98.1% | 98.2% | 96.8% | 97.0% | 97.6% | 97.7% | 97.6% |

The values in the Table 6 depict the average prediction accuracy of running the model a hundred times using the same number of hidden layers and epoch values on the testing data set. The highest and the lowest average prediction accuracy for the testing dataset are 98.6% and 96.1%, respectively. Considering the average value of averages, which is 97.5%, is the highest prediction accuracy in comparison to other methods used for predicting test dataset, i.e., Logistic Regression (94.69%) and Random Forest (77.88%).

## 5. Discussion

In our experiments, we divided the dataset into 70% and 30% for training and testing. The decision to use the model was not made only on the basis of better performance indicators; rather, it involved a thorough assessment that took the model's ability to generalise into account. It is critical to base credit choices in the context of credit scoring on models that are resistant against overfitting, highlighting the necessity of dataset segmentation. Deep learning outperformed other modelling techniques in terms of accuracy-based model performance, followed by random forest, logistic regression, and deep learning. These results are consistent with those of other investigations, including those by (Mancisidor et al. 2022; Thomas 2000).

This study's main objective was to investigate how machine learning approaches could be used to categorise credit applicants, with a special attention paid to the underlying mathematical ideas that underlie these methodologies. Deep learning, random forest, and logistic regression were among the specific techniques examined. Since, the The dataset used to study the issue was unbalanced, therefore we employed cost-sensitive learning to make the algorithms more harsh when misclassifying members of the minority class.

## 6. Limitations and Future Scope

The research's conclusions came from modelling exercises using a single dataset. Through the addition of a wider and more varied set of datasets, a more thorough evaluation may be carried out in order to examine the resilience and reliability of classification algorithms in the field of credit scoring. The analysis's robustness can be greatly increased by using additional datasets.

Our research focused mostly on well-recognized machine learning methodologies, with just a scant examination of cutting-edge methods that might support the fundamental algorithms. It should be noted that the potential uses of convolutional neural networks and selected ensemble methods in the context of credit scoring were not thoroughly explored in our research. These novel methods need more investigation and can provide insightful information for upcoming field research projects.

## 7. Conclusion

In recent years, for the behavioural credit scoring process, the use of sophisticated techniques have shown rapid growth. The methods adopted for credit scoring can be considered as an innovative approach that is inspired by artificial intelligence and machine learning, such as, deep learning, random forest and ensemble methods. The future of machine learning models in the further exploring behavioural credit scoring model will be more prevalent. Various reasons support the use of machine learning models, such as these models might help improve risk differentiation by improving the discriminatory power of the model and providing tools for identifying all risk drivers. Machine learning models also help confirm the data features and give a data-driven perspective for the feature selection process.

## Conflicts of Interest Statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

## References

[1]. Ala'raj, M., Abbod, M., & Radi, M. The applicability of credit scoring models in emerging economies: an evidence from Jordan. *International Journal of Islamic and Middle Eastern Finance and Management*, *11*(4), (2018). 608–630. https://doi.org/10.1108/IMEFM-02-2017-0048

[2]. Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society*, *73*(1), 70–90. (2022). https://doi.org/10.1080/01605682.2021.1922098

[3]. Bueff, A. C., Cytryński, M., Calabrese, R., Jones, M., Roberts, J., Moore, J., & Brown, I. Machine learning interpretability for a stress scenario generation in credit scoring based on counterfactuals. *Expert Systems with Applications*, *202*, (2022). 117271. https://doi.org/10.1016/j.eswa.2022.117271

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

[5]. de Moraes, C. O., & Costa, Á. Credit behavior and Financial Stability in an Emerging Economy. *Economic Systems*, 100999. (2022). https://doi.org/10.1016/j.ecosys.2022.100999

[6]. Demaris, A. A Tutorial in Logistic Regression Published by : National Council on Family Relations. *National Council on Family Relations*, *57*(4), 956–968, (1995).

[7]. Dumitrescu, E.-I., Hué, S., Hurlin, C., & Tokpavi, S. Machine Learning or Econometrics for Credit Scoring: Let's Get the Best of Both Worlds. *SSRN Electronic Journal*, *Fre 2014*, 1–42, (2020). https://doi.org/10.2139/ssrn.3553781

[8]. Goel, A., & Rastogi, S. Understanding the impact of borrowers' behavioural and psychological traits on credit default: review and conceptual model. *Review of Behavioral Finance*, (2021). https://doi.org/10.1108/RBF-03-2021-0051

[9]. Goh, R. Y., & Lee, L. S. Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches. *Advances in Operations Research*, 1–30, (2019). https://doi.org/10.1155/2019/1974794

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

[11]. Hsieh, N.-C. An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, *27*(4), 623–633, (2004). https://doi.org/10.1016/j.eswa.2004.06.007

[12]. Kennedy, K., Mac Namee, B., Delany, S. J., O'Sullivan, M., & Watson, N. A window of opportunity: Assessing behavioural scoring. *Expert Systems with Applications*, *40*(4), 1372–1380, (2013). https://doi.org/10.1016/j.eswa.2012.08.052

[13]. Khalilia, M., Chakraborty, S., & Popescu, M. Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, *11*(1), (2011). 51. https://doi.org/10.1186/1472-6947-11-51

[14]. Kokate, S., & Chetty, M. S. R. Credit Risk Assessment of Loan Defaulters in Commercial Banks Using Voting Classifier Ensemble Learner Machine Learning Model. *International Journal of Safety and Security Engineering*, *11*(5), 565–572. (2021). https://doi.org/10.18280/ijsse.110508

[15]. Lappas, P. Z., & Yannacopoulos, A. N. A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Applied Soft Computing*, *107*, 107391, (2021). https://doi.org/10.1016/j.asoc.2021.107391

[16]. LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. *Nature*, *521*(7553), 436–444, (2015). https://doi.org/10.1038/nature14539

[17]. Mancisidor, R. A., Kampffmeyer, M., Aas, K., & Jenssen, R. Generating customer's credit behavior with deep generative models. *Knowledge-Based Systems*, *245*, 108568, (2022). https://doi.org/10.1016/j.knosys.2022.108568

[18]. Mhlanga, D. Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment. *International Journal of Financial Studies*, *9*(3), 39, (2021). https://doi.org/10.3390/ijfs9030039

[19]. Moscato, V., Picariello, A., & Sperlí, G. A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, *165*((2021). 113986. https://doi.org/10.1016/j.eswa.2020.113986

[20]. Neto, R., Jorge Adeodato, P., & Carolina Salgado, A. A framework for data transformation in Credit Behavioral Scoring applications based on Model Driven Development. *Expert Systems with Applications*, *72*, 293–305, (2017). https://doi.org/10.1016/j.eswa.2016.10.059

[21]. Nikitin, N. O., Kalyuzhnaya, A. V., Bochenina, K., Kudryashov, A. A., Uteuov, A., Derevitskii, I., & Boukhanovsky, A. V. Evolutionary Ensemble Approach for Behavioral Credit Scoring. In *Computational Science – ICCS 2018* (pp. 825–831). Springer, Cham. (2018). https://doi.org/10.1007/978-3-319-93713-7_81

[22]. Onay, C., & Öztürk, E. A review of credit scoring research in the age of Big Data. *Journal of Financial Regulation and Compliance*, *26*(3), 382–405, (2018). https://doi.org/10.1108/JFRC-06-2017-0054

[23]. Óskarsdóttir, M., & Bravo, C. Multilayer network analysis for improved credit risk prediction. *Omega*, *105*, 102520, (2021). https://doi.org/10.1016/j.omega.2021.102520

[24]. Remolina, N. The Role of Financial Regulators in the Governance of Algorithmic Credit Scoring. *SSRN Electronic Journal*, 1–25, (2022). https://doi.org/10.2139/ssrn.4057986

[25]. Simumba, N., Okami, S., Kodaka, A., & Kohtake, N. Multiple objective metaheuristics for feature selection based on stakeholder requirements in credit scoring. *Decision Support Systems*, *155*, 113714, (2022). https://doi.org/10.1016/j.dss.2021.113714

[26]. Sum, R. M., Ismail, W., Abdullah, Z. H., Shah, N. F. M. N., & Hendradi, R. A New Efficient Credit Scoring Model for Personal Loan Using Data Mining Technique for Sustainability Management. *Journal of Sustainability Science and Management*, *17*(5), 60–76, (2022). https://doi.org/10.46754/jssm.2022.05.005

[27]. Thomas, L. C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, *16*(2), 149–172, (2000). https://doi.org/10.1016/S0169-2070(00)00034-0

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

[29]. Tripathi, D., Shukla, A. K., Reddy, B. R., Bopche, G. S., & Chandramohan, D. Credit Scoring Models Using Ensemble Learning and Classification Approaches: A Comprehensive Survey. *Wireless Personal Communications*, *123*(1), 785–812, (2022). https://doi.org/10.1007/s11277-021-09158-9

[30]. Trivedi, S. K. A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society*, *63*, 101413, (2020). https://doi.org/10.1016/j.techsoc.2020.101413

[31]. Yuping, Z., Jílková, P., Guanyu, C., & Weisl, D. New Methods of Customer Segmentation and Individual Credit Evaluation Based on Machine Learning. *Proceedings of the "New Silk Road: Business Cooperation and Prospective of Economic Development" (NSRBCPED 2019)*, *131*, 925–931, (2020). https://doi.org/10.2991/aebmr.k.200324.170

[32]. Zhu, B., Yang, W., Wang, H., & Yuan, Y. A hybrid deep learning model for consumer credit scoring. *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 205–208, (2018). https://doi.org/10.1109/ICAIBD.2018.8396195