International Journal on Applications in Information and communication Engineering
Volume 6 : Issue 4 : December 2020, pp 12 – 20 www.aetsjournal.com ISSN (Online) : 2394 – 6237

CREDIT RISK PREDICTION MODEL USING MACHINE LEARNING ALGORITHMS

KANIMOZHI . P , SIVAGAMI.S , ARUN BALAJI.S , NIRANJANI.K

Abstract— Deep learning has achieved great importance in exhibiting good performance in diverse fields like computer vision and image recognition. This study explores integrating deep learning techniques into consume the credit scoring, a important component of credit risk management in the rapidly expanding consumer finance industry. Various data mining approaches have been proposed and deployed, with a focus on gaining the advantages of deep learning for enhanced credit scoring accuracy. The paper aims to use the potential of these techniques, aligning them with the evolving landscape of credit risk assessment. Credit risk analysis is used to discriminate between good and bad applicants, which will help to reduce the bank loss. Through a comparative analysis, this research paper involves the utilization of diverse data sources, involving financial indicators, customer behavior patterns, and macroeconomic factors. Based on the analysis of credit risk prediction model, can able to save our Indian economy from the huge economical crisis, which will otherwise negatively impact on our Indian economy. This study suggests an better predictive accuracy algorithm by analyzing the various machine learning algorithms like SVM, Naïve Bayes, KNN and ANN. The findings suggest that deploying the ANN with sampling technique enhances the performance of the credit risk prediction model. The ANN, using this sampling approach, demonstrates an impressive accuracy of 97%. This study provides valuable insights into the effective utilization of neural networks and neurons within the proposed algorithm, shedding light on their integrating relationship with sampling techniques and offering a deeper understanding of their functioning.

I. INTRODUCTION

Credit default is the word used to describe the failure of a borrower to meet their contractual obligations, specifically the repayment of a debt or loan as mentioned in the loan agreement. This failure can manifest in various ways, ranging from missed or late payments to a complete indebtedness to repay the borrowed amount.

The implications of credit defaults are significant for both borrowers and lenders alike. When a borrower defaults, it signals either an inability or unwillingness to fulfill their financial commitments, thereby exposing the creditor to potential financial risks. For creditors, the aftermath often involves damage to their credit scores, creating hurdles in securing loans in the future. On the flip side, lenders—especially financial institutions like banks—face tangible financial losses and a negative impact on their overall financial health/liablity.

To manage and mitigate the risks associated with credit defaults, creditors employ various strategies. This includes entire credit assessments to evaluate the borrower's creditworthiness, setting collateral requirements to secure the loan, and deploying robust risk management practices. These measures are important for maintaining a balanced lending portfolio and safeguarding the financial stability of both creditors and debtors in the dynamic landscape of financial transactions.

Non-Performing Assets (NPAs) are loans or advances that have stopped generating profit for the lender because of the borrower's default. In the context of banks, NPAs are loans on which the interest or principal repayments remain overdue for a certain period, typically 90 days or more. When a loan becomes an NPA, the bank or financial institution may face hurdles in recovering the funds, and it can have crucial implications for its financial stability.

The classification of assets into performing and non-performing categories is important for assessing the health of a financial institution's loan portfolio. The higher the proportion of NPAs in a bank's portfolio, the higher the financial strain on the institution. This is because NPAs represent

Kanimozhi.P, Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India
Sivagami.S, Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India
Arun balaji.S, Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India
Niranjani.K Faculty of Information Technology, Dhanalakshmi Srinivasan Engineering College, Tamil Nadu, India

funds that are tied up and not earning income, impacting the profitability for the creditor.

Regulatory bodies, such as in the case of Indiathe Reserve Bank of India (RBI), set guidelines for the classification and management of NPAs. Financial institutions are required to regularly report their NPA levels and take measures to address and reduce non-performing assets via recovery mechanisms, restructuring, or other remedial measures.

In summary, credit default occurs when a borrower fails to repay a debt, and NPAs are loans that have become non-performing due to failure of non payment of the debtor. Managing and mitigating credit defaults and NPAs are crucial for the stability and health of financial institutions and the overall financial system.

A. Impact of credit default:

1) Non-Performing Assets (NPAs):

One of the most significant negative impacts of credit risk is the addtion of Non-Performing Assets in the banking sector. When lendors default on their loans, it leads to financial stress for banks, affecting their ability to lend further and disrupting the stability of the country's economy.

2)Financial Instability:

Excessive credit risk, especially when not adequately managed, can lead to economic instability. If a higher number of borrowers default simultaneously, it can lead to domino effect, affecting the entire economy. This instability can have severe repercussions on the overall economic health.

3)Interest Rates and Inflation:

To compensate for potential credit losses, creditors may increase interest rates on loans. Higher interest rates can disrupt borrowing and spending, impacting economic growth. Additionally, inflation can be fueled if credit is extended increasingly, leading to an overheated economy.

B. Mitigation Strategies:

1. Effective Risk Management: Financial institutions must implement robust risk management practices, including thorough credit assessments, monitoring of lender's financial health, and establishing risk mitigation strategies.

2. Regulatory Oversight: Government and regulatory bodies play a crucial role in ensuring that financial institutions adhere to prudent lending practices. Stringent regulations help in maintaining the stability of the financial system.

3. Promoting Financial Literacy: Educating borrowers about responsible borrowing and financial management is important. Informed borrowers are more likely to honor their financial commitments, decreasing the overall credit risk.

Mitigating credit defaults is vital for lessening their adverse effects on the Indian economy. This effective credit prediction model helps to address credit defaults and it can contribute to maintaining financial stability and minimizing economic repercussions. Proactive measures to reduce the impact of credit defaults are essential for sustaining a healthy economic environment in India.

II. RELATED WORKS:

Galindo et.al [1], Renewed interest in financial intermediary risk assessment, driven by past crises, emphasizes the need for accurate risk estimation in credit portfolios. Our study compares over 9,000 statistical and machine learning models on a mortgage loan dataset. CART decision-tree models deploy superior default estimation with an 8.31% error rate (2,000 records), potentially reaching 7.32% with more data. Neural Networks follow with an 11.00% average error, outperforming K-Nearest Neighbor (14.95%). These findings suggest that, the potential integration of precise predictive models in institutional and global risk frameworks.

Ma et.al [2], Scientific and technological progress is reshaping consumer finance, yielding new financial credit risks, notably in the evolving realm of Internet finance. The MLIA algorithm, an advanced machine learning approach, decomposes the objective function into weighted sums of basis functions. Comparative analysis against logistic prediction algorithms revealincreased performance by MLIA across three test functions. The study, assessing MLIA's financial credit risk prediction model with AUC, underscores machine learning's efficacy in forecasting financial risks, providing crucial insights for future research in this field.

Clements et.al [3], Machine learning is important in minimizing banking losses by predicting potentially billion-dollar credit risks. Despite the dominance of gradient boosted decision tree models, their efficacy levels off without expensive new data or features. This paper presents an innovative deep learning method for credit risk assessment, using a distinctive credit card transaction sampling technique with deep recurrent and causal convolution-based neural networks. The sequential approach, especially a temporal convolutional network, outperforms non-sequential tree-based models, providing substantial financial savings, early risk detection, and adaptability for efficient online learning in production environments.

Veeramanikandan et.al [4], Financial Credit Scoring (CS) is a crucial area of finance research, crucial for evaluating the creditworthiness of individuals and organizations. Data mining proves beneficial in banking, facilitating the creation of risk-optimized products and services. This study introduces an advanced Deep Neural Network (DNN) framework with Stacked Autoencoders (SA) for precise credit score prediction. SA extracts features, classified via a SoftMax layer, and the network is fine-tuned with Truncated Backpropagation Through Time (TBPTT) in a supervised approach. The SADNN model attains outstanding performance on a German credit dataset, with a 96.10% accuracy, 97.25% F-score, and a kappa value of 90.52%, showcasing its efficacy in predicting credit scores and assessing borrowers' repayment capabilities.

Bhatore et.al [5], Credit risk, arising from borrower default, requires careful credit assessment during loan allocation and ongoing monitoring to mitigate non-performing assets (NPA) and fraud risks. The escalation of NPAs and fraud incidents accelerates the necessity for robust predictive mechanisms in evaluating loan performance. Over the past two decades, artificial intelligence, particularly machine learning (ML), has advanced, enhancing tools for credit risk assessment. Analyzing 136 papers from 1993 to March 2019, the survey highlights aincreasing preference for Ensemble and Hybrid models, incorporating neural networks and SVM, for credit scoring, NPA prediction, and fraud detection, while acknowledging the persistent challenge of limited comprehensive public datasets in this research domain.

Khandani et.al [6], Employing machine-learning techniques, we develop nonlinear forecasting models for consumer credit risk, increasing customer transactions and credit bureau data from January 2005 to April 2009. Out-of-sample forecasts crucially improve credit-card-holder delinquency and default classification rates, yielding an 85% linear regression R2 for forecasted/realized delinquencies. Conservative estimates on adjusting credit lines based on machine-learning predictions indicate potential cost savings of 6% to 25% of total losses. Furthermore, time-series patterns of estimated delinquency rates during the recent financial crisis suggest the applicability of aggregated consumer-credit risk analytics for forecasting systemic risk.

Li et.al [7], Credit scoring, vital for assessing customers' repayment capability, is a key concern for loan institutions. The rapid advancement of machine learning techniques has introduced various classification methods to characterize customer repayment behavior. In this study, we deploy the XGBoost algorithm to distinguish between good and bad customers based on their repayment habits. A comprehensive analysis compares the performance of XGBoost with logistic regression, revealing superior results for the XGBoost algorithm in identifying customers who fail to repay.Kruppa et.al [8], Consumer credit scoring, notably a binary classification task, is enhanced by machine learning methods that estimate detailed default probabilities. Logistic regression is commonly used for this, but our study introduces a comprehensive framework utilizing nonparametric regression approaches such as random forests (RF), k-nearest neighbors (kNN), and bagged k-nearest neighbors (bNN). Applying these algorithms to a large dataset of short-term installment credit histories, we showcase the superiority of regression RF over optimized logistic regression, kNN, and bNN in probability estimation and credit risk assessment. Our approach employs Random Jungle, a C++ RF package with a specialized framework for efficient tree growth, probability estimation, and classification.

Bao et.al [9], Credit scoring, essential for credit risk assessment, distinguishes "bad" from "good" applicants in financial institutions. While supervised machine learning is generally used, the integration of unsupervised learning is often overlooked. This study proposes a novel strategy, combining unsupervised and supervised learning at two stages: the consensus and dataset clustering stages. Comparing model performance across credit datasets, integration at either stage improves credit scoring models. Notably, combining both stages achieves the best performance, affirming the superiority of the proposed integration in enhancing credit risk assessment with confidence for extension to diverse financial datasets.

Addo et.al [10], Leveraging advanced technology and Big Data, banks are reshaping their business models, emphasizing credit risk predictions, model reliability, and efficient loan processing. In this study, binary classifiers are developed using machine and deep learning models on real data to predict loan default probability. The top 10 important features are identified, and their stability is tested by evaluating performance on separate data sets. The findings reveal that tree-based models exhibit greater stability compared to multilayer artificial neural networks, prompting considerations regarding the widespread adoption of deep learning systems in enterprises.

III. METHODOLOGY:

In machine learning (ML), classification techniques are algorithms designed to categorize or label data into distinct classes or groups based on some patterns and features. These techniques are fundamental for tasks where the goal is to predict the category of a given input. Here we used machine learning algorithms for credit prediction model.

1) Logistic Regression:

For binary classification problems, a linear model used, to predicting outcomes as either 0 or 1.

2) Decision Trees:

Tree-like models that make decisions based on features, branching to different outcomes, suitable for bothmulti-class classification andbinary.

3)Random Forest:

An ensemble method that builds multiple decision trees and combines their predictions to enhance accuracy and reduce overfitting.

4)Support Vector Machines (SVM):

This supervised machine learning algorithm used to find a hyperplane to separate data into different classes, effective in high-dimensional spaces.

5)Naive Bayes:

A probabilistic algorithm based on Bayes' theorem that is particularly useful for classification based problems.

6)K-Nearest Neighbors (KNN):

A simple ML algorithm that classifies data points based on the majority class of their k-nearest neighbors.

7)Neural Networks:

Deep learning models composed of layers of interconnected nodes (neurons) that can be used for complex classification based tasks.

The choice of classification technique based on factors such as the nature of the data, the problem at hand, and the desired model interpretability. It's general to experiment with multiple algorithms and fine-tune parameters to achieve optimal performance for a given classification task.

8)Logistic Regression:

Random Forest is an ensemble learning method in machine learning that deploy a multitude of decision trees during training. It introduces diversity by selecting random subsets of both data instances and features for each tree, decreasing overfitting. Each tree independently makes predictions, and the final output is determined through majority voting for classification or averaging for regression. Random Forest excels in capturing complex relationships in high-dimensional datasets and is robust to noisy data. It gives insights into feature

International Journal on Applications in Information and communication Engineering
Volume 6 : Issue 4 : December 2020, pp 12 – 20 www.aetsjournal.com ISSN (Online) : 2394 – 6237

importance, aiding interpretability. The algorithm's versatility and effectiveness make it widely used across diverse domains, from finance,education to healthcare. The combination of decision trees in an ensemble enhances predictive accuracy and generalization, making Random Forest a powerful and popular tool for various real-world applications.

Logistic Regression employs the sigmoid function (also known as the logistic function) to transform a linear combination of input features into a value between 0 and

$$f(x) = \frac{1}{1+e^{-(x)}}$$

The linear combination is expressed as $Z=b_{0+}b_1x_{1+}b_2x_{2+} \ . \ . \ . \ +b_nx_{n,}$ where b represents coefficients and x represents features.

The sigmoid function transforms the linear combination into a probability score.

The predicted probability P(Y=1) is given by $\sigma(z)$, and P(Y=0) = 1-P(Y=1).

A decision threshold (commonly 0.5) is chosen. Instances with predicted probabilities above this threshold are classified as class 1, and those below as class 0.

Logistic Regression is favored in credit risk prediction for its simplicity, interpretability, and efficiency. It provides a clear understanding of the factors influencing credit risk and offers a transparent decision-making process, important in the financial domain.

9)Decision tree:

The Decision Tree algorithm is a crucial tool in machine learning, for used at handling both classification and regression tasks. Operating through recursive data partitioning, it creates a tree structure with decision nodes representing features and their conditions. Terminal nodes, or leaf nodes, hold final predicted outcomes. Splitting criteria, such as Gini impurity for classification, guide the algorithm in determining optimal feature conditions. The training process starts with the root node, recursively splitting data until stopping conditions, like a maximum tree depth, are met. Decision Trees offer interpretability and make no strict assumptions about data distribution. However, they are prone to overfitting and exhibit instability with small data alteration. Common use cases span finance, healthcare, and marketing for tasks like credit scoring and disease diagnosis. Techniques like regularization and ensemble methods, such as Random Forests, are employed to reducing the overfitting concerns.

Decision Trees are important in credit risk prediction as they facilitate transparency, allowing financial institutions to explain credit decisions to applicants. However, for addressing overfitting, techniques like pruning and ensemble methods are often employed.

10) Random forest:

Random Forest is an ensemble learning algorithm that enhances the strength of multiple decision trees to increase predictive performance. During training, it constructs an ensemble of decision trees by randomly selecting subsets of the training data and features. Each tree independently makes a prediction, and the final output is determined by aggregating the results, either through voting for classification or regression. This ensemble approach address overfitting and increases robustness compared to individual decision trees. Random Forest is resilient to outliers and noise, making it suitable for complex, high-dimensional datasets. The algorithm's versatility extends to handling missing values and maintaining interpretability. Random Forest's widespread use spans various domains, including finance, healthcare, and image recognition, where its ability to balance accuracy and generalization proves advantageous.

The Random Forest algorithm's ability to handle non-linear relationships, manage missing data, and provide interpretable insights into feature importance makes it well-suited for credit risk prediction. By leveraging the strengths of ensemble learning, Random Forest increases the robustness and reliability of credit risk assessments in real-world scenarios.

11) Support vector machine:

Support Vector Machines (SVM) is a valuable supervised learning algorithm used for both classification and regression based tasks. The key objective is to find a hyperplane that best separates data points into different classes while maximizing the margin between those different classes. SVM works effectively in high-dimensional spaces, making it important for complex datasets. It relies on support vectors, which are data points closest to the decision boundary, to define the optimal hyperplane. SVM is versatile, capable of handling linear and non-linear relationships through kernel functions. The algorithm is robust to overfitting, and its decision boundary is influenced by a subset of critical data points. SVM finds applications in various fields, including image recognition, text classification, image classification and bioinformatics, where its ability to handle intricate data distributions is particularly valuable.

SVMs are crucial in credit risk prediction due to their ability to handle non-linear relationships and high-dimensional feature spaces. The algorithm provides a flexible decision boundary, making it suitable for scenarios where the decision boundary is complex and not easily represented by a linear model. However, parameter tuning and careful consideration of kernel selection are important to optimizing SVM performance in credit risk prediction tasks.

12)Top of Form Naïve bayes:

Naive Bayes, a probabilistic classification algorithm based on Bayes' theorem, assumes feature independence given the class, simplifying computations. It calculates class probabilities for a given set of features, selecting the one with the highest probability. Despite its "naive" assumption, Naive Bayes excels in tasks like text classification, spam filtering, and sentiment analysis. Known for minimal training data requirementsandcomputational efficiency, it is robust to irrelevant features but may struggle with complex relationships. Widely embraced for simplicity and ease of implementation, Naive Bayes proves effective in scenarios where the assumption of feature independence holds reasonably well.

Naive Bayes is specifically useful when dealing with a higher number of features and assuming independence among them simplifies the modeling process. Despite its "naive" assumption, Naive Bayes often performs good in credit risk prediction scenarios, providing a computationally efficient and interpretable solution. It is particularly valuable when computational resources are a concern and a quick, reliable model is needed.

13) KNN:

K-Nearest Neighbors (KNN), as a straightforward yet potent machine learning algorithm applicable to classification and regression tasks. Its classification process relies on determining the majority class among the k-nearest neighbors in the feature space for a given instance. Built on the premise that similar instances share similar classes, the choice of 'k' shapes the algorithm's sensitivity to local patterns. Despite its robustness and ease of implementation, KNN may lead to computational challenges with large datasets. It excels in scenarios featuring non-linear decision boundaries and clustered data. The versatility of KNN finds applications across recommendation systems, image recognition, and anomaly detection, showcasing its adaptability to diverse data distributions and problem domains.

14) Neural Networks:

Neural networks, inspired by the structure and function of the human brain, are a category of machine learning models. These networks comprise interconnected nodes arranged in layers, encompassing an input layer, one or more hidden layers, and an output layer. Their application spans various domains, with finance being a notable example where they are deployed for tasks like credit prediction.

When it comes to predicting credit, neural networks can be employed for credit scoring, which includes assessing the creditworthiness of individuals or entities based on their historical financial behavior.

It's important to note that the success of a neural network in credit prediction depends on the quality and relevance of the data, the appropriateness of the chosen features, and the fine-tuning of the model during training. Additionally, interpretability and explainability of the model's decisions are important in financial applications, so understanding the factors influencing the credit decision is crucial.

International Journal on Applications in Information and communication Engineering
Volume 6 : Issue 4 : December 2020, pp 12 – 20 www.aetsjournal.com ISSN (Online) : 2394 – 6237

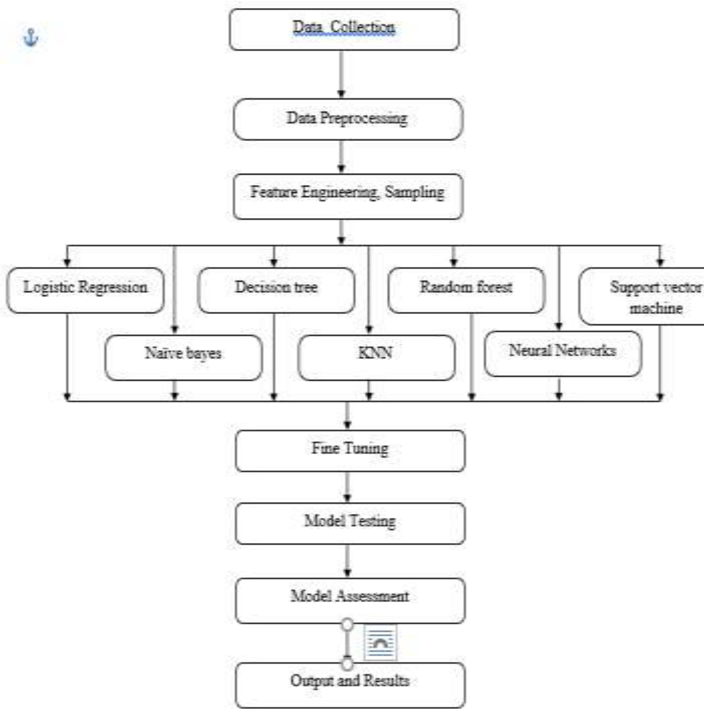
IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Architecture diagram:

Certainly! Developing a credit prediction model involves several key steps. Here's an explanation of each step:

1) Data Collection:

- Gather relevant data for building the credit risk prediction model. This data may include information about applicants, their financial history, employment status, debt, and other relevant factors. Sources might include credit bureaus, financial institutions, or other databases.



2)Data Preprocessing:

- Clean and organize the collected data. This involves handling missing values, removing duplicates, and dealing with noise. Ensure that the data is in a format suitable for analysis, addressing issues like data types and formatting.

3)Feature Engineering:

- Identify and create relevant features that can help improve the predictive power of the model. This may involve transforming existing variables, creating new variables, or combining features to extract more meaningful information.

4) Sampling:

- Based on the size of the dataset, consider sampling techniques to create a representative subset for model training and testing. Common methods include random sampling or stratified sampling to ensure a balanced representation of classes.

5)Machine Learning Classification Algorithms:

- Choose appropriate machine learning algorithms for credit prediction. Some common algorithms for classification tasks include logistic regression, decision trees, random forests, support vector machines, and neural networks. Train these algorithms on the training dataset.

6)Fine Tuning:

- Optimize the model's hyperparameters to achieve better performance. This involves addressing settings such as learning rates, regularization parameters, and tree depths to enhance the model's ability to generalize well to new, unseen data.

7)Model Testing:

- Evaluate the model's performance on a separate testing dataset that it has not seen earlier. This helps ensure that the model can make accurate predictions on new, unseen data. Assess metrics such as accuracy, precision, recall, and F1 score to gauge performance.

8)Model Assessment:

- Analyze the model's strengths and weaknesses. Consider factors such as interpretability, computational efficiency, and the model's ability to handle different scenarios. This step may involve iterating on the model architecture or fine-tuning based on the assessment results.

9) Output and Results:

- Once the model is deemed satisfactory, use it to make predictions on new credit applications. The output can include a probability score or a binary decision indicating whether an applicant is likely to be a credit risk. Communicate the results effectively, and, if necessary, deploy the model in a real-world setting.

It's important to note that these steps are iterative, and the process may involve going back and forth between them as you discover insights, refine the model, and improve its performance.

B. Results and Discussion:

Algorithm	Training accuracy	Testing accuracy
Logistic Regression	82.56	86.89
Decision Trees	79.83	85.65
Random Forest	86.98	88.87
Support Vector Machine	90.34	91.98
Naïve Bayes	92.34	92.50
KNN	89.56	93.23
Neural Network	**96.56**	**97.30**

Algorithm	Training Loss	Testing Loss
Logistic Regression	0.583	0.5638
Decision Trees	1.2356	1.5825
Random Forest	2.4654	2.5487
Support Vector Machine	0.5498	0.5684
Naïve Bayes	0.2464	0.2354
KNN	3.1542	3.4648
Neural Network	**2.5689**	**2.5684**

In this model , we have used naïve bayes, logistic regression, decision tree, random forest, SVM, KNN, neural network algorithm with feature engineering and sampling technique for credit prediction model. We took accuracy and loss for measuring the performance of this model. Compare with above algorithms, neural network achieves a 96.56 training accuracy and testing accuracy 97.3.

V. Conclusion:

This study investigates credit risk prediction through the application of various machine learning and deep learning techniques. A comprehensive analysis of various machine learning algorithms reveals that the neural networks model consistently outperforms other models, demonstrating superior accuracy. This observation underscores the neural network model's effectiveness in increasing the identification capabilities of individual fast credit risk. In conclusion, the findings strongly support the assertion that the neural network model stands out as the most suitable and effective choice for developing a credit prediction model.

REFERENCES

[1] Galindo, Jorge, and Pablo Tamayo. "Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications." Computational economics 15 (2000): 107-143.

[2] Ma, Xiaomeng, and Shuliang Lv. "Financial credit risk prediction in internet finance driven by machine learning." Neural Computing and Applications 31 (2019): 8359-8367.

[3] Clements, Jillian M., et al. "Sequential deep learning for credit risk monitoring with tabular financial data." arXiv preprint arXiv:2012.15330 (2020).

[4] Veeramanikandan, Varadharajan, and Mohan Jeyakarthic. "Parameter-tuned deep learning model for credit risk assessment and scoring applications." Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science) 14.9 (2021): 2958-2968.

[5] Bhatore, Siddharth, Lalit Mohan, and Y. Raghu Reddy. "Machine learning techniques for credit risk evaluation: a systematic literature review." Journal of Banking and Financial Technology 4 (2020): 111-138.

[6] Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo. "Consumer credit-risk models via machine-learning algorithms." Journal of Banking & Finance 34.11 (2010): 2767-2787.

[7] Li, Yu. "Credit risk prediction based on machine learning methods." 2019 14th International Conference on Computer Science & Education (ICCSE). IEEE, 2019.

[8] Kruppa, Jochen, et al. "Consumer credit risk: Individual probability estimates using machine learning." Expert systems with applications 40.13 (2013): 5125-5131.

[9] Bao, Wang, Ning Lianju, and Kong Yue. "Integration of unsupervised and supervised machine learning algorithms for credit risk assessment." Expert Systems with Applications 128 (2019): 301-315.

[10] Addo, Peter Martey, Dominique Guegan, and Bertrand Hassani. "Credit risk analysis using machine and deep learning models." Risks 6.2 (2018): 38.

[11] Kumar, Ashwani, D. Shanthi, and Pronaya Bhattacharya. "Credit Score Prediction System using deep learning and K-means algorithms." Journal of Physics: Conference Series. Vol. 1998. No. 1. IOP Publishing, 2021.

[12] Addo, Peter Martey, Dominique Guegan, and Bertrand Hassani. "Credit risk analysis using machine and deep learning models." Risks 6.2 (2018): 38.

[13] Du, Peng, and Hong Shu. "Exploration of financial market credit scoring and risk management and prediction using deep learning and bionic algorithm." Journal of Global Information Management (JGIM) 30.9 (2022): 1-29.

[14] Wu, Hsu-Che, Ya-Han Hu, and Yen-Hao Huang. "Two-stage credit rating prediction using machine learning techniques." Kybernetes 43.7 (2014): 1098-1113.

[15] Mahbobi, Mohammad, Salman Kimiagari, and Marriappan Vasudevan. "Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks." Annals of Operations Research (2021): 1-29.

[16] Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo. "Consumer credit-risk models via machine-learning algorithms." Journal of Banking & Finance 34.11 (2010): 2767-2787.

[17] Ha, Van-Sang, et al. "Improving credit risk prediction in online peer-to-peer (P2P) lending using feature selection with deep learning." 2019 21st International Conference on Advanced Communication Technology (ICACT). IEEE, 2019.

[18] Sayjadah, Yashna, et al. "Credit card default prediction using machine learning techniques." 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA). IEEE, 2018.

[19] Trivedi, Shrawan Kumar. "A study on credit scoring modeling with different feature selection and machine learning approaches." Technology in Society 63 (2020): 101413.

[20] Ala'raj, Maher, et al. "A deep learning model for behavioural credit scoring in banks." Neural Computing and Applications (2022): 1-28.