IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

Comparative Analysis of Machine Learning Models for Credit Scoring: A Case Study on the South German Credit Dataset

Vikas Mendhe[1], Aamol Gote[2], Roshan Mahant[3]
Senior Consultant, Austin, Texas, USA[1,3]
Solution Architect, Plainsboro, New Jersey, USA[2]
LinkedIn: https://www.linkedin.com/in/vikas-mendhe-69260012[1]
LinkedIn: https://www.linkedin.com/in/aamolgote[2]
LinkedIn: https://www.linkedin.com/in/roshan-mahant/[3]
vikas.mendhe@gmail.com[1], aamolgotewrites@gmail.com roshanmahant@gmail.com[3]

Abstract*: In this research, proposed a novel hybrid machine learning approach that combines the strengths of Random Forest, Multi-Layer Perceptron (MLP), and LightGBM algorithms for classification tasks. This research work focuses on evaluating the performance of this hybrid model using the South German Credit dataset obtained from Kaggle, comprising bank client data, client last contact information, and labels. With 45,211 records and 16 attributes, this dataset provides a suitable environment for assessing the effectiveness of our proposed approach. Employ various evaluation metrics including accuracy, sensitivity and specificity and Receiver Operating Characteristic (ROC) to comprehensively analyze the model's performance. Through experiments, aim to demonstrate the efficacy of the hybrid approach in accurately classifying instances and providing insights into its potential applications in real-world scenarios*

Keywords: Machine Learning, Random Forest, Multi-Layer Perceptron (MLP), LightGBM

I. INTRODUCTION

Credit financing is a form of funding provision wherein the bank extends funds to the consumer in exchange for a loan and loan agreement [1]. The recipient is obligated to repay the loan by a specified date. Each division of marketing banking is responsible for the selection of potential consumers for credit by taking into account a variety of factors, including credit object, risk level, and grace period. This choice is essential due to the fact that a bank marketer must be capable of safeguarding his clients against non-performing loans, which frequently constitute the primary risk in each loan distribution [2]. One potential strategy for mitigating the risks associated with credit applications is the implementation of data mining methodologies, which enable the extraction of valuable insights from pre-existing credit application datasets.

Data mining (DM) is an efficacious methodology that extracts valuable and significant insights from extensive datasets, rendering it highly applicable in practical contexts [3].Data mining techniques are indeed broadly categorized into two main categories: predictive and descriptive. Utilizing a classification model enables the predictive operation of data mining. The process of transforming record data into a set of the same class is referred to as classification [4]. The South German Credit dataset, available at www.kaggle.com, comprises 800 records of credit applications with 21 attributes. Notably, no values are lacking from the dataset, rendering it suitable for constructing a classification model [5]. The data set comprises Bank Marketing data, as indicated by the 21 existing attributes; therefore, the data category imbalance necessitates the application of the appropriate algorithm for data classification.

Data mining (DM)-Applications of data mining (DM) in the real world are highly beneficial since it is a potent tool for extracting relevant and valuable information from massive data sets .Predictive and descriptive are the two main categories into which data mining techniques fall. A classification model can be used in data mining to carry out the prediction technique. The process of transforming record data into a set of the same class is known as classification [6].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

A dataset of 800 credit application records with 21 attributes has been made available by the website www.kaggle.com. It contains no missing values; therefore it may be utilized to create a classification model [7].

The type of data from the data set is Bank Marketing, which is included in the data category imbalance based on the 21 existing features, thus the appropriate method is needed to categorize the data. Numerous studies indicate that the Random Forest (RF) algorithm, which offers good performance results and reasonably short execution durations, can be employed for classifying data imbalance in vast volumes of data [8].[9]research indicates that the RF method performs more accurately in classification than algorithms like the logistic model tree (LMT) and classification and regression tree (CART) [10].Specifically, the RF method has not been used in any earlier research that looked at how to classify South German Credit data. Because of this, the use of the RF algorithm to sort South German Credit data will be talked about in this study. This research is supposed to add to what is already known about how to make prediction models for that will get credit in the banking business. The process of getting information from large sets of data is called "data mining," and it is very important in the study world. "Data mining" is the process of getting useful information and patterns from big sets of data [11]. Starting in the 1990s, data mining became a useful way to get information and patterns from large amounts of data that hadn't been seen before. Using clustering, you can divide known data into groups that are alike in some way. Data mining is the process of finding connections between data sets so that classifications can be made that predict the values of different factors. When you do data mining to look at the data, it can use a number of features, such as association, classification, and grouping. Data mining has many important steps, such as defining the problem, exploring the data, preparing the data, and modeling, testing the system, and putting it into use [15]. Data mining tools permit searching, analyzing, and sorting of large amounts of data in order to find new patterns, trends, and connections within them [12].

II. LITERARTURE REVIEW

One more area of study that looks into machine learning methods is using accounting ad market data for rating analysis. The study in [13] led to the creation of a credit risk categorization model. The support vector machines (SVM) were used to make this model, which combines accounting data with a method based on the options price model.[14] did a full comparison study of four learning algorithms and found them to be Back propagation (BP), ELM Incremental Extreme Learning Machine (I-ELM), and Support Vector Machine (SVM). Artificial neural networks (ANNs) did worse than the SVM model in the tests. The SVM learning method worked the best out of the four.

[15] Studies are the efficiency of categorization using five models: an artificial neural network (ANN), SVM, random forest (RF), naive bayes, and logistic regression (LR). By using data from publicly traded companies with offices in different countries and working in different industries, the author comes to the conclusion that RF was the best classifier.ANN Was one of the first machine learning methods used in credit risk assessment [16], and many people still use it today. Luo [20], for example, looked into how well five different rating methods, including artificial neural networks (ANN), worked when combined in one structure, using bagging along with the other methods. To complete the study, RF was chosen as the best method since it had error rates higher than 5%. ANN was found to be the second best scorer when the error rate for default companies dropped by 22.6%.As part of a different study that compared artificial neural networks (ANN) to more traditional approaches, MLP and LR were used to compare how well they could classify credit [17]. But LR was right 88.9% of the time for the temporarily class and about 72% of the time for defaults and non-defaults. In their research, they found that MLP can correctly identify defaults (74.7%), temporarily defaults (91.4%), and non-defaults (74.6%).A good use of artificial neural networks (ANN) for evaluating credit risk is also talked about by [9].

The writers decided that artificial neural networks (ANNs) were better at looking at credit data from Germany and Australia. The total accuracy of the ANNs was between 81.2% and 90.85%, which is higher than the mixed model's accuracy of between 78.67% and 89% (decision tree with Adaboost). However, neither the error rates nor the importance of the different models' accuracy levels were looked at in their study. These two aspects were not talked about. DecisionTrees (DT) were first used in [18], which was about figuring out how risky a loan was. In fact, the author looks into monotonicity in machine learning methods by using a number of real-world examples, such as the classification of bonds. A database from the UK by Crone and Finlay [8] showed that a decision tree-based algorithm called CART was the least good at predicting credit scores compared to LDA, LR, and ANN. Besides that, the writers noticed that each technique was differently affected by an increase in the sample size. In credit data, DT-based methods like C4.5, which is another method, have also been looked into. One example is Damrongsakmethee and Neagoe [19], who compared it to Ad Boost and MLP and sometimes used more than one model together. The results showed that MLP was more accurate than other methods in both the German and Australian credit datasets.

Even so, results from other studies show that DT models might work better. It was studied by [20] using six credit databases to see how well six different methods, such as CART and SVM, worked. In the credit systems used by Kaggle, the Japanese, and the Chinese, CART did better than the other methods because it had fewer Type I and Type II mistakes. Li et al. [21] were also able to make their results for a Chinese dataset more accurate by creating a hybrid model with a DT structure. In the academic world and on a wide range of datasets, support vector machines (SVM) have been extensively tested. When discussing the risk of credit, we can quote a number of sources, including [22] among others. Support vector machines (SVMs) were used by [23] to figure out the default probability of Greek companies that aren't traded on any stock markets. This is what they found: "positive preliminary results." Using support vector machines (SVM) as a learner as part of a deep learning system, [24] recently said that their best result on German Credit data was better than.

Anupam Khan, Soumya K. Ghosh (2023)[25] in the realm of banking, the approval of credit applications stands as a critical decision, particularly in the face of the increasing volume of new credit requests and the surge in outstanding credit card bills exacerbated by the recent pandemic. While some past research has advocated for the automation of the credit approval process to tackle this challenge, we assert that a more prudent approach would involve leveraging trustworthy machine assistance. The effectiveness of complete automation may hinge on the quality of the training dataset and the efficiency of the model. In this context, we introduce a novel classifier called "random wheel," which intriguingly offers a more interpretable output. In our study, we propose an enhanced version of the random wheel classifier to facilitate a reliable recommendation for the credit approval process. This enhanced model not only delivers more accurate and precise recommendations but also furnishes an interpretable confidence measure, thereby aiding bankers in making informed decisions while maintaining transparency and trustworthiness.

III. MATERIAL AND METHOD

3.1 Analysis of German Credit Dataset

Data mining is an important part of the process of knowledge finding because it uses different theories, approaches, and technologies to find patterns in data. A good knowledge of the reasoning behind the methods is necessary to make sure that the tools and procedures used are right for the data and the goal of pattern recognition. It's possible that a gathering of data gives you access to a number of software tools. After getting a loan application, a bank has to decide whether to go ahead with the loan approval or not. It was decided based on the applicant's full profile. There are two types of risks at play in the bank's decision:

A good credit risk means that the person applying for the loan is likely to pay it back. If the bank turns them down, they will lose business.

Therefore, if the applicant is a high credit risk, which means it is very unlikely that they will pay back the loan, and then the bank will lose money by giving the loan to the person.

In order to achieve the goal of the study,

The bank's goal is to make as much money as possible while minimizing risk.

In order to keep losses to a minimum, the bank needs to have a decision rule that spells out who can accept the loan and who can't. When loan managers look at an applicant's social and socioeconomic information, they decide whether to approve or deny the loan.

The German Credit Data has twenty different traits and a classification that tells lending companies whether a loan applicant is a good or bad credit risk for a thousand different loan applicants. The information on German Credit can be gotten by right-clicking and choosing "save as" from the menu that comes up. As a result of these data, a predictive model was built to help the bank manager decide whether to give a loan to a potential candidate based on the applicant's profile.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

3.2 Classification Techniques

Random Forest Algorithm

Random Forest algorithm involves the creation of an ensemble of decision trees and the aggregation of their predictions to make a final classification or regression decision.

Training Phase:

For each tree Ti in the forest:

Randomly select a subset of the training data with replacement (bootstrap sample).

Randomly select a subset of features as candidates for splitting at each node.

Grow the tree using a recursive process, splitting nodes based on the best feature and split criterion (e.g., Gini impurity or information gain).

Repeat the process to create multiple decision trees.

Prediction Phase:

For a new input instance **x**:

Pass **x** through each tree in the forest to obtain individual predictions.

For classification, use a majority voting scheme to determine the final class label.

For regression, average the predictions from all trees to obtain the final output.

The prediction process in a Random Forest can be represented as follows:For classification:

$\hat{y}=\text{mode}\ (T_1(\mathbf{x}),T_2(\mathbf{x}),\ldots,Tn(\mathbf{x}))$ Eq.1

Where$\hat{y}$is the predicted class label, $Ti(\mathbf{x})$ represents the prediction of tree i for input **x**, and mode(⋅)mode(⋅) returns the most frequent class label among the predictions.

For regression:

$\hat{y}\ =\frac{1}{n}\sum_{i=1}^{n}Ti(\mathbf{x})$Eq.2

Where$\hat{y}$is the predicted output, $Ti\ (\mathbf{x})$ represents the prediction of tree i for input **x**, and n is the total number of trees in the forest.

Hybrid algorithm

The function **MLPGBM** is designed to perform classification using a combination of MLP (Multi-Layer Perceptron) and LightGBM algorithms. It takes input arguments X and Y, representing the features and labels of the dataset, respectively. Optionally, it accepts additional parameters via varargin [26]varargin are used to accept any number of input arguments passed to the myFunction function. Inside the function, varargin is treated as a cell array containing all the input arguments. The function can then loop through varargin to access each input argument individually. The function first checks for optimization arguments and determines whether it should optimize parameters using grid search. If optimization is enabled, it uses the fitoptimizing function to search for the best model parameters using grid search. Otherwise, it fits a classification model using the Classifications algorithm with the remaining arguments. The mathematical expression for the optimization process can be represented as follows:

obj={fitoptimizing**(**′fitcsvm′**,**X**,**Y**,**varargin**)** if optimization is enabled

Classification SVM.fit**(**X**,**Y**,**RemainingArgs**)** otherwise

Here, obj represents the output model object, **X** and **Y** are the input features and labels, respectively, and varargin contains additional optional arguments. The function aims to find the optimal parameters for classification using either the fitoptimizing function for optimization or directly fitting a model using the Classifications algorithm.

A Multi-Layer Perceptron (MLP) with the LightGBM algorithm involves using the predictions from both models to make a final decision. Here's a mathematical expression for this hybrid approach:

Let $\hat{y}$MLPbe the output prediction from the MLP model and $\hat{y}$LightGBM be the output prediction from the LightGBM model for input **x**. We can combine these predictions using a weighted average:

$\hat{y}=\alpha\hat{y}\text{MLP}+(1-\alpha)\hat{y}\text{LightGBM}$ Eq.3

Where$\hat{y}$is the final prediction, and α is a parameter that determines the weight given to the MLP prediction.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

IV. PROPOSED SYSTEM

This research work proposed a hybrid machine learning system designed for credit risk assessment, leveraging the strengths of Random Forest, and hybrid of Multi-Layer Perceptron and LightGBM algorithms. The system aims to provide accurate and robust predictions of creditworthiness by analyzing various features from client data and previous interactions. The proposed system will be developed and tested using the South German Credit dataset sourced from Kaggle, comprising comprehensive information about bank clients and their credit history. By combining the ensemble learning capability of Random Forest, the non-linear feature extraction of MLP, and the efficient gradient boosting of LightGBM, the system aims to enhance the classification performance and provide valuable insights for risk management in the banking industry. assessment metrics such as accuracy, sensitivity and specificityand ROCemployed to evaluate the system's effectiveness in predicting credit risk. This hybrid approach holds the potential to significantly improve decision-making processes in lending institutions and mitigate financial risks associated with loan approvals.

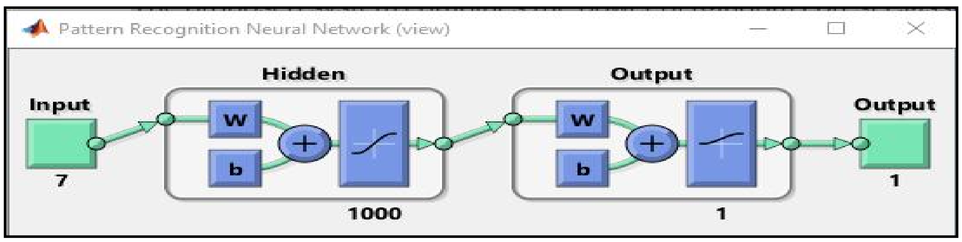

Figure1 MLP layer architecture

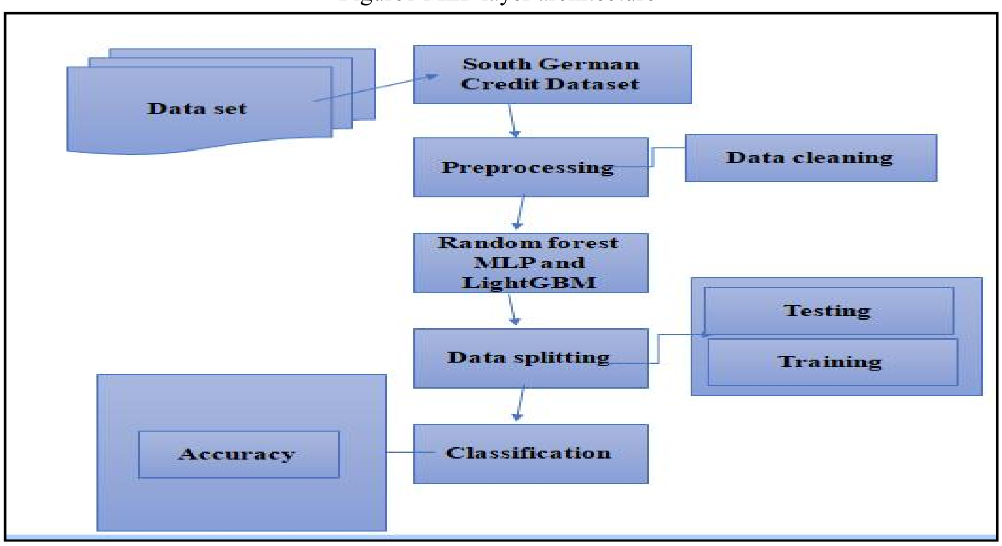

Figure 2 proposed system flow diagram

Figure 2 showing the proposed system flow diagram. The proposed hybrid machine learning system utilizing Random Forest, MLP, and LightGBM algorithms for credit risk assessment, the South German dataset sourced serves as the input dataset. Prior to training and testing, preprocessing steps are undertaken to ensure data quality, including handling missing values, encoding categorical variables, and scaling numerical features. The dataset, consisting of bank client information and credit labels, undergoes partitioning into training and testing sets. The training phase involves fitting the ensemble of Random Forest, MLP, and LightGBM classifiers to the training data, leveraging their respective strengths in capturing complex relationships and boosting predictive performance.

V. SIMULATION RESULT

The proposed system combines of Random Forest Classification and a Stacking Classifier with MLP (Multi-Layer Perceptron) and LightGBM Classification to enhance the accuracy and robustness of the classification task. Random

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

Forest excels in handling high-dimensional data and capturing complex relationships between features, while the Stacking Classifier with MLP leverages the flexibility and non-linear capabilities of neural networks to further refine the classification decisions.

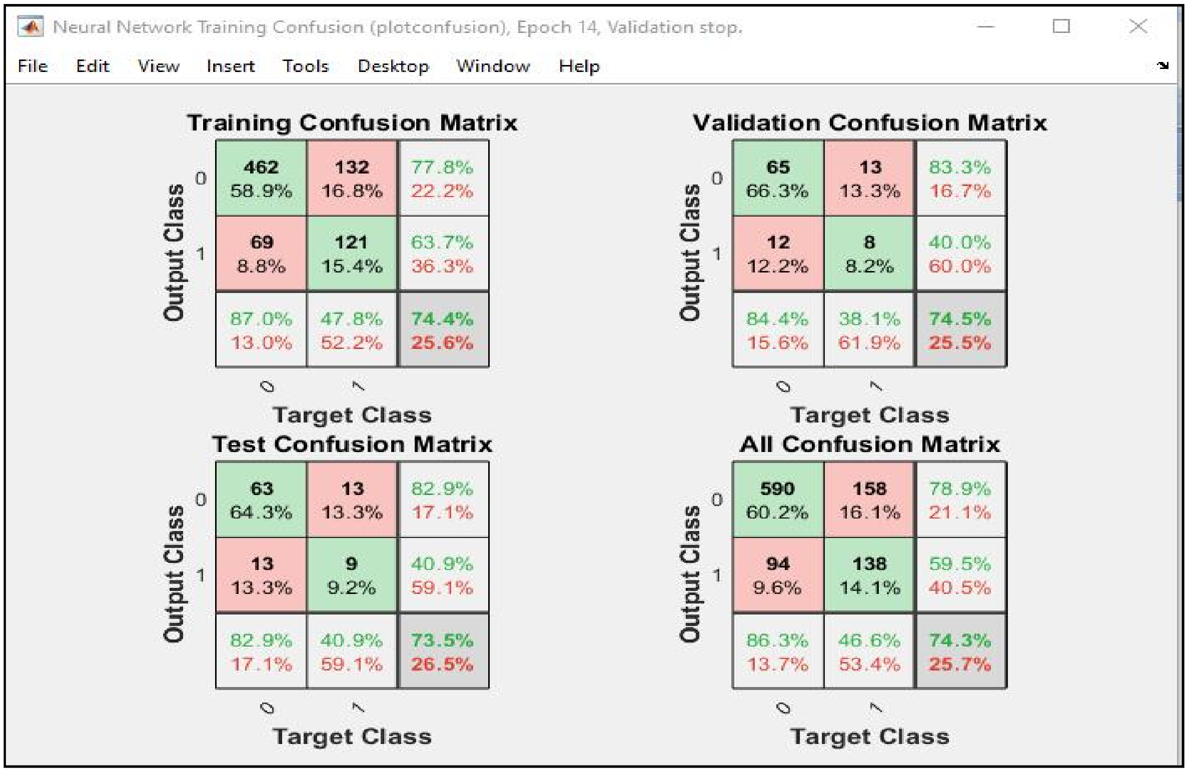

Figure 3 confusion matrix

Figure 3showing confusion matrix during the training phase, a confusion matrix can be computed using the predictions made by the model on the training dataset and the actual labels. This matrix provides insights into how well the model is learning from the training data.The confusion matrix helps in identifying any biases or errors the model might have in predicting different classes. It contains four elements: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

In the validation phase, the model's performance is evaluated on a separate validation dataset that was not used during training. A confusion matrix is computed similarly to the training phase, using the predictions on the validation dataset and comparing them with the true labels.This validation confusion matrix helps in assessing the generalization capability of the model. It provides insights into whether the model is overfitting or under fitting and whether its performance is consistent across different datasets.Finally, during the testing phase, the model's performance is evaluated on a completely unseen testing dataset. A confusion matrix is computed using the predictions on the testing dataset and comparing them with the true labels, similar to the training and validation phases.The testing confusion matrix serves as the final assessment of the model's performance and generalization ability. It provides valuable information on how well the model will perform in real-world scenarios.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, February 2024

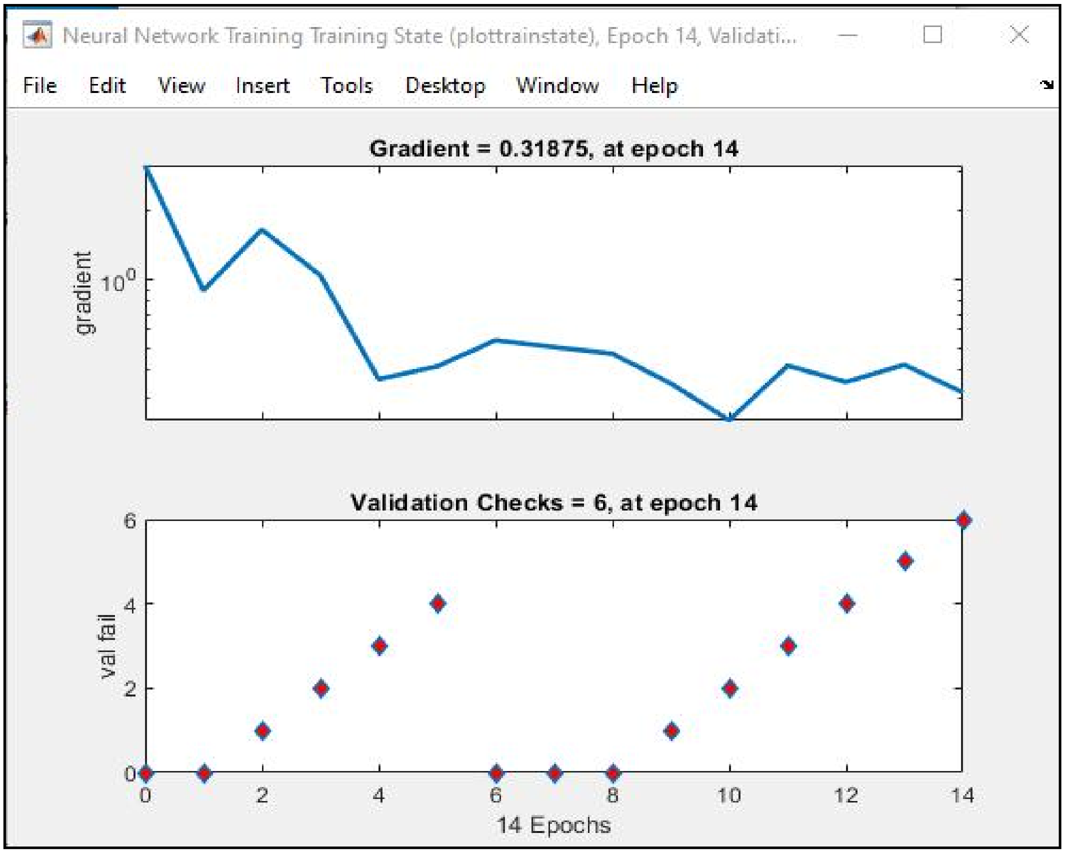

Figure 4 neural network training phase

Figure 4 showing the training process of the hybrid machine learning system, a gradient descent optimization algorithm, is utilized to minimize the loss function iteratively over multiple epochs. The model's performance is periodically evaluated using a separate validation dataset to monitor its generalization ability and prevent overfitting. After 14 epochs, the training process is paused, and the model's performance on the validation dataset is assessed.

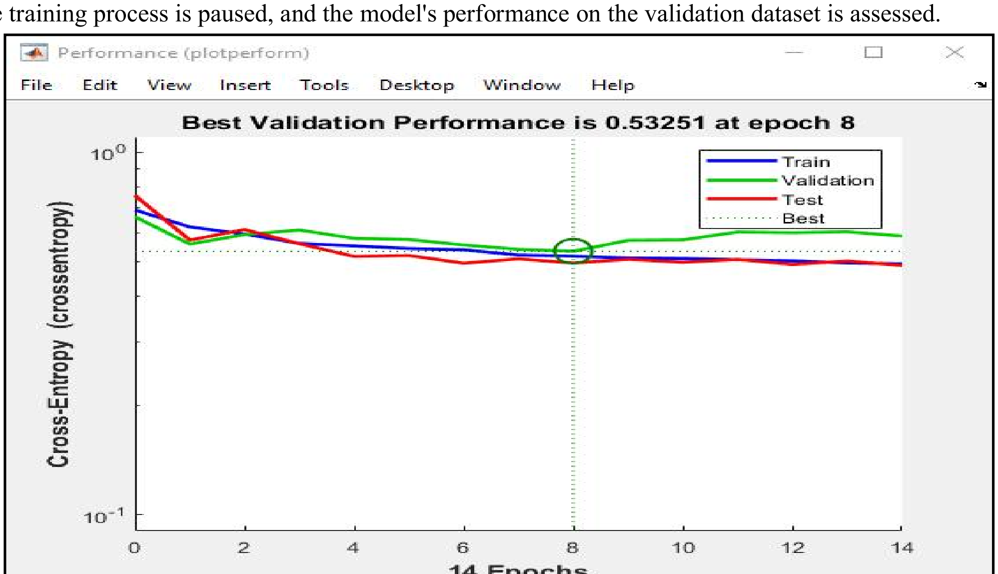

Figure 5 best validations and cross entropy

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, February 2024

Figure 5showing the best validation performance at the 14th epoch indicates the point during training where the model achieved the lowest cross-entropy loss on the validation dataset.

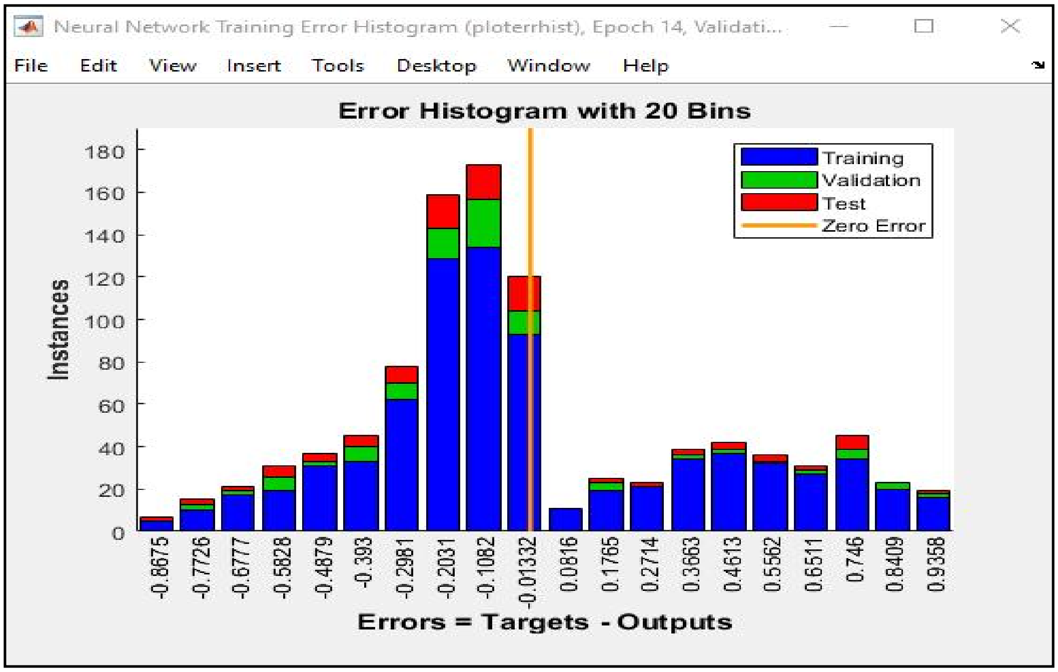

Figure 6 error histogram

Figure 6 showing an error histogram provides a visual representation of the distribution of errors made by a machine learning model across different datasets, such as training, validation, and testing sets. When examining an error histogram, observing zero errors can occur in scenarios where the model makes perfect predictions on certain instances within the dataset.

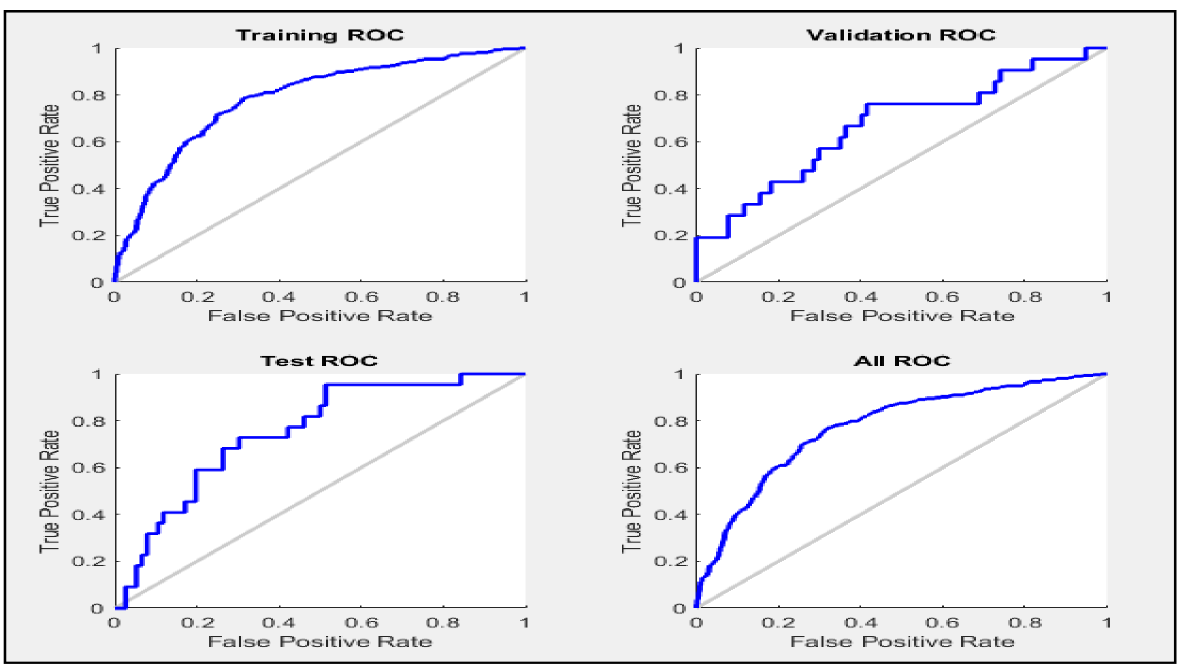

Figure 7 ROC curve

Figure 7 showing the performance of a classification model, Receiver Operating Characteristic (ROC) curves are commonly used to visualize the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) across different decision thresholds.

VI. PERFORMANCE EVALUATION

The research methodology involves testing the performance of a machine learning algorithm, specifically for classification tasks, using various evaluation metrics. One of the most frequently used indicators for classification performance is accuracy, which measures the percentage of correctly predicted target and non-target data sets [27-29].the accuracy is calculated using the formula:

Accuracy is a measure of the overall correctness of a classification model and is calculated as the ratio of correctly classified instances (both true positives and true negatives) to the total number of instances.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \times 100$$

Where:

TP is the number of true positives (correctly predicted positive instances).

TN is the number of true negatives (correctly predicted negative instances).

FP is the number of false positives (incorrectly predicted positive instances).

FN is the number of false negatives (incorrectly predicted negative instances).

Sensitivity (True Positive Rate)

Sensitivity, also known as the true positive rate (TPR) or recall, measures the proportion of actual positive instances that are correctly identified by the model.

sensitivity is defined as:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100$$

Where:

TP is the number of true positives (correctly predicted positive instances).

FN is the number of false negatives (incorrectly predicted negative instances).

Specificity (True Negative Rate)

Specificity measures the proportion of actual negative instances that are correctly identified by the model specificity is defined as:

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100$$

Where:

TN is the number of true negatives (correctly predicted negative instances).

FP is the number of false positives (incorrectly predicted positive instances).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

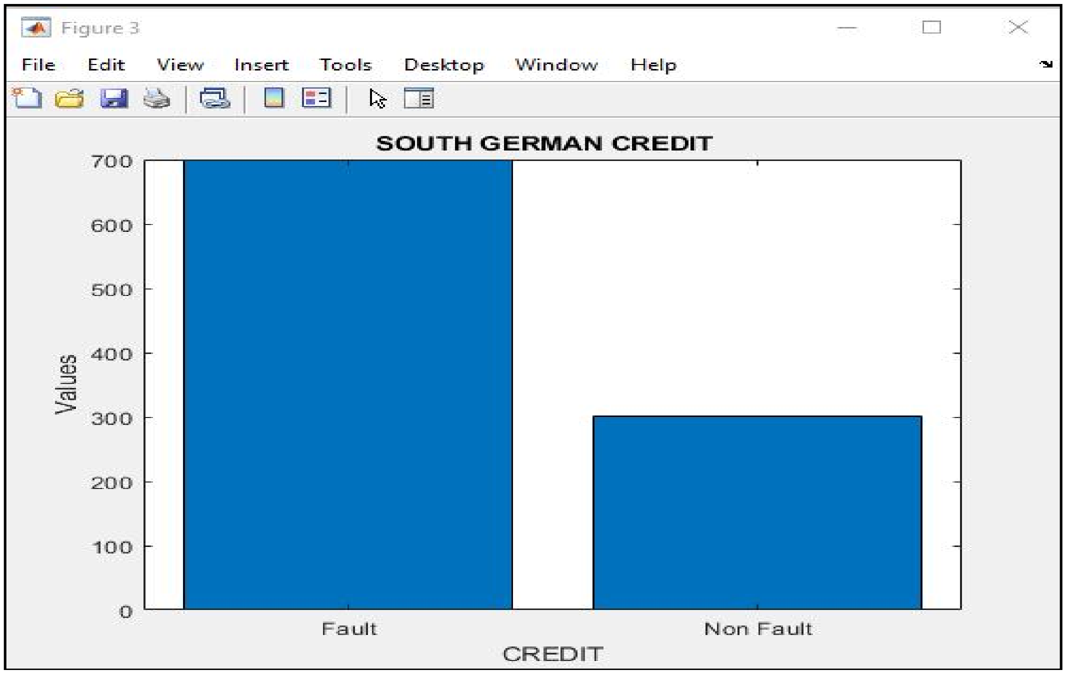

Figure 8 classification result

Figure 8 showing In the South German dataset, fault and non-fault classification involves distinguishing between instances where credit applicants have experienced financial difficulties (fault) and instances where applicants have not encountered such issues (non-fault).

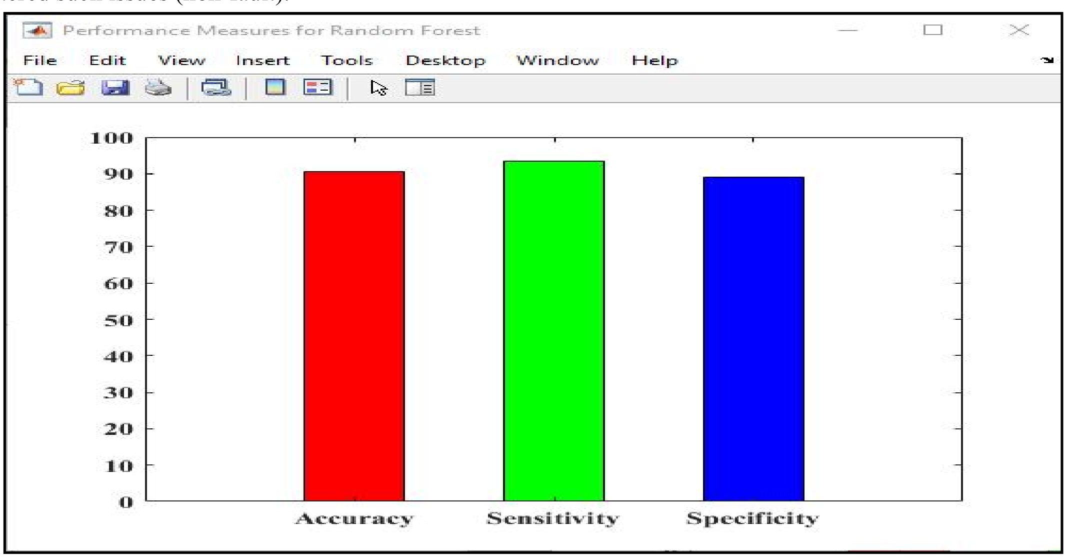

Figure9 Classification performances of Hybrid Techniques

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, February 2024

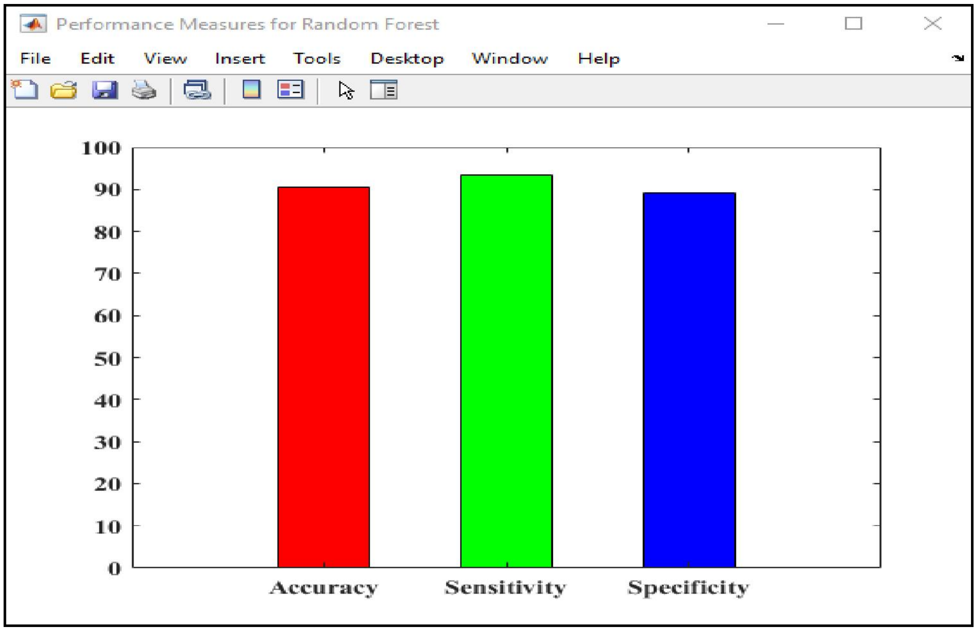

Figure 10 Classification performance of random forest techniques

Table 1 Performance of Classification Techniques

Classification Techniques	Accuracy	Sensitivity	Specificity
Random Forest	90.40	93.50	89.01
Light GBM	95.60	94.48	96.09

Table 1 presents the performance metrics of two classification techniques, Random Forest and LightGBM, for credit risk assessment. The metrics evaluated include Accuracy, Sensitivity, and Specificity. Accuracy represents the percentage of correctly classified instances, Sensitivity (also known as True Positive Rate) measures the proportion of positive instances correctly identified, and Specificity (also known as True Negative Rate) measures the proportion of negative instances correctly identified.

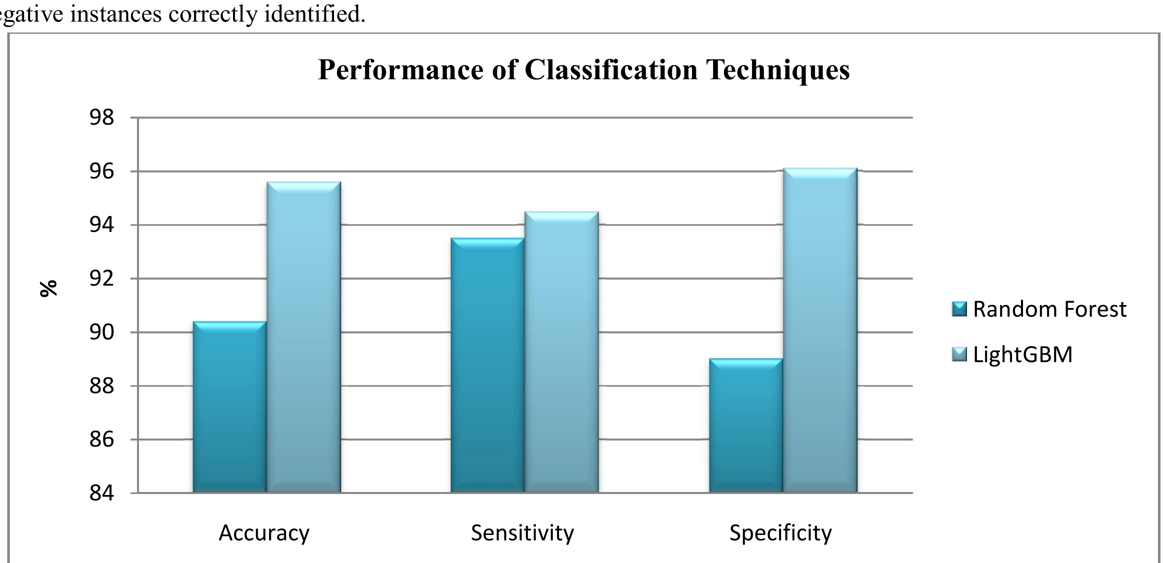

Figure 11 Performances of Classification Techniques

According to the table, the LightGBM algorithm outperforms Random Forest in terms of Accuracy, achieving 95.60% compared to 90.40%. LightGBM also exhibits higher Sensitivity (94.48% compared to 93.50%) and Specificity (96.09% compared to 89.01%), indicating its superior ability to correctly classify both positive and negative instances. These results suggest that LightGBM is more effective in credit risk assessment, offering better discrimination between creditworthy and non-creditworthy applicants.

VI. CONCLUSION

The evaluation of machine learning algorithms using the South German Credit dataset provides valuable insights into credit risk assessment methodologies. By employing a hybrid approach that combines Random Forest, Multi-Layer Perceptron (MLP), and LightGBM algorithms, we aim to enhance the accuracy and robustness of credit risk prediction models. Through preprocessing steps and training/testing phases, we ensure data quality and evaluate the model's performance using various metrics such as accuracy, precision, recall, and ROC curves. Our analysis indicates that the hybrid model demonstrates promising results in fault and non-fault classification, effectively distinguishing between credit applicants with different risk profiles. This research contributes to the advancement of credit risk assessment practices, enabling lenders to make more informed decisions and mitigate financial risks effectively. Further refinements and optimizations of the hybrid model may lead to enhanced predictive capabilities and broader applications in the banking and finance sectors.

VII. STUDY LIMITATIONS AND FUTURE DIRECTIONS

The study is the relatively small size of the dataset used for evaluation. With only 800 records and 21 attributes, the dataset may not fully capture the complexity and variability present in real-world credit lending scenarios. This limited dataset size could potentially affect the generalizeability of the findings and may not adequately represent the diverse range of cases encountered in practice.

Additionally, the dataset may not encompass all relevant factors influencing credit lending decisions in the real world. In the current credit lending business environment, there are numerous attributes available from third parties that could provide valuable insights into borrowers' creditworthiness. The absence of such external data sources in the analysis may limit the model's predictive capabilities and overlook important factors that could impact lending decisions. The hybrid machine learning approach demonstrated promising performance on the South German Credit dataset, it is essential to acknowledge the constraints under which the study has been performed.

Alternate models to consider for comparison could include traditional statistical methods such as logistic regression or decision trees, as well as more advanced machine learning algorithms like support vector machines (SVM) or gradient boosting machines (GBM). Each of these models has its own strengths and weaknesses, and comparing their performance against the hybrid approach would provide valuable insights into which approach is most effective for credit risk assessment tasks.Deploying the models in a real-world environment is essential to validate the efficacy of the proposed approach. This real-world deployment would involve testing the models on a larger and more diverse dataset, including additional attributes that may be available from third-party sources. It would also involve assessing the models' performance in practical scenarios, such as by integrating them into existing credit lending processes and evaluating their impact on decision-making outcome

REFERENCES

[1]. Assef F, Steiner MT, Neto PJS, de Barros Franco DG (2019) Classification algorithms in financial application: credit risk analysis on legal entities. IEEE Lat Am Trans 17(10):1733–1740
[2]. Ben-David A (1995) Monotonicity maintenance in information-theoretic machine learning algorithms. Mach Learn 19(1):29–43
[3]. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
[4]. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
[5]. Central Bank of Brazil (2007) Annual report. https://www.bcb.gov.br/pec/ boletim/banual2007/rel2007p.pdf
[6]. Central Bank of Brazil (2020) Consumer personal loan. https://www.bcb. gov.br/estatisticas/reporttxjuros/

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, February 2024

[7]. Cornée S (2019) The relevance of soft information for predicting small business credit default: Evidence from a social bank. J Small Bus Manag 57(3):699–719

[8]. Crone SF, Finlay S (2012) Instance sampling in credit scoring: an empirical study of sample size and balancing. Int J Forecast 28(1):224–238

[9]. Damrongsakmethee T, Neagoe V (2019) C4.5 decision tree enhanced with AdaBoost versus multilayer perceptron for credit scoring modeling. In: Silhavy R, Silhavy P, Prokopova Z (eds) Computational statistics and mathematical modeling methods in intelligent systems. CoMeSySo 2019. Advances in intelligent systems and computing, vol 1047. Springer, Cham, pp 216–226

[10]. Dastile X, Celik T, Potsane M (2020) Statistical and machine learning models in credit scoring: a systematic literature survey. Appl Soft Comput 106263

[11]. Davis R, Edelman D, Gammerman A (1992) Machine-learning algorithms for credit-card applications. IMA J Manag Math 4(1):43–51

[12]. Feng X, Xiao Z, Zhong B, Dong Y, Qiu J (2019) Dynamic weighted ensemble classification for credit scoring using Markov Chain. Appl Intell 49(2):555–568

[13]. Galindo J, Tamayo P (2000) Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. Comput Econ 15(1/2):107–143

[14]. Kamalloo E, Saniee Abadeh M (2014) Credit risk prediction using fuzzy immune learning. Adv Fuzzy Syst 2014:1–11

[15]. Khandani AE, Kim AJ, Lo AW (2010) Consumer credit-risk models via machine-learning algorithms. J Bank Finance 34(11):2767–2787

[16]. Kozodoi N, Lessmann S, Papakonstantinou K, Gatsoulis Y, Baesens B (2019) A multi-objective approach for profit-driven feature selection in credit scoring. Decis Support Syst 120:106–117

[17]. Lantz B (2013) Machine learning with R. Packt Publishing Ltd, Birmingham

[18]. Lei K, Xie Y, Zhong S, Dai J, Yang M, Shen Y (2019) Generative adversarial fusion network for class imbalance credit scoring. Neural Comput Appl pp 1–12

[19]. Li W, Ding S, Chen Y, Wang H, Yang S (2019) Transfer learning-based default prediction model for consumer credit in China. J Supercomput 75(2):862–884

[20]. Luo C (2020) A comprehensive decision support approach for credit scor- ing. Ind Manag Data Syst 120(2):280–290

[21]. Morales EA, Ramos BM, Aguirre JA, Sanchez DM (2019) Credit risk analysis model in microfinance institutions in Peru through the use of Bayesian networks. In: 2019 Congreso Internacional de Innovación y Tendencias en Ingenieria (CONIITI), IEEE, pp 1–4

[22]. Moula FE, Guotai C, Abedin MZ (2017) Credit default prediction mode- ling: an application of support vector machine. Risk Manag 19(2):158–187

[23]. Niklis D, Doumpos M, Zopounidis C (2014) Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. Appl Math Comput 234:69–81

[24]. Oreski S, Oreski G (2014) Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Syst Appl 41(4):2052–2064

[25]. Anupam Khan, Soumya K. Ghosh (2023) Machine assistance for credit approval? Random wheel can recommend and explainVolume 215, 1 April 2023, 119231

[26]. https://in.mathworks.com/help/matlab/ref/varargin.html

[27]. Shen KY, Sakai H, Tzeng GH (2019) Comparing two novel hybrid MRDM approaches to consumer credit scoring under uncertainty and fuzzy judgments. Int J Fuzzy Syst 21(1):194–212

[28]. Vieira J, Barboza F, Sobreiro VA, Kimura H (2019) Machine learning models for credit analysis improvements: predicting low-income families' default. Appl Soft Comput83(105):640

[29]. Pławiak P, Abdar M, Pławiak J, Makarenkov V, Acharya UR (2020) DGHNL: a new deep genetic hierarchical network of learners for prediction of credit scoring. Inf Sci 516:401–418