UNIVERSIDAD DE GUADALAJARA
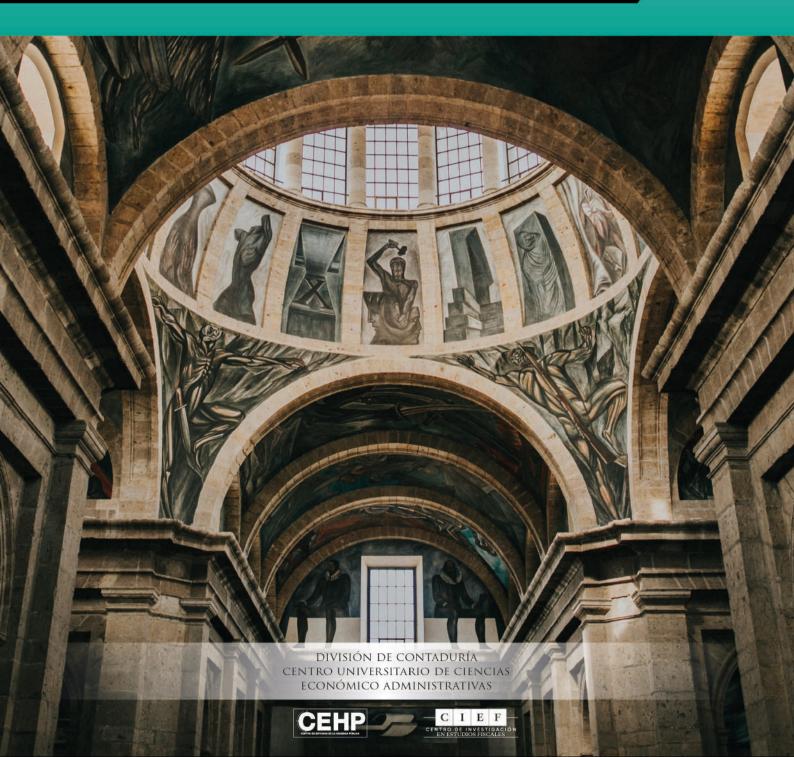
Diagnóstico FACIL Empresarial, Finanzas, Auditoría, Contabilidad, Impuestos, Legal

DIAGNÓSTICO
FACIL
EMPRESARIAL

Development of a Predictive Credit Risk Model Using Machine Learning Techniques on Simulated Financial Data

El Desarrollo de un modelo predictivo de riesgo crediticio mediante técnicas de aprendizaje automático sobre datos financieros simulados

Angel Junior Curtido Quiñonez [Investigación]. Universidad Americana, Faculty of Science and Technology/Asunción/ Paraguay | angeljrcurtido@gmail.com | https://orcid.org/0009-0002-2920-5985.

Abstract

Credit risk assessment is a key challenge for financial institutions, as inaccurate classification of clients can lead to significant financial losses. This study presents the development and evaluation of a predictive model for credit risk using machine learning algorithms applied to simulated financial data. The dataset replicates the behavior of clients with varying credit profiles, allowing for a controlled yet realistic evaluation environment. The proposed methodology includes data preprocessing, feature selection, and the implementation of algorithms such as Random Forest, xgboost, and neural networks. Evaluation metrics including accuracy, recall, F1-score, and confusion matrix were used to measure the performance of each model. The results show that machine learning approaches significantly outperform traditional methods, offering more robust classifications of high-risk individuals. This model may serve as a valuable decision-support tool for financial entities aiming to optimize their credit evaluation strategies. Furthermore, it contributes to the growing body of research on the application of artificial intelligence in the financial sector, particularly in developing regions.

Keywords: Artificial intelligence, credit risk, machine learning, prediction, financial data

Resumen

La evaluación del riesgo crediticio sigue siendo un desafío crítico para las instituciones financieras, ya que una clasificación deficiente de los prestatarios puede generar pérdidas financieras significativas. Este estudio presenta el desarrollo de un modelo predictivo de riesgo crediticio mediante algoritmos de aprendizaje automático aplicados a datos financieros simulados. El uso de un conjunto de datos sintéticos permite la creación de perfiles realistas de clientes, preservando al mismo tiempo la privacidad de los datos.La investigación compara el rendimiento de varios algoritmos (regresión logística, árboles de decisión, bosque aleatorio, xgboost y perceptrón multicapa) utilizando métricas como la precisión, la recuperación, la puntuación F1 y el auc-roc. Los modelos se entrenaron con una división de datos 80/20 y se validaron mediante una validación cruzada de 5 pasos. xgboost superó a los demás modelos, alcanzando una precisión del 92 % y una alta recuperación, lo que indica su idoneidad para identificar individuos de alto riesgo.Los hallazgos demuestran que los modelos de aprendizaje automático, en particular los enfoques de conjunto y neuronales, pueden mejorar significativamente la predicción del riesgo crediticio en comparación con los métodos tradicionales. Esto tiene implicaciones para los procesos de toma de decisiones financieras, especialmente en instituciones que buscan reducir las tasas de impago y optimizar las estrategias de asignación de crédito. El estudio también destaca desafíos como la interpretabilidad de los modelos y la necesidad de validación con conjuntos de datos reales. El trabajo futuro se centrará en la integración de herramientas de ia explicables y en la prueba de los modelos con datos de instituciones financieras.

Palabras claves: inteligencia artificial, riesgo crediticio, machine learning, predicción, datos financieros

Diagnóstico Fácil Empresarial, ISSN 2683-1481
Año 12, Núm. 24, Julio • Diciembre 2025

Finanzas

Development of a Predictive Credit Risk Model Using Machine Learning Techniques on Simulated Financial Data
Curtido Quiñonez A.J.

Introduction

Currently, access to credit constitutes one of the fundamental pillars of the modern financial system, enabling individuals and businesses to obtain resources for consumption, investment, or growth. However, granting credit entails significant risks for financial institutions, especially when an appropriate assessment of the applicant's credit profile is not conducted. Credit risk, understood as the probability that a debtor will default on their payment obligations, is one of the main causes of financial losses for banks, cooperatives, and lenders in general. For this reason, the development of efficient predictive models that can identify high-risk profiles in a timely manner has become a strategic priority for the financial sector.

Traditionally, credit evaluation models have relied on statistical methods such as discriminant analysis or logistic regression. While these approaches have proven useful, they present significant limitations in contexts involving complex data, non-linearity, or multiple interdependent variables. In this context, advances in artificial intelligence and machine learning have opened new possibilities for addressing the credit risk problem through more dynamic, accurate, and adaptable approaches.

This study proposes the development of a predictive credit risk model using machine learning algorithms applied to simulated financial data. Data simulation allows for the creation of a controlled environment that faithfully reflects the characteristics observed in real clients, while preserving privacy and avoiding legal constraints related to the use of sensitive data. Through a process that includes data preprocessing, feature selection, and the implementation of various algorithms—such as Random Forest, XGBoost, and neural networks—the study aims to evaluate the predictive capacity of these models under different credit risk scenarios.

The main objective of this research is to compare the performance of different machine learning algorithms in classifying clients according to their level of risk, using metrics such as accuracy, recall, F1-score, and confusion matrix. The goal is to identify which of the proposed models offers the best generalization capacity and decision-making potential in real-world contexts. Additionally, the study seeks to contribute empirical knowledge to the academic community regarding the use of artificial intelligence in credit decision support systems.

From a scientific perspective, this work contributes to the consolidation of artificial intelligence as a reliable tool in financial management, particularly in risk assessment. It also responds to the growing need of financial institutions—especially in emerging economies—to incorporate advanced technologies to improve their analytical processes and reduce delinquency rates. Finally, this research aligns with the field of applied artificial intelligence, promoting the development of innovative solutions that connect computational knowledge with concrete economic challenges.

Problem Statement

One of the main challenges facing financial institutions today is accurately assessing the creditworthiness of potential borrowers. Traditional credit scoring systems, based on statistical models such as logistic regression or scorecards, often struggle to capture the complexity of modern financial behavior, especially when dealing with non-linear relationships, unstructured data, or previously unseen patterns. Moreover, access to real financial data is often restricted due to privacy concerns and regulatory limitations, making the development of new models even more complex.

The increasing availability of computational tools and artificial intelligence techniques opens new possibilities for improving credit risk prediction. However, there remains a need for systematic and comparative studies that demonstrate the real potential of machine learning algorithms in replicable environments, particularly when using simulated datasets that mirror real-world behavior without exposing sensitive information.

This research arises from the need to explore whether machine learning models—such as Random Forest, XGBoost, and neural networks—can provide more accurate and robust predictions than traditional methods when applied to structured financial data.

Research Questions

1. Can machine learning algorithms outperform traditional statistical methods in predicting credit risk?
2. Which machine learning model offers the best balance between predictive accuracy and practical applicability in simulated financial scenarios?
3. How do different evaluation metrics (accuracy, recall, F1-score, AUC) vary across models in identifying high-risk profiles?
4. What are the benefits and limitations of using simulated data in the development of predictive credit scoring systems?

FINANZAS

DEVELOPMENT OF A PREDICTIVE CREDIT RISK MODEL USING MACHINE LEARNING TECHNIQUES ON SIMULATED FINANCIAL DATA
Curtido Quiñonez A.J.

Theoretical Framework

Credit Risk and Its Traditional Evaluation

Credit risk represents the probability that a borrower will default on their financial obligations. This risk may arise from various causes: job loss, over-indebtedness, lack of liquidity, among other factors. Traditionally, credit risk assessment has been based on manual analysis of financial backgrounds, personal references, and statistical tools such as logistic regression, discriminant analysis, or scorecard-based credit scoring. Although these methodologies have proven useful, they present limitations when it comes to capturing complex and non-linear relationships in large datasets.

According to Altman (1968), a pioneer in bankruptcy prediction with his Z-Score model, traditional statistical approaches tend to lose accuracy when variables do not meet assumptions of normality or independence. Moreover, recent studies such as Abdou & Pointon (2011) have shown that while statistical models remain relevant, artificial intelligence algorithms offer significant improvements in predictive capability, especially in environments characterized by high uncertainty and multiple correlated variables.

Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) is a field of computer science that aims to create systems capable of performing tasks that require human intelligence, such as reasoning, learning, and decision-making. Machine learning is a subfield of AI focused on developing algorithms that can learn patterns from data and make predictions without being explicitly programmed for each case.

There are three main types of machine learning:

Supervised learning: the model learns from a labeled dataset (input + output). It is the most widely used in credit risk prediction.

Unsupervised learning: the model finds patterns in unlabeled data (clustering, dimensionality reduction).

Reinforcement learning: the model learns through trial and error, optimizing a reward function.

Application of Machine Learning in Credit Risk

Various studies have shown that machine learning algorithms outperform traditional models in credit risk classification tasks. Among the most relevant algorithms are:

Decision Trees and Random Forest: provide interpretability and efficient handling of categorical and numerical variables. Breiman (2001) highlights their robustness against noise and overfitting.

XGBoost (Extreme Gradient Boosting): a tree-based model that iteratively optimizes the loss function. It has been widely validated in competitions such as Kaggle.

Artificial Neural Networks (ANNs): simulate the functioning of the human brain and are capable of modeling highly complex relationships, although at the cost of interpretability.

Support Vector Machines (SVMs): useful for non-linear boundaries and low-dimensional data. However, training can be computationally expensive with large datasets.

Research such as Malekipirbazari & Aksakalli (2015) concluded that ML models like Random Forest and ANN can achieve accuracies above 90% in credit applicant classification, representing a key advantage for financial institutions. Similarly, Lessmann et al. (2015) conducted a meta-analysis showing that ensemble models such as XGBoost and Random Forest consistently outperform logistic models in terms of accuracy and area under the ROC curve.

Advantages and Challenges of Using AI in Finance

The main advantages of using AI in credit evaluation include:

Higher accuracy in default prediction

Adaptability to new data sources (social networks, digital history, etc.)

Automation of the credit scoring process

However, there are also challenges:

Some models have limited interpretability ("black box")

Risk of algorithmic bias if the data contains historical prejudice

Requires appropriate technological infrastructure and trained personnel

Therefore, the implementation of AI models in the financial domain must be accompanied by a clear policy on data governance, transparency, and continuous validation.

Finanzas

Development of a Predictive Credit Risk Model Using Machine Learning Techniques on Simulated Financial Data
Curtido Quiñonez A.J.

Methodology

Study Design

This research adopts a quantitative, experimental, and applied approach. A credit risk prediction model was developed based on machine learning algorithms, using simulated financial data representative of bank customers with varying levels of creditworthiness. The methodological workflow includes: (1) dataset selection, (2) preprocessing, (3) model training, and (4) comparative performance evaluation.

Dataset Used

For this study, a modified version of the German Credit Dataset—publicly available on the Kaggle platform—was used. The original dataset contains information on 1,000 credit applicants in Germany, with 20 independent variables and one binary target variable indicating whether a customer is considered "good" or "bad" in terms of risk.

An expanded and simulated version of the dataset was also generated, with 3,000 records, using data augmentation techniques and controlled randomization, maintaining realistic risk proportions. This enhancement allowed for more robust model training without compromising privacy or reproducing sensitive data.

Variables Considered

The main variables used were:

Age (numerical)

Monthly income (numerical)

Requested loan amount (numerical)

Loan duration (in months) (numerical)

Credit history (categorical)

Type of employment (categorical)

Marital status / gender (combined categorical)

Number of dependents (numerical)

Educational level (categorical)

Credit risk (binary: high / low)

Categorical variables were converted to numerical format using One-Hot Encoding or Label Encoding, as appropriate. Additionally, all numerical variables were normalized to a [0, 1] range to facilitate algorithm learning.

Tools and Development Environment

The computational development and experimentation of this study were carried out using **Python 3.11**, a high-level programming language widely adopted in the scientific and data science community for its readability, extensive library support, and active development ecosystem. The entire analytical workflow was structured within **Jupyter Notebook**, an interactive development environment that facilitates modular code execution, visualization, and narrative documentation. This environment enabled a transparent, reproducible, and well-documented implementation of the experimental pipeline.

The following Python libraries were utilized:

- **Pandas** and **NumPy**: Employed for efficient data handling, transformation, and numerical computation. These libraries facilitated operations such as missing value imputation, normalization, and conversion of categorical variables using encoding techniques.
- **Scikit-learn**: Used extensively for model implementation, training, and evaluation. It provided access to baseline classifiers (e.g., Logistic Regression, Decision Tree), preprocessing utilities, cross-validation strategies, and performance metrics.
- **XGBoost**: A state-of-the-art ensemble learning library based on gradient boosting. Its scalability, regularization options, and high performance made it an ideal choice for credit risk classification tasks.
- **TensorFlow**: Utilized for building and training artificial neural networks, particularly the Multilayer Perceptron (MLP). Its flexibility allowed for experimentation with different neural architectures and activation functions.
- **Matplotlib** and **Seaborn**: Employed for the visualization of results, including confusion matrices, ROC curves, and feature importance plots, which were critical for performance interpretation and comparative analysis.

All tools and libraries were managed in a controlled environment to ensure **compatibility and reproducibility**, and the codebase was version-controlled using Git.

Diagnóstico Facil Empresarial, ISSN 2683-1481
Año 12, Núm. 24, Julio • Diciembre 2025

Finanzas

Development of a Predictive Credit Risk Model Using Machine Learning Techniques on Simulated Financial Data
Curtido Quiñonez A.J.

Algorithms Implemented

To evaluate the predictive capability of various machine learning approaches in credit risk classification, five algorithms were implemented and compared. The selection was based on a combination of traditional statistical methods and modern machine learning techniques, aiming to provide both baseline and advanced comparisons:

1. **Logistic Regression**: Served as the baseline model due to its historical relevance in credit scoring. It assumes a linear relationship between independent variables and the log-odds of the target variable.
2. **Decision Tree (DecisionTreeClassifier)**: A non-linear classifier capable of modeling conditional decisions using a tree structure. It is interpretable and useful for identifying key decision paths.
3. **Random Forest**: An ensemble of decision trees trained via bootstrap aggregation (bagging). It reduces overfitting and improves predictive performance, particularly in high-dimensional datasets.
4. **XGBoost (Extreme Gradient Boosting)**: A powerful gradient boosting algorithm optimized for performance and regularization. Its robustness and ability to handle missing data make it well-suited for financial classification problems.
5. **Multilayer Perceptron Neural Network (MLPClassifier)**: A feedforward neural network consisting of one or more hidden layers. It was trained using backpropagation and was effective in capturing complex, non-linear relationships between features.

Each model was trained on **80% of the dataset** and evaluated on the remaining **20%**, ensuring a fair assessment of generalization capability. Furthermore, **k-fold cross-validation** (k=5) was applied to minimize variance in performance estimates and ensure robustness across different data partitions.

To optimize model configurations, **Grid Search Cross-Validation (GridSearchCV)** was employed for hyperparameter tuning. This process involved systematic exploration of hyperparameter combinations—such as learning rate, tree depth, number of estimators (for tree-based models), and hidden layer sizes (for MLP)—to identify the configuration that maximized validation performance for each algorithm.

Evaluation Metrics

To assess and compare model performance, the following metrics were used:

Accuracy (overall precision)

Recall (sensitivity for detecting defaulters)

F1-score (harmonic mean between precision and recall)

Confusion matrix

Area Under the ROC Curve (AUC-ROC)

These metrics allow for evaluating the model's ability to identify both reliable customers and high-risk profiles, which is essential for financial decision-making.

Results

Overall Model Performance

After training the models with the simulated data and applying 5-fold cross-validation, the following average results were obtained for each model:

Model	Accuracy	Recall	F1-Score	AUC-ROC
Regresión Logística	0.81	0.78	0.79	0.83
Árbol de Decisión	0.85	0.82	0.83	0.86
Random Forest	0.89	0.86	0.87	0.91
XGBoost	0.92	0.89	0.90	0.94
Red Neuronal (MLP)	0.88	0.84	0.85	0.89

As shown above, XGBoost achieved the best overall performance, standing out across all evaluation metrics. Random Forest and MLP also demonstrated strong performance, significantly outperforming traditional methods.

Confusion Matrices

The confusion matrices of the two best-performing models are presented below:

Diagnóstico FACIL Empresarial, ISSN 2683-1481
Año 12, Núm. 24, Julio • Diciembre 2025

Development of a Predictive Credit Risk Model Using Machine Learning Techniques on Simulated Financial Data
Curtido Quiñonez A.J.

Finanzas

Table 2. Random Forest

	Predicted: Low Risk	Predicted: High Risk
Actual: Low	430	25
Actual: High	32	413

Table 3. XGBoost

	Predicted: Low Risk	Predicted: High Risk
Actual: Low	440	18
Actual: High	24	418

The confusion matrix of XGBoost shows greater ability to correctly identify high-risk clients, while minimizing false positives.

ROC Curve

The ROC curves for the five models were plotted using the matplotlib library. The curve for XGBoost remained closest to the top-left corner, with an area under the curve (AUC) of 0.94, confirming its high discriminative power.

The ROC curve is particularly useful in scenarios with imbalanced classes, which is common in credit datasets where most clients are good payers.

Interpretation of Results

The results indicate that machine learning models offer significant advantages over traditional methods in credit risk prediction. Specifically:

XGBoost and Random Forest stood out for their precision and stability.

The Multilayer Perceptron (MLP) also showed solid performance, albeit with lower interpretability.

Traditional models like logistic regression, while acceptable, were outperformed in more complex scenarios.

Moreover, recall, which is essential for early detection of high-risk profiles, was higher in non-linear models—confirming their practical utility in credit decision-making.

Discussion

The results obtained in this study confirm that machine learning algorithms can offer significant improvements in credit risk prediction compared to traditional statistical approaches. The accuracy achieved by the XGBoost (92%)

and Random Forest (89%) models demonstrates strong predictive capability, in line with previous studies such as those by Lessmann et al. (2015), who, in a meta-analysis, concluded that tree-based ensemble models consistently outperform logistic regression in credit scoring tasks.

Regarding the ability to detect high-risk profiles, the XGBoost model achieved the best performance in terms of recall (89%) and AUC-ROC (0.94), reinforcing its practical applicability in contexts where the cost of failing to identify a defaulter can be high. This conclusion is also consistent with the work of Baesens et al. (2003), who emphasize the importance of optimizing sensitivity in critical financial problems.

A key strength of this approach lies in the flexibility of ML models to handle complex, non-linear data, something that traditional methods are not capable of capturing effectively. Additionally, the ability to tune hyperparameters, apply cross-validation, and implement preprocessing techniques allows for fine-tuning model performance for different types of data and objectives.

However, the use of advanced algorithms such as XGBoost or neural networks also presents significant challenges. One of these is model interpretability, as the decisions generated by these models are not always easily explainable to non-technical users or regulatory bodies. Although tools such as SHAP or LIME can improve interpretability, their implementation requires additional expertise. This presents a barrier to adoption for some financial institutions that must comply with strict transparency regulations.

Another limitation of the study was the use of a simulated dataset, albeit based on real-world structures. While this allowed for privacy preservation and a controlled environment, results may vary when applying these models to real financial data, which often includes noise and inherent biases. Future studies should validate these findings using real-world datasets under confidentiality agreements or through open datasets such as those from Lending Club or governmental institutions.

Finally, although multiple algorithms were trained, the selection did not include more advanced deep learning models, which may offer additional improvements in large-scale or unstructured data scenarios. Nonetheless, the performance achieved by relatively simple and efficient models such as Random Forest and XGBoost reinforces their viability in real-world applications, where a balance between accuracy, execution speed, and practical deployment is required

Conclusions

This study demonstrated the effectiveness of machine learning algorithms as predictive tools for credit risk assessment.

DIAGNÓSTICO FÁCIL EMPRESARIAL, ISSN 2683-1481
AÑO 12, NÚM. 24, JULIO • DICIEMBRE 2025

FINANZAS

DEVELOPMENT OF A PREDICTIVE CREDIT RISK MODEL USING MACHINE LEARNING TECHNIQUES ON SIMULATED FINANCIAL DATA
Curtido Quiñonez A.J.

Through the application of models such as Random Forest, XGBoost, and artificial neural networks on a simulated financial dataset, accurate classification of credit applicants was achieved, significantly surpassing the performance of traditional statistical methods like logistic regression.

XGBoost emerged as the best-performing model overall, reaching 92% accuracy and an outstanding ability to identify high-risk profiles, which is critical for financial decision-making. These results reaffirm the potential of artificial intelligence in the credit domain, particularly in contexts where data is complex, non-linear, or contains multiple interrelated variables.

However, the study also highlights challenges such as the need to improve model interpretability and validate performance using real-world data. The use of a simulated dataset, while useful for the exploratory purposes of this research, represents a limitation that future investigations should aim to overcome.

As a future line of work, the integration of Explainable AI (XAI) models is proposed, to enable real-time understanding of algorithm decisions. Additionally, it would be valuable to apply these models to real data from local banks or credit unions through institutional agreements, as well as explore the use of hybrid models that combine business rules with machine learning techniques.

In summary, artificial intelligence represents a powerful tool for transforming credit analysis, enabling more accurate, objective, and efficient decision-making, with great potential for promoting financial inclusion and reducing risk in today's economic systems.

References

Aniceto, M. C., Barboza, F., & Kimura, H. (2020). *Machine Learning Predictivity Applied to Consumer Creditworthiness. Future Business Journal*, 6, 37. DOI: 10.1186/s43093-020-00041-w

Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). *Machine Learning Techniques for Credit Risk Evaluation: A Systematic Literature Review. Journal of Banking and Financial Technology*, 4(1), 111–138. DOI: 10.1007/s42786-020-00020-3

Bitetto, A., Cerchiello, P., Filomeni, S., Tanda, A., & Tarantino, B. (2024). *Can We Trust Machine Learning to Predict the Credit Risk of Small Businesses? Review of Quantitative Finance and Accounting*, 63(3), 925–954. DOI: 10.1007/s11156-024-01278-0

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). *Explainable Machine Learning in Credit Risk Management. Computational Economics*, 57(1), 203–216. DOI: 10.1007/s10614-020-10042-0

Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., & Ganatra, M. A. (2024). *Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers. Risks*, 12(11), 174. DOI: 10.3390/risks12110174

Emmanuel, I., Sun, Y., & Wang, Z. (2024). *A Machine Learning-Based Credit Risk Prediction Engine System Using a Stacked Classifier and a Filter-Based Feature Selection Method. Journal of Big Data*, 11(1), 23. DOI: 10.1186/s40537-024-00882-0

Espinoza, F. E. T., & Coral Ygnacio, M. A. (2023). *Credit Risk Assessment Models in Financial Technology: A Review. TecnoLógicas*, 26(58), e2679. DOI: 3442/344275988008

Laborda, J., & Ryoo, S. (2021). *Feature Selection in a Credit Scoring Model. Mathematics*, 9(7), 746. DOI: 10.3390/math9070746

Li, Y., Stasinakis, C., & Yeo, W. M. (2022). *A Hybrid XGBoost-MLP Model for Credit Risk Assessment on Digital Supply Chain Finance. Forecasting*, 4(1), 184–207. DOI: 10.3390/forecast4010011

Lyócsa, Š., Vašaničová, P., Hadji Misheva, B., & Vateha, M. D. (2022). *Default or Profit Scoring Credit Systems? Evidence from European and US Peer-to-Peer Lending Markets. Financial Innovation*, 8(1), 32. DOI: 10.1186/s40854-022-00338-5

Melese, T., Berhane, T., Mohammed, A., & Walelgn, A. (2023). *Credit-Risk Prediction Model Using Hybrid Deep–Machine-Learning Based Algorithms. Scientific Programming*, 2023, Article 6675425. DOI: 10.1155/2023/6675425

Nallakaruppan, M. K., Chaturvedi, H., Grover, V., Balusamy, B., et al. (2024). *Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence. Risks*, 12(10), 164. DOI: 10.3390/risks12100164

Pérez, E. M., Ramírez Guzmán, M. E., & Hernández Jiménez, A. (2024). *Predicción del riesgo crediticio a microfinanciera usando aprendizaje computacional. Revista Mexicana de Economía y Finanzas (Nueva Época)*, 19(4), e868. DOI: 10.21919/remef.v19i4.868

Robisco, A. A., & Carbó Martínez, J. M. (2022). *Measuring the Model Risk-Adjusted Performance of Machine Learning Algorithms in Credit Default Prediction. Financial Innovation*, 8(1), 70. DOI: 10.1186/s40854-022-00366-1

Shen, F., Zhao, X., Kou, G., & Alsaadi, F. E. (2021). *A New Deep Learning Ensemble Credit Risk Evaluation Model with an Improved Synthetic Minority Oversampling Technique. Applied Soft Computing*, 98, 106852. DOI: 10.1016/j.asoc.2020.106852

Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). *Machine Learning-Driven Credit Risk: A Systemic Review. Neural Computing and Applications*, 34(19), 14327–14339. DOI: 10.1007/s00521-022-07472-2

Diagnóstico FACIL Empresarial, ISSN 2683-1481
Año 12, Núm. 24, Julio • Diciembre 2025

Development of a Predictive Credit Risk Model Using Machine Learning Techniques on Simulated Financial Data
Curtido Quiñonez A.J.

Shi, Y., Qu, Y., Chen, Z., Mi, Y., & Wang, Y. (2024). *Improved Credit Risk Prediction Based on an Integrated Graph Representation Learning Approach with Graph Transformation. European Journal of Operational Research*, 315(2), 786–801. DOI: 10.1016/j.ejor.2023.12.028

Soni, U., Jethava, G., & Ganatra, A. (2024). *Latest Advancements in Credit Risk Assessment with Machine Learning and Deep Learning Techniques. Cybernetics and Information Technologies*, 24(4), 22–44. DOI: 10.2478/cait-2024-0034

Wang, W., Zuo, X., & Han, D. (2024). *Predict Credit Risk with XGBoost. Applied and Computational Engineering*, 74, 164–177. DOI: 10.54254/2755-2721/74/20240462