

## Article

# Machine Learning for Credit Risk Prediction: A Systematic Literature Review

Jomark Pablo Noriega <sup>1,2,\*</sup> , Luis Antonio Rivera <sup>1,3,†</sup>  and José Alfredo Herrera <sup>1,4,†</sup> 

<sup>1</sup> Departamento Académico de Ciencia de la Computación, Universidad Nacional Mayor de San Marcos, Decana de América, Lima 15081, Peru; lriviera@unmsm.edu.pe or rivera@uenf.br (L.A.R.); jherreraqu@unmsm.edu.pe (J.A.H.)

<sup>2</sup> Financiera QAPAQ, Lima 150120, Peru

<sup>3</sup> Centro de Ciências Exatas e Tecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes 28013-602, Brazil

<sup>4</sup> Programme in Biotechnology, Engineering and Chemical Technology, Universidad Pablo de Olavide, 41013 Sevilla, Spain

\* Correspondence: jomark.noriega@unmsm.edu.pe or jnoriega@qapaq.pe; Tel.: +51-987-521-006

† These authors contributed equally to this work.

**Abstract:** In this systematic review of the literature on using Machine Learning (ML) for credit risk prediction, we raise the need for financial institutions to use Artificial Intelligence (AI) and ML to assess credit risk, analyzing large volumes of information. We posed research questions about algorithms, metrics, results, datasets, variables, and related limitations in predicting credit risk. In addition, we searched renowned databases responding to them and identified 52 relevant studies within the credit industry of microfinance. Challenges and approaches in credit risk prediction using ML models were identified; we had difficulties with the implemented models such as the black box model, the need for explanatory artificial intelligence, the importance of selecting relevant features, addressing multicollinearity, and the problem of the imbalance in the input data. By answering the inquiries, we identified that the Boosted Category is the most researched family of ML models; the most commonly used metrics for evaluation are Area Under Curve (AUC), Accuracy (ACC), Recall, precision measure F1 (F1), and Precision. Research mainly uses public datasets to compare models, and private ones to generate new knowledge when applied to the real world. The most significant limitation identified is the representativeness of reality, and the variables primarily used in the microcredit industry are data related to the Demographic, Operation, and Payment behavior. This study aims to guide developers of credit risk management tools and software towards the existing ability of ML methods, metrics, and techniques used to forecast it, thereby minimizing possible losses due to default and guiding risk appetite.

**Keywords:** loan; credit risk; prediction; machine learning; systematic literature review



**Citation:** Noriega, J.P.; Rivera, L.A.; Herrera, J.A. Machine Learning for Credit Risk Prediction: A Systematic Literature Review. *Data* **2023**, *8*, 169. <https://doi.org/10.3390/data8110169>

Academic Editors: Henry Han, Qiannong (Chan) Gu, Diane Li, Tie Wei and Jeffrey Yi-Lin Forrest

Received: 5 August 2023

Revised: 13 October 2023

Accepted: 3 November 2023

Published: 7 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The digitalization of processes and AI are already part of our daily lives and have been developing in all areas with which we interact, especially during the period of the COVID-19 pandemic [1,2]. This trend continued with the promotion of online loans and Internet sales [3], and in turn, with the increase in demand for credit and crowdfunding through the Internet in the short term [4–6], and considering the pandemic as an external factor, it has considerably changed the economy, increasing uncertainty for financial institutions, and consequently the need to generate new models to manage it [4]. The World Bank (WB) emphasizes that the banking sector is crucial because it improves the well-being of a country's population and is essential for the growth of the economy [7].

In this competitive environment, financial institutions seek to differentiate themselves, generate shareholder value, improve the customer experience, and promote financial

inclusion. In this sense, they face the challenge of adopting data-driven innovation (DDI) to manage credit, customer and operational risk appetite, and seeking efficiency [8] while being supported by information technology—cloud services, Internet of Things (IoT), BIG DATA, AI and mobile telephony—known as the fourth industrial revolution [9]. This is followed by its sustainable evolution towards virtuous interaction between humans, technology, and the environment, called the fifth industrial revolution [10].

There are still many use cases for DDI and ML in financial institutions to solve [11,12]. There are gaps in the evaluation and precariousness in the results when processing large volumes of information [13] using ML algorithms for real-time applications [8]. On the other hand, using machine learning techniques, through association, collaborative filtering, along with recommended and personalized content, systems can identify individual preferences that could enrich the risk assessment [8]. Likewise, BIG DATA tools and technology could help improve forecasts in changing markets, considering their analysis potential [11,14].

Utilizing machine learning in business intelligence to reduce the uncertainty of payment behavior, also called credit risk, is a necessity in the microfinance industry since it allows for the analysis of large volumes of information generated, especially in the context of post-pandemic COVID-19 and technological development [4,15]. The challenges are determining which configurations of attributes and algorithms are best suited for the tasks, together with identifying limitations in the applications. Consequently, we propose the research topics shown in Table 1.

**Table 1.** Research Topics.

Motivation	Research Topics
We wish to know what models the industry and academics use to predict credit risk.	The algorithms, methods, and models used to predict credit risk.
We wish to know what metrics to use in the industry for academics to evaluate the performance of algorithms, methods, or models to predict credit risk.	The metrics to evaluate the performance of algorithms, methods, or models.
We wish to know the metrics accuracy, precision, F1 measure, and AUC of algorithms, methods, or models to predict credit risk.	The models' accuracy, precision, F1 measure, and AUC.
We wish to know what datasets to use in the industry and for academics to predict credit risk.	The datasets are used in the prediction of credit risk.
We wish to know what variables or features to use in the industry and for academics to predict credit risk.	The variables or features used in the prediction of credit risk.
We wish to know the main problems or limitations to predicting credit risk.	The main problems or limitations of predicting credit risk.

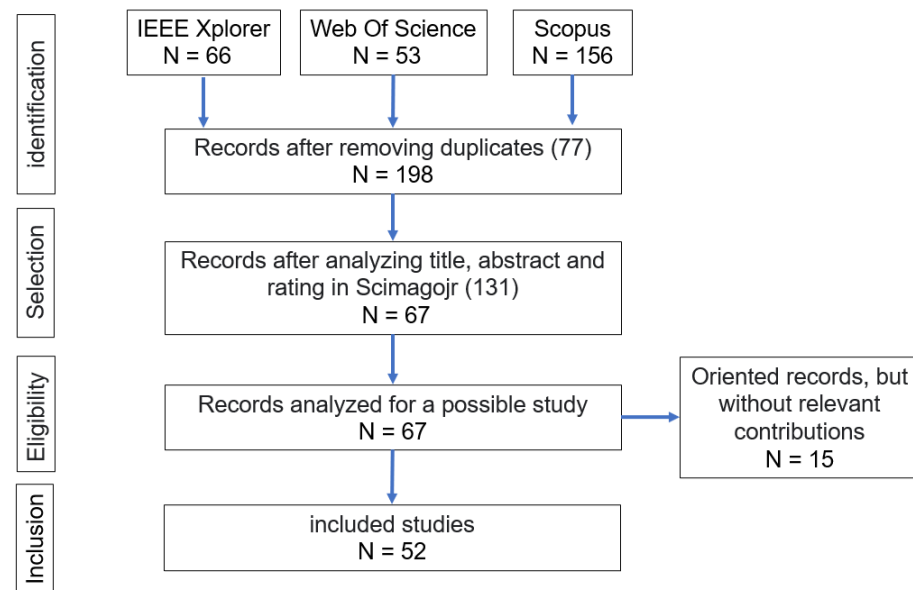
## 2. Materials and Methods

Based on the research topics, we pose the questions in Table 2 as a first step. Using the most relevant words related to our research topic, we build our search string and apply it to the recognized databases: IEEE Xplore, Scopus, and Web Of Science (WOS). We consider the studies belonging to articles from journals and conferences published from 2019 to May 2023 and are related to the computer science area as shown in Figure 1.

We identified 275 studies. There were 77 eliminations for being duplicates and 131 eliminations from applying the eligibility criteria (without full access to the document, without ranking in Scimagojr, review article) and 15 eliminations for not having relevance after the complete analysis of the document. We obtained as a result 52 revealing articles; we show in Table 3 the result of the application of the inclusion and exclusion criteria.

**Table 2.** Research Questions.

Id.	Metrics
RQ1	What are the algorithms, methods, and models used to predict credit risk?
RQ2	Which are the metrics to evaluate the performance of algorithms, methods, or models?
RQ3	What are these models' accuracy, precision, F1 measure, and AUC?
RQ4	What datasets are used in the prediction of credit risk?
RQ5	What variables or features are used to predict credit risk?
RQ6	What are the main problems or limitations of predicting credit risk?

**Figure 1.** PRISMA Research Strategy.

Research string:

TITLE-ABS-KEY (("credit" OR "loan") AND ("machine learning") AND ("model" OR "algorithm" OR "method") AND ("financial") AND ("credit risk")) AND (LIMIT-TO (PUBYEAR, 2023) OR LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) ) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SUBJAREA, "COMP")).

**Table 3.** Application of inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria	#	%
Article of conference		2	0.73%
Article of journal		50	18.18%
	Article duplicated	77	28.00%
	Not related	15	5.45%
	Review article	1	0.36%
	Without access to the full document	57	20.73%
	Without rank in Scimagojr	73	26.55%
	Total	275	100.00%

### Research Strategy

To determine the importance of studies, we assess that they include relevant conclusions or results; attributes or features of datasets; descriptions of the proposed model or algorithm; the metrics with which the models were evaluated; preprocessing techniques; identified problems or limitations; and future studies. We also consider when the article

includes more than one set of data or experiments; we select one of them to present the metrics considering the most relevant in terms of ACC, AUC, or another metric in the event the indicated ones are not used. According to our criteria and in the cases which it applies, we decided on the German Dataset, which is the dataset of the University of California Irvine (UCI), considering that it is imbalanced [16], of frequent use, and it will allow us to compare results between investigations that include it.

The main limitation when carrying out a systematic review is bias, and considering how the decisions of researchers influence the application of the method framed in experiences and previous knowledge. For example, the choice of topic, choice of electronic resources, proposal of research questions, methodology for data collection, and their evaluation. We have tried to meticulously follow the PRISMA method, taking into account the application of the most appropriate criteria and procedures in the construction of this document.

### 3. Current Research

The demand for online credit generates considerable information, with which, when analyzed using BIG DATA [17], designs new products, machine learning models, and credit risk assessment methods [18]. Consequently, in scenarios of increased demand, credit risks also escalate considerably, in a non-linear manner, considering the level of risk, rate, and terms of credit [19]. In the same manner, there is an expectation of an increase in fraud in the following years [20]. Another problem to consider is the consistency of the information recorded at the different stages of the process, such as sales data [21], cultural variables, environmental data [22], macroeconomics [23], innovation capacity management and development, exchange rate evolution, Gross Domestic Product (GDP) growth trends [23,24], economic activity, and experience [25].

The relevant problems are addressed by various research papers, using different ML approaches for the respective interpretations and decision-making [26]. However, there are also difficulties with the implemented models [27], which generally follow the black box model; that is, to predict, for example, the good and bad payers [28–30]. These models have presented problems, especially in difficult times, such as “the financial crisis of 2008” [1,31], since financial institutions focus on loans that generate the most income, being, therefore, of higher risk due to payment defaults [19,22]. Automatic evaluation models, based on credit data, could confuse good paying customers for bad ones [20], and apply penalties on possible benefits [32]. The low explainability of advanced non-linear ML models is slowing down their implementation [33]. The challenge is in the development of Explainable artificial intelligence (XAI), whose objective is to provide predictive models with inherent transparency, especially in complex cases [34,35]. In that sense, one could use Shapley Additive exPlanations (SHAP) or the Generalized Shapley Choquet integral (GSCI) to expose the parameter dependency [36,37], and a model interpretable payment risk forecast from a penalized logistic tree to enhance the performance of Logistic Regression (LR), and LR + Slack Based Measure (SBM) [38,39]. The MLIA model and variable logistic regression coefficient more intuitively reveal the contribution of a variable to the predicted outcome [40]. The “non-payment” problem is important because it could generate significant losses for financial entities [41,42].

Here, the challenge of machine learning is to consider the existing multicollinearity in the input data [43,44] where there is a high correlation of variables and some that is not useful for classification [2]. This imbalance in the data actually used, with probable overfit [20], could generate biases in machine learning [45,46], causing chaotic reputation management, and malicious or criminal deception [47]. In other words, the challenge is to determine the effective, relevant features [48], for example, for the training of neural networks (NNs) (which is performed end-to-end by interactions). These are additional constraints with desirable features known a priori in order to reduce the number of interactions and prioritize smoothness; they are factors of explanatory interest, in the domain, control, and generative modeling of features [49]. Other authors recommend the use of genetic

algorithms (GAs) to guide training, with data sequences that have the best result [50]; the use of hive intelligence (HI) is also highlighted for this purpose [51–53]; Boost Category models (BCat) such as Adaboost (ADAB), XGBoost (XGB), LightGBM (LGMB), and Bagging (BAG) [30,40,54–56]; multi-classification (MC) and information fusion models (MIFCA) [44]; there is also noise removal using fuzzy logic (FL) and contribution to the identification of main attributes [28,37,57]. It is worth noting the possibility to evaluate the interaction of borrowers within their supply chain to enrich predictive models [17,58]. The use of images, interviews, text and social information, and interaction with mobile technology would give the credit risk assessment a multidimensional and multi-focus characteristic [15,59,60]; this indicates the inclusion of integrated accounting information with statistical variables of profitability, liquidity, asset quality, management indices, capital adequacy, efficiency, SCORECARD scorecard and the maximization of the internal rate of return (IRR), risks of the industry, and GDP [1,23].

Some authors maintain that the most relevant characteristics are gender, educational level, mortgage credit, microfinance, debt balance, and days past due [61]. Other authors maintain that the most relevant variables are the days of default, especially those greater than 90 days, to determine non-payment behavior and consider that discriminatory variables such as gender, age, and marital status should not be considered [55,61]. Consequently, the challenge of validating the quality of the features arises [39,62]. For this part, [17] maintains that the number of influential variables for risk assessment has increased, and the linear and non-linear time series relationships have increased their complexity.

There are also methods to address the imbalance problem: Random Undersampling (RUS), Near Miss (NMISS), Cluster Centroid (CC), Random Oversampling (ROS), Adaptive Synthetic (ADASYN), K-Nearest Neighbor (KNN) SMOTE (KN-SMOTE), Synthetic Minority Oversampling Technique (SMOTE), Borderline-Smot (B-SMOT), and Smotetomek (SMOTE-T) [63]. Other authors propose the use of CS-classifiers [64], and KFold [29,55]. However, other authors propose considering imbalance and missing data as characteristics to take into account in the [54,56], evaluation. As an initial part of the model evaluations in the experiments, the importance of optimizing the hyperparameters highlight using, for example, GA [15], K-Fold CV [38], random search (RS) [20], grid search (GS) [57,63], and other methods [56].

In Table 4, we summarize the challenges in credit risk prediction and the ML methods and techniques proposed by the authors to address them. The main challenge is high uncertainty. The authors propose various classification models, including Boosted category models, neural networks, deep learning enhanced with fuzzy logic, genetic algorithms, and hive intelligence. For the low explainability of the results, the authors propose applying Explainable artificial intelligence, Shapley Additive exPlanations, Generalized Shapley Choquet integral, and MLIA. To address the complexity of ML models, the authors propose optimizing the hyperparameters of the models supported with KFold CV, genetic algorithms, and grid search. For the multivariate origin of the data, the authors propose applying BIG DATA to take advantage of its prominent volume characteristics and its unstructured nature. To address the natural characteristics of unbalanced data, the authors propose the application of SMOTE, RUS, ROS, KFold, and ADASYN.

**Table 4.** Challenge of Credit Risk Prediction.

It.	Challenges	ML Methods & Techniques	#	Author
1	Uncertainty	NN, DL, BCAT, MC, MIFCA, FL, GA, HI	28	[2,15,17,19,20,28,30,36–58]
2	Explainability	XAI, SHAP, GSCI, MLIA	18	[1,19,20,22,27–40]
3	Complexity	GA, K-Fold CV, RS, GS	9	[15,17,20,23,24,38,56,57,63]
4	Multivariate Data	BIG DATA	9	[1,15,17,23–25,59–61]
5	Unbalanced Data	SMOTE, RUS, ROS, KFold, CS, ADASYN	7	[20,29,54–56,63,64]



## 4. Results

### 4.1. Answer to RQ1: What Are the Algorithms, Methods, and Models Used to Predict Credit Risk?

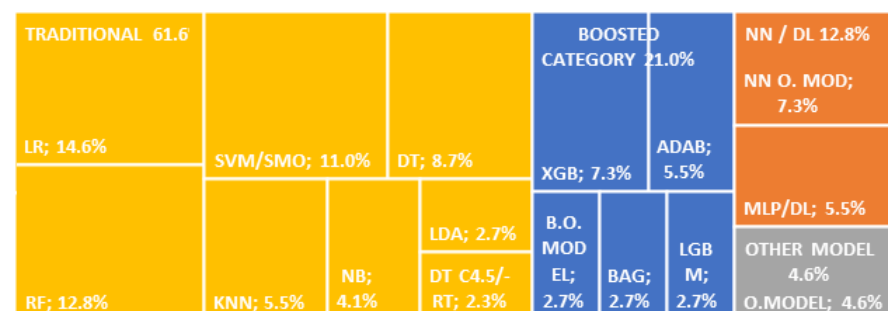
In their research to forecast credit risk, the authors use ML models: 72.76% non-assembled (N-Ass) and 27.24% assembled (Ass), which are shown in Table 5.

**Table 5.** Family of algorithms, methods, and models.

It.	Family	#			%		
		Ass	N-Ass	Total	Ass	N-Ass	Total
1	Boosted Category	36	46	82	11.96%	15.28%	27.24%
2	Collective Intelligence	7	-	7	2.33%	0.00%	2.33%
3	Fuzzy Logic	10	-	10	3.32%	0.00%	3.32%
4	NN/DL	8	28	36	2.66%	9.30%	11.96%
5	Other Model	3	10	13	1.00%	3.32%	4.32%
6	Traditional	18	135	153	5.98%	44.85%	50.83%
Total		82	219	301	27.24%	72.76%	100.00%

For better presentation we group them into the following families: Boosted Category, the models related to the Boosted algorithm; Collective Intelligence, models related to collective or swarm intelligence including Ant Colony Optimization (ACO), Bat Algorithm (BAT), and Particle Swarm Optimization (PSO); Fuzzy Logic, models related to Fuzzy Logic including Clustering-Based Fuzzy Classification (CBFC); NN/DL, models related to neural networks or Deep Learning (DL) including Back Propagation Neural Network (BPNN), Artificial Neural Network (ANN), Multilayer Perceptron (MLP), Wide and Deep Neural Networks (Wide&Deep), Gated Recurrent Unit (GRU), Geometric Deep Learning (GDL), Graph Neural Networks (GNNs), Deep Genetic Hierarchical Network of Learners (DG), and Convolutional Neural Networks (CNNs); Other Models, for the models not cataloged; and Traditional, for the models cataloged but not related to previous models including Decision Tree (DT), Decision Tree C4.5 (C4.5), Classification And Regression Tree (CART), K-Means (KM), Linear Discriminant Analysis (LDA), Non-Linear Logistic Regression (NLR), Naive Bayes (NB), Random Forest (RF), Random Tree (RT), Support Vector Machine (SVM), and Sequential Minimal Optimization (SMO). Of the total models used, 50.83% correspond to the Traditional model family, 27.24% to the Boosted Category, and 11.96% to NN/DL.

Analyzing the groups separately corresponding to N-Ass as shown in Figure 2, the Category Boosted family was used in 21% of the studies, NN/DL has a use rate of 12.8%, Traditional has a use rate of 61.6%, and Other Models have a use rate of 4.6%.



**Figure 2.** Non-Assembled Models.

In Figure 3, the Category Boosted family was used 43.9% of the time, NN/DL has a use rate of 9.8%, Traditional has a use rate of 22%, Other Models have a use rate of 3.7%, Fuzzy Logic has a use rate of 12.2%, and Collective Intelligence is used in 8.5% of the total. These results will demonstrate that N-Ass models are mostly used in credit risk prediction. However, the authors use these as a baseline to compare them with Ass models generated by nesting N-Ass models and improving them using fuzzy logic and hive intelligence.

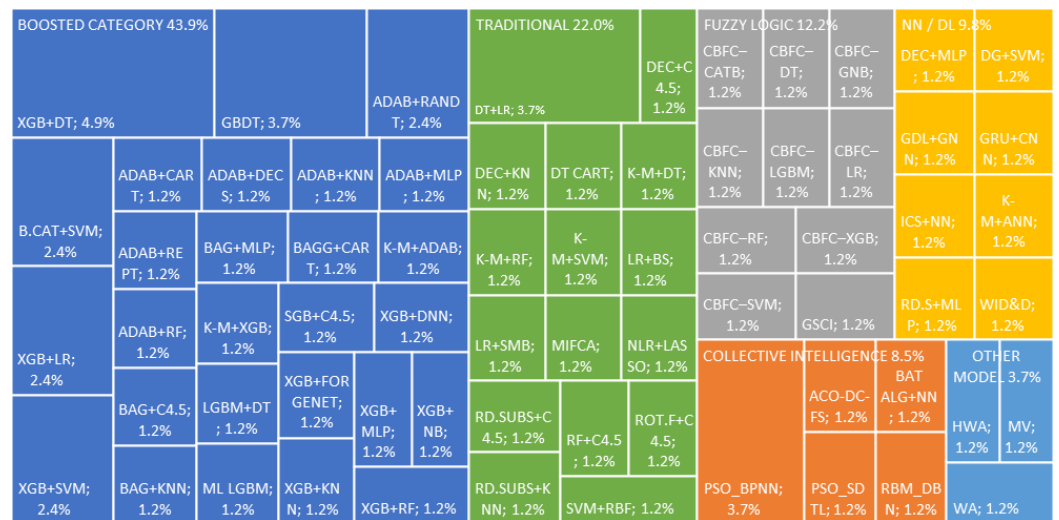


Figure 3. Assembled Models.

#### 4.2. Answer to RQ2: Which Are the Metrics to Evaluate the Performance of Algorithms, Methods, or Models?

We collected the metrics used by the researchers in their articles, taking into account that for the cases in which they evaluate more than one dataset and more than one model is applied, a pair is searched, looking for the best ACC and AUC values, or another value in the cases where these are not used, as shown in Figure 4.

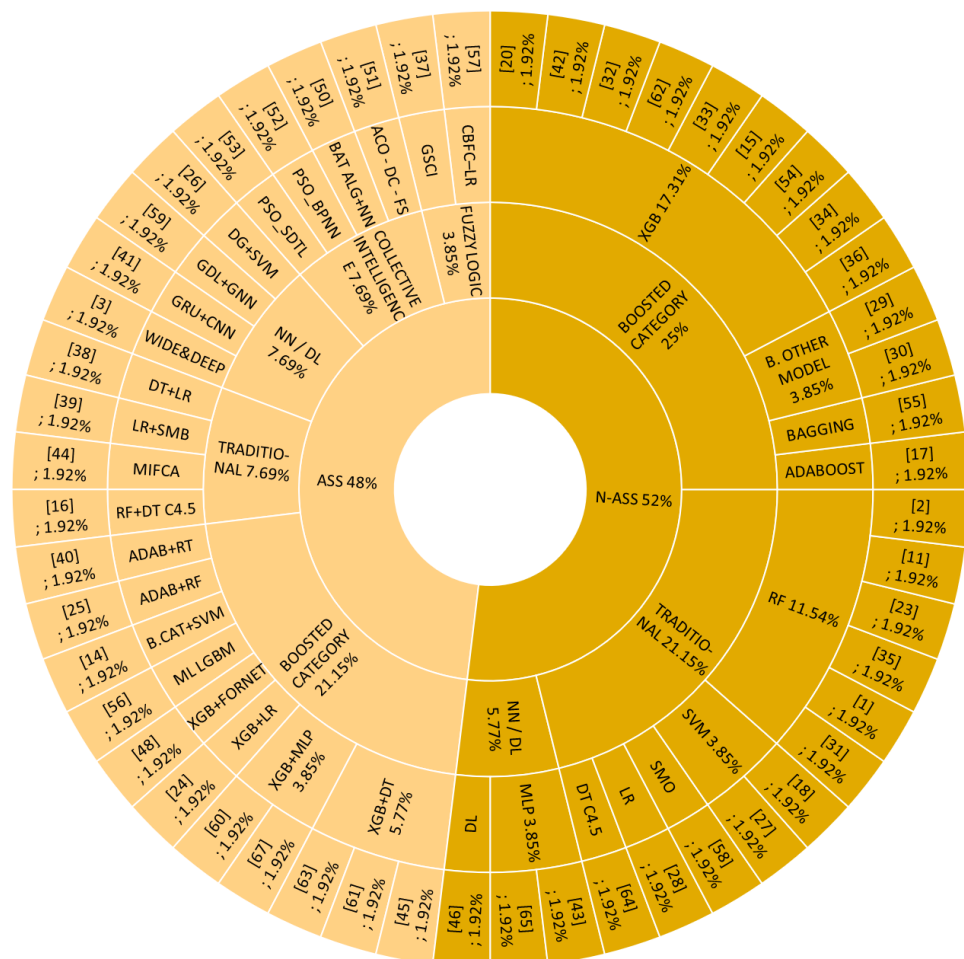


Figure 4. Best Models with family and author.

From this simplification, the authors propose 48% assembled models and 52% non-assembled models.

Of the total assembled models, the Boosted Category has 21%; Collective Intelligence has 8%; NN/DL has 8%, Traditional has 8%, and Fuzzy Logic has 4%. Of the non-assembled models, the Boosted Category makes up 25%; Traditional is 21%; and NN/DL is 6%.

In these models, the metrics used by the authors are divided into AUC with 16%, ACC with 14%, F1 measure with 11%, others with 11%, Precision with 10%, Recall with 9%, True Positive Rate (TPR) with 7%, True Negative Rate (TNR) with 6%, Geometric Mean (GMEAN) with 4%, Kolmogorov–Smirnov (KS) with 3%, Brier Score (BS) with 3%, GINNI Score (GINNI) with 2%, and Root Mean Squared Error (RMSE), KAPPA coefficient (KAPPA), and Mean Absolute Error (MAE), whose group participation is 2%; the details are shown in Table 6.

These results could demonstrate that the most-used metrics for credit risk prediction are AUC and ACC since they allow for the comparison of different models. However, AUC is prioritized because the distribution of the classes does not influence it and has better behavior when using unbalanced data.

**Table 6.** Metrics.

It.	Metrics	#	%	Author
1	AUC	34	16.11%	[1,2,11,16,17,20,24,25,27,28,30,32–43,46,50,54–58,61–64]
2	ACC	30	14.22%	[3,11,15,18,25–32,35,36,37,39,44,46,48,50–53,55,57,58,60–63,65]
3	F1	24	11.37%	[11,17,18,24,27,29,30,32,36,37,39,41,42,44,48,51,56,58,60–65]
4	Precis.	22	10.43%	[3,11,17,18,24,27–30,32,36,39,41,42,44,46,58,60–62,65]
5	Recall	19	9.00%	[11,24,27–30,36,39,41,42,44,46,58,60–63,65,66]
6	TPR	14	6.64%	[16,18,24,25,31,32,35,37,45,50,51,55,57,65]
7	TNR	13	6.16%	[16,25,31,32,35,37,39,45,50,51,55,57,65]
8	GMEAN	9	4.27%	[20,31,35,39,45,48,50,63,64]
9	KS	7	3.32%	[24,36,38,39,43,55,56]
10	BS	6	2.84%	[33,37,38,55,57,62]
11	GINNI	5	2.37%	[2,24,38,43,46]
12	RMSE	2	0.95%	[53,67]
13	KAPPA	1	0.47%	[51]
14	MAE	1	0.47%	[67]
15	Other	24	11.37%	[2,14,23,25,28,33,35,36,38,41,43,45,46,51,53,54,57,59–62,64,65,67]
Total		211	100.00%	

#### 4.3. Answer to RQ3: What Are These Models' Accuracy, Precision, F1 Measure, and AUC?

We listed the values of the five metrics most used by the authors in their research: AUC, ACC, F1 measure, Precision, and Recall, as shown in Table 6. Furthermore, we have taken the values in each case according to the tuple defined in the question. To compare experiments, the characteristics of the dataset and metric to be evaluated should correspond. Below, we show only the metrics' values the authors used and evaluated in their research in Table 7. If the metrics have an empty value, the authors do not consider them in their experiment design.

**Table 7.** Metrics author.

It.	Dataset	Author	Metrics' Values				
			ACC	Precision	F1	Recall	AUC
1	UCI Taiwan	[29]	85.00	70.00	50.00	62.00	-
2	UCI German	[63]	83.50	82.10	84.40	86.80	91.00
3	UCI German	[25]	82.80	-	-	-	91.20
4	UCI German	[50]	81.18	-	-	-	85.38
5	UCI German	[51]	76.60	-	84.74	-	-



Table 7. Cont.

It.	Dataset	Author	Metrics' Values				
			ACC	Precision	F1	Recall	AUC
6	UCI German	[28]	75.80	54.20	-	82.00	85.90
7	UCI German	[57]	74.90	-	-	-	75.80
8	UCI German	[16]	-	-	-	-	79.40
9	Lending Club	[32]	92.60	97.90	92.20	-	97.00
10	Lending Club	[65]	84.40	88.99	91.42	93.98	-
11	Lending Club	[30]	76.10	75.98	75.95	76.35	76.80
12	Lending Club	[48]	88.77	-	94.14	-	-
13	Lending Club	[31]	74.90	-	-	-	-
14	Lending Club	[35]	64.00	-	-	-	71.70
15	Lending Club	[36]	63.60	85.30	73.50	64.50	67.40
16	Lending Club	[46]	-	18.25	-	46.88	63.63
17	Lending Club	[56]	-	-	2.72	-	75.86
18	K Prosper	[3]	78.50	54.70	-	-	-
19	K Prosper	[17]	-	79.00	71.00	65.00	80.00
20	K Give Me	[61]	88.30	78.50	77.60	76.70	93.30
21	RenRenDai	[37]	93.35	-	73.12	-	82.64
22	BR	[55]	96.68	-	-	-	89.63
23	AVG Used	[11]	92.80	31.60	33.40	35.50	82.80
24	AVG Used	[64]	-	-	91.89	-	96.19
25	UCI Austr...	[26]	97.39	-	-	-	-
26	Tsinghua	[52]	91.23	-	-	-	-
27	Tsinghua	[62]	77.20	75.90	77.54	79.38	85.01
28	Private Data	[18]	98.34	100.00	96.00	-	-
29	Private Data	[53]	98.00	-	-	-	-
30	Private Data	[60]	97.80	98.90	98.70	98.90	-
31	Private Data	[15]	90.10	-	-	-	-
32	Private Data	[27]	84.29	82.63	84.68	86.83	84.29
33	Private Data	[44]	84.15	82.15	83.40	84.68	-
34	Private Data	[58]	83.00	83.50	83.00	83.00	83.30
35	Private Data	[39]	77.49	79.87	85.59	92.18	79.00
36	Private Data	[24]	-	87.15	84.56	83.91	83.59
37	Private Data	[54]	-	-	-	-	46.10
38	Private Data	[1]	-	-	-	-	75.40
39	Private Data	[38]	-	-	-	-	85.68
40	Private Data	[33]	-	-	-	-	93.39
41	Private Data	[34]	-	-	-	-	93.00
42	Private Data	[42]	-	42.81	52.00	67.01	78.00
43	Private Data	[40]	-	-	-	-	71.32
44	Private Data	[2]	-	-	-	-	91.40
45	Private Data	[41]	-	88.00	88.00	88.00	93.00
46	Private Data	[43]	-	-	-	-	77.56
47	Private Data	[20]	-	-	-	-	95.50

Considering that there are experiments in which the same dataset is used, we can compare their prediction capabilities. This occurs in the case of the UCI German Dataset, in which the XGB + DT model has an ACC of 84 [63], against the LR models [28], and Random Forest (RF) + C4.5 [16]. Other less-used datasets on which we can compare the metrics are the Tsinghua—Pekin U RESSET database and the Kaggle Prosper dataset, as shown in Table 7. It is essential to indicate that the Lending Club dataset is possible to compare; however, it is necessary to consider the different date ranges for the investigations that the authors used [35,36].

#### 4.4. Answer to RQ4: What Datasets Are Used in the Prediction of Credit Risk?

The datasets used by the authors can be divided into 53.85% public and 46.15% private, seeking mainly in the first case to validate new models. To be able to compare them with experiments, we show the data in Figure 5—private datasets used to extract knowledge by revalidating existing models in real scenarios. In the public group, the most used is the UCI German dataset with a usage of 15.38%, possibly due to its characteristics [16]; the second most used is the Lending Club platform, which is a P2P loan platform used in 15.38% of the studies. Some authors in their experiments validate both public and private data, and Ass and N-Ass models, and determine their behavior in different scenarios.

The financial industry needs machine learning tools to support credit risk prediction, as many experiments with private datasets demonstrate. However, since these use sensitive data, their access is limited, which could slow down community efforts to improve prediction using actual cases.

#### 4.5. Answer to RQ5: What Variables or Features Are Used to Predict Credit Risk?

The authors propose many variables, using different methods to identify the variables with the best predictive capacity. GA, FL, hive intelligence, statistical methods, and functions are used to determine the correlation; in Table 8 we display the details.

To simplify the analysis, we have grouped the proposed variables into Demographic, which has a 54.09% share; Operation with 29.18%; Payment behavior with 7.62%; External factors with 6.69%; Unstructured data with 1.30%, and Transaction with 1.12%, as shown in Table 9.

The preference in the use of demographic variables for the prediction of credit risk would have to be explained because these can represent the behavior, preferences, and socioeconomic profile of the client or the segments to which they belong; their inclusion contributes positively to the models, but it is not sufficient. It must be accompanied by Payment behavior variables, characteristics of the operation, and environmental variables that could influence the results.

PRIVATE DATASET 46.15%				LENDING CLUB 15.38%		UCI GERMAN 15.38%
ZHAO(22); 1.92%	CHROŚCICKI(22); 1.92%	LOMBARDO(22); 1.92%	DUMITRESCU(22); 1.92%	JEMAI(23); 1.92%	SI(22); 1.92%	MOSCATO(21); 1.92%
LI(22); 1.92%	MERČEP(21); 1.92%	BUSSMANN(21); 1.92%	MOUSAVI(20); 1.92%	CHEN(19); 1.92%	LIU(22); 1.92%	MANCIS..(20); 1.92%
FAN(20); 1.92%	TIAN(22); 1.92%	RISHEHCHI FAYYAZ(21); 1.92%	WANG(22); 1.92%	BISWAS(22); 1.92%	LI(21); 1.92%	ARIZAG(20); 1.92%
ZHENG(19); 1.92%	LI(22); 1.92%	KARN(22); 1.92%	MUNOZ-C(23); 1.92%	AVG USED 3.85%	YILDIRIM(21); 1.92%	PEKIN URESSET 3.85% FENG(22); 1.92%
MA(19); 1.92%	ZIEMBA(21); 1.92%	MANCI(21); 1.92%	CHO(23); 1.92%	SONG(19); 1.92%	PROSPER 3.85%	WANG(21); 1.92%
MACHADO(22); 1.92%	TELES(21); 1.92%	LIU(22); 1.92%	ZOU(22); 1.92%	LIU(21); 1.92%	QIAN(22); 1.92%	BR-DA PMCMV DE.CAST.V.(19); 1.92%
						K.GIVE MESC 1.92%
						UCI AUS. 1.92%
						UCI TAIWAN MITRA(22); 1.92%

Figure 5. Dataset used.

**Table 8.** Features.

It.	Features Group	Feature	#	%
1	Demographic	External Debt Value/historical	27	5.02%
2	Demographic	Domestic Debt Value/historical	27	5.02%
3	Operation	Loan value	24	4.46%
4	Demographic	Average/Total revenue	20	3.72%
5	Demographic	Residence/Registered Assets	19	3.53%
6	Demographic	Economic Activity/Experience	18	3.35%
7	Demographic	Family Income	18	3.35%
8	Payment behavior	Days in arrears/Range Days in arrears	17	3.16%
9	Operation	Historical use of debt	16	2.97%
10	Operation	Destination of the Credit/Purpose	16	2.97%
11	Operation	Interest Rate	16	2.97%
12	External factors	Debt Profitability	16	2.97%
13	Demographic	Total Debt/Income/DTI	15	2.79%
14	Demographic	Gender/Sex	14	2.60%
15	Demographic	Risk Segment/Buro Rating/Score	14	2.60%
16	Demographic	Age/Date of Birth	13	2.42%
17	Operation	Checking/Savings Account	13	2.42%
18	Operation	Credit Line Limit	13	2.42%
19	Demographic	Civil Status	12	2.23%
20	Demographic	Mortgage Debt	12	2.23%
21	Operation	Monthly Fees	12	2.23%
22	Payment behavior	Collection status	11	2.04%
23	Payment behavior	Unpaid Installment Number	11	2.04%
24	Demographic	Financial maturity	9	1.67%
25	Demographic	Residence type	9	1.67%
26	Demographic	Fee value	9	1.67%
27	External factors	Inventory turnover	9	1.67%
28	Demographic	Labor Old	7	1.30%
29	Demographic	Education Level	7	1.30%
30	Others	Others	114	21.21%
Total			538	100.00%

**Table 9.** Features Group.

It.	Features Group	#	%
1	Demographic	291	54.09%
2	Operation	157	29.18%
3	Payment behavior	41	7.62%
4	External factors	36	6.69%
5	Unstructured data	7	1.30%
6	Transaction	6	1.12%
Total		538	100.00%

#### 4.6. Answer to RQ6: What Are the Main Problems or Limitations of Predicting Credit Risk?

In the reviewed articles, the authors state limitations or problems that they have faced during the experiments, although in each case there are nuances. We have grouped them into the following: Representativeness of reality, seen in 32% of the studies, refers to the fact that many of the existing variables do not reflect the true nature of the information. Unbalanced data seen in 28%, refers to the fact that, according to some authors, the usage of highly unbalanced data significantly reduces the performance of the models. Inconsistency in information recording is noted in 17%, where reference is made to the fact that the existing records have been entered with errors, bias, and noise that generates the need to apply cleaning techniques, with the risk of losing certain information. Lack of ability to explain the proposed results in 13% refers to the explainability limitation that most robust

ML models have. Availability of information and centralized processing in 6% refers to the need to process information centrally, which can generate additional losses, noise, and delays. Adaptability in processing structured and unstructured information in 4%, refers to the need to process structured and unstructured data within the operation process. We show the results in Table 10.

**Table 10.** Limitations.

It.	Limits Identified	#	%
1	Representativeness of reality	39	31.71%
2	Unbalanced data	35	28.46%
3	Inconsistency in information recording	21	17.07%
4	Lack of ability to explain the proposed results	16	13.01%
5	Availability of information and centralized processing	7	5.69%
6	Adaptability in processing struct. and unstruct. information	5	4.07%
Total		123	100.00%

For credit risk prediction models, the difficulties center on the consistency of available information. Furthermore, the capacity of the models to process it is a limitation, since in this industry, its fundamental nature corresponds to unbalanced data.

## 5. Additional Findings

During the development of this RSL, we identified preprocessing techniques that the authors refer to during their experiments.

The techniques most used by the authors for estimating the hyperparameters are KFold in 58.33% of the studies and Grid Search in 22.22%. We display the details in Table 11.

The most-used algorithms to mitigate the problem of imbalance in the datasets are SMOTE at 29.55%, followed by KFold with 20.45%, RUS at 11.76%, and ROS similarly at 11.36%. We show the detail in Table 12. Also used are ADASYN, some variants of SMOTE like B-SMOTE, SMOTE-T, adapted classification algorithms like KN-SMOTE, Under-Bagging, and techniques to identify missing values such as NMISS, CC, CS-Classifiers, and RESAMPLE.

Credit risk prediction can be schematized as a classification problem with unbalanced samples [20]; in that sense, the authors' preprocessing techniques can be considered the baseline for new research.

**Table 11.** Techniques for determination of hyperparameters.

It.	Method	#	%	Author
1	KFold CV	21	58.33%	[1,2,15,18,26,28,29,31,33,35,38,40,41,43,45,46,50,55,59,62,67]
2	Grid Search Method	8	22.22%	[11,15,16,44,48,56,57,63]
3	LightGBM Bayesian Optimization	2	5.56%	[20,56]
4	Genetic Algorithm	2	5.56%	[26,52]
5	Random Search	1	2.78%	[20]
6	Ant Colony Optimization	1	2.78%	[53]
7	Other	1	2.78%	[65]
Total		36	100.00%	

**Table 12.** Dataset balancing techniques.

It.	Method	#	%	Author
1	SMOTE	13	29.55%	[2,11,15,20,25,30,32,35,36,44,45,63,64]
2	KFold	9	20.45%	[14,16,24–26,29,50,55,57]
3	ROS	5	11.36%	[2,20,35,45,63]
4	RUS	5	11.36%	[2,20,45,63,64]
5	ADASYN	2	4.55%	[35,63]

Table 12. Cont.

It.	Method	#	%	Author
6	SMOTEBoost	2	4.55%	[20,64]
7	B-SMOT	1	2.27%	[63]
8	CC	1	2.27%	[63]
9	CS-Classifiers	1	2.27%	[64]
10	KN-SMOTE	1	2.27%	[63]
11	NMISS	1	2.27%	[63]
12	RESAMPLE	1	2.27%	[48]
13	SMOTE-T	1	2.27%	[63]
14	Under-Bagging	1	2.27%	[64]
Total		44	100.00%	

## 6. Discussion

The most widely used ML models to assess credit risk correspond to the Traditional family, possibly due to their easy implementation. On the other hand, those with the best prediction results correspond to the Boosted Category, both in the Ass and in the N-Ass groups. This trend is evidenced in Table 13, where this family obtains 24 evaluations out of 52 and is constantly growing through recent years. Another trend identified is that better results are acquired with the N-Ass models [67], for example, in the experiments which obtained AUC 91.00 with the AdaB + RF model [63] and 91.20 with XGB + DT [25], respectively. As shown in Table 7 and Figure 4, these results could be explained by gradient-based optimization features, parallelism, high-volume throughput, and missing values. The most-used metrics correspond to AUC, ACC, F1 measure, Recall, and Precision; the authors propose more specialized metrics according to the situation being evaluated. AUC and ACC in 16.11% and 14.22% of the studies, are mainly due to their ability to measure the capacity of different types of ML models. In the first case, it does not vary before the transformation by normalization, which allows analyzing unbalanced data with a high cost of classification; in the second case, it works better with balanced data, and that of easy explanation.

Table 13. Model trends.

It.	Family	2019	2020	2021	2022	2023	Total
1	Boosted Category	4	4	5	10	1	24
2	Traditional	4	1	5	4	1	15
3	NN/DL	1	1	2	2	1	7
4	Collective Intelligence		2		2		4
5	Fuzzy Logic		1			1	2
Total		9	9	12	18	4	52

Likewise, of the experiments carried out, the most significant databases used are 53.85% public datasets, while 46.15% correspond to private ones; the first serves to evaluate the predictive capacity of the new models proposed by the authors, comparing the results with previous experiences and the second to generate new knowledge through its application in the real world. In the experiments, the authors identify as the main problem in the datasets for the design of ML models the misrepresentation of reality due to possible bias, inconsistency, or error when recording the information. The second problem corresponds to the imbalance of the data, which can impair the excellent performance of the models. To face this problem, the SMOTE algorithm is mainly used, and for the optimization of the hyperparameters, the Kfold CV and Grid Search Method techniques are utilized. However, some authors propose hive intelligence [53] and genetic algorithms [26,52] to guide optimization. Finally, the most-used variables that best represent the intention and ability to pay, which in turn originate credit risk, correspond to the Demographic, Transaction, and

Payment behavior features. These encompass the main characteristics expected to predict it; see Table 8. However, the corresponding external features and unstructured data must be considered, bearing in mind the former, the influence in the hyperconnected world, the growing development of DDI, and the processing capacity of BIG DATA.

## 7. Conclusions and Future Research

In this systematic review article of the literature on credit risk prediction using ML, we reach the following conclusions:

- The Boosted Category is a family of ML models being investigated. They are the most used in Ass and N-Ass situations, highlighting the XGB model, with the tendency being its use in Ass. This category in the experiments obtained better results than other models due to its ability to process categorical variables—numerical, with noise, missing, and unbalanced data—and applying regularization could avoid overfitting. However, since they are complex models, they are challenging to interpret and are not very tolerant of applying atypical values.
- The five most-used metrics are AUC, ACC, Recall, F1 measure, and Precision, although, in practice, the problem must be considered when choosing the most appropriate metrics. However, AUC stands out for its ability to not be influenced by the distribution of classes and preferable behavior when processing unbalanced data.
- Public datasets are more utilized; of this group, the commonly used are UCI German Dataset and Landing Club Dataset. Their main use is to validate the behavior against other models under the same conditions. Private datasets generate knowledge from the application to a specific situation.
- For the evaluation of credit risks through ML, demographic variables are mainly used, which represent behavior, preferences, socioeconomic profile, and operations that represent the characteristics of the financial product acquired. However, this information is insufficient and is complemented by external variables and those related to unstructured data such as images, video, or others generated from hyperconnectivity, which is supported by DDI and BIG DATA development and processing.
- The main problems are the representativeness of reality, the imbalance of data for the training, and the inconsistency in recording information. All cases arise due to biases, errors, or problems in recording information.
- The most widely used method to solve the imbalance problem is SMOTE to optimize the performance of ML models, while the methods to determine the hyperparameters are KFold-CV and Grid Search to guide their optimization.

The credit risk prediction contribution corresponds to the stage where the credit originates. In this sense, we propose to extend the application of ML to precise credit datasets from specialized companies, including these models in other processes, such as credit collection and customer retention, considering the regulatory impositions governments are implementing to mitigate possible losses in the industry. Credit risk prediction can be enhanced with BIG DATA analysis, especially on unstructured data such as images, text, writing, sentiment, and hive intelligence, to assess adaptability to changing scenarios [17]. Finally, and in the same sense, including variables that represent the state of the environment could contribute to reducing uncertainty in this sector in the face of unexpected external events.

**Author Contributions:** Conceptualization, J.P.N. and J.A.H.; methodology, J.P.N.; validation, J.P.N., L.A.R. and J.A.H.; formal analysis, J.P.N. and J.A.H.; investigation, J.P.N.; resources, J.P.N.; writing—original draft preparation, J.P.N.; writing—review and editing, J.P.N., L.A.R. and J.A.H.; visualization, J.P.N.; supervision, J.A.H.; project administration, J.P.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.



**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lombardo, G.; Pellegrino, M.; Adosoglou, G.; Cagnoni, S.; Pardalos, P.M.; Poggi, A. Machine Learning for Bankruptcy Prediction in the American Stock Market: Dataset and Benchmarks. *Future Internet* **2022**, *14*, 244. [CrossRef]
2. Ziemba, P.; Becker, J.; Becker, A.; Radomska-Zalas, A.; Pawluk, M.; Wierzba, D. Credit decision support based on real set of cash loans using integrated machine learning algorithms. *Electronics* **2021**, *10*, 2099. [CrossRef]
3. Liu, C.; Ming, Y.; Xiao, Y.; Zheng, W.; Hsu, C.H. Finding the next interesting loan for investors on a peer-to-peer lending platform. *IEEE Access* **2021**, *9*, 111293–111304. [CrossRef]
4. Chen, C.; Lin, K.; Rudin, C.; Shaposhnik, Y.; Wang, S.; Wang, T. A holistic approach to interpretability in financial lending: Models, visualizations, and summary-explanations. *Decis. Support Syst.* **2022**, *152*, 113647. [CrossRef]
5. Shih, D.H.; Wu, T.W.; Shih, P.Y.; Lu, N.A.; Shih, M.H. A Framework of Global Credit-Scoring Modeling Using Outlier Detection and Machine Learning in a P2P Lending Platform. *Mathematics* **2022**, *10*, 2282. [CrossRef]
6. Zhang, Z.; Jia, X.; Chen, S.; Li, M.; Wang, F. Dynamic Prediction of Internet Financial Market Based on Deep Learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 1465394. [CrossRef]
7. BM Panorama General. Available online: <https://www.bancomundial.org/es/topic/financialsector/overview> (accessed on 22 December 2021).
8. Hani, U.; Wickramasinghe, A.; Kattiyapornpong, U.; Sajib, S. The future of data-driven relationship innovation in the microfinance industry. In *Annals of Operations Research*; Springer: Dordrecht, The Netherlands, 2022; pp. 1–27.
9. Zhang, C.; Zhong, H.; Hu, A. A Method for Financial System Analysis of Listed Companies Based on Random Forest and Time Series. *Mob. Inf. Syst.* **2022**, *2022*, 6159459. [CrossRef]
10. Majerník, M.; Daneshjo, N.; Malega, P.; Drábik, P.; Barilová, B. Sustainable development of the intelligent industry from Industry 4.0 to Industry 5.0. *Adv. Sci. Technol. Res. J.* **2022**, *16*. [CrossRef]
11. Yıldırım, M.; Okay, F.Y.; Özdemir, S. Big data analytics for default prediction using graph theory. *Expert Syst. Appl.* **2021**, *176*, 114840. [CrossRef]
12. Bi, W.; Liang, Y. Risk Assessment of Operator's Big Data Internet of Things Credit Financial Management Based on Machine Learning. *Mob. Inf. Syst.* **2022**, *2022*, 5346995. [CrossRef]
13. Hariri, R.H.; Fredericks, E.M.; Bowers, K.M. Uncertainty in big data analytics: Survey, opportunities, and challenges. *J. Big Data* **2019**, *6*, 44. [CrossRef]
14. Chen, Z.; Chen, W.; Shi, Y. Ensemble learning with label proportions for bankruptcy prediction. *Expert Syst. Appl.* **2020**, *146*, 113155. [CrossRef]
15. Fan, S.; Shen, Y.; Peng, S. Improved ML-based technique for credit card scoring in internet financial risk control. *Complexity* **2020**, *2020*, 8706285. [CrossRef]
16. García, V.; Marques, A.I.; Sánchez, J.S. Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction. *Inf. Fusion* **2019**, *47*, 88–101. [CrossRef]
17. Wang, M.; Yang, H. Research on personal credit risk assessment model based on instance-based transfer learning. In Proceedings of the Intelligence Science III: 4th IFIP TC 12 International Conference, ICIS 2020, Durgapur, India, 24–27 February 2021; Revised Selected Papers 4; Springer: Dordrecht, The Netherlands, 2021; pp. 159–169.
18. Teles, G.; Rodrigues, J.J.; Rabêlo, R.A.; Kozlov, S.A. Comparative study of support vector machines and random forests machine learning algorithms on credit operation. *Softw. Pract. Exp.* **2021**, *51*, 2492–2500. [CrossRef]
19. Orlova, E.V. Decision-making techniques for credit resource management using machine learning and optimization. *Information* **2020**, *11*, 144. [CrossRef]
20. Zou, Y.; Gao, C.; Gao, H. Business failure prediction based on a cost-sensitive extreme gradient boosting machine. *IEEE Access* **2022**, *10*, 42623–42639. [CrossRef]
21. Fritz-Morgenthal, S.; Hein, B.; Papenbrock, J. Financial risk management and explainable, trustworthy, responsible AI. *Front. Artif. Intell.* **2022**, *5*, 779799. [CrossRef]
22. Sun, M.; Li, Y. Credit Risk Simulation of Enterprise Financial Management Based on Machine Learning Algorithm. *Mob. Inf. Syst.* **2022**, *2022*, 9007140. [CrossRef]
23. Mousavi, M.M.; Lin, J. The application of PROMETHEE multi-criteria decision aid in financial decision making: Case of distress prediction models evaluation. *Expert Syst. Appl.* **2020**, *159*, 113438. [CrossRef]
24. Zhao, L.; Yang, S.; Wang, S.; Shen, J. Research on PPP Enterprise Credit Dynamic Prediction Model. *Appl. Sci.* **2022**, *12*, 10362. [CrossRef]
25. Pandey, M.K.; Mittal, M.; Subbiah, K. Optimal balancing & efficient feature ranking approach to minimize credit risk. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100037.
26. Plawiak, P.; Abdar, M.; Acharya, U.R. Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Appl. Soft Comput.* **2019**, *84*, 105740. [CrossRef]

27. Cho, S.H.; Shin, K.s. Feature-Weighted Counterfactual-Based Explanation for Bankruptcy Prediction. *Expert Syst. Appl.* **2023**, *216*, 119390. [\[CrossRef\]](#)
28. Bao, W.; Lianju, N.; Yue, K. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Syst. Appl.* **2019**, *128*, 301–315. [\[CrossRef\]](#)
29. Mitra, R.; Goswami, A.; Tiwari, M.K. Financial supply chain analysis with borrower identification in smart lending platform. *Expert Syst. Appl.* **2022**, *208*, 118026. [\[CrossRef\]](#)
30. Jemai, J.; Zarrad, A. Feature Selection Engineering for Credit Risk Assessment in Retail Banking. *Information* **2023**, *14*, 200. [\[CrossRef\]](#)
31. Chen, S.F.; Chakraborty, G.; Li, L.H. Feature selection on credit risk prediction for peer-to-peer lending. In Proceedings of the New Frontiers in Artificial Intelligence: JSAI-isAI 2018 Workshops, JURISIN, AI-Biz, SKL, LENLS, IDAA, Yokohama, Japan, 12–14 November 2018; Revised Selected Papers; Springer: Dordrecht, The Netherlands, 2019; pp. 5–18.
32. Si, Z.; Niu, H.; Wang, W. Credit Risk Assessment by a Comparison Application of Two Boosting Algorithms. In *Fuzzy Systems and Data Mining VIII*; IOS Press: Amsterdam, The Netherlands, 2022; pp. 34–40.
33. Merćep, A.; Mrćela, L.; Birov, M.; Kostanjčar, Z. Deep neural networks for behavioral credit rating. *Entropy* **2020**, *23*, 27. [\[CrossRef\]](#)
34. Bussmann, N.; Giudici, P.; Marinelli, D.; Papenbrock, J. Explainable machine learning in credit risk management. *Comput. Econ.* **2021**, *57*, 203–216. [\[CrossRef\]](#)
35. Moscato, V.; Picariello, A.; Sperlí, G. A benchmark of machine learning approaches for credit score prediction. *Expert Syst. Appl.* **2021**, *165*, 113986. [\[CrossRef\]](#)
36. Ariza-Garzón, M.J.; Arroyo, J.; Caparrini, A.; Segovia-Vargas, M.J. Explainability of a machine learning granting scoring model in peer-to-peer lending. *IEEE Access* **2020**, *8*, 64873–64890. [\[CrossRef\]](#)
37. Chen, X.; Li, S.; Xu, X.; Meng, F.; Cao, W. A novel GSCI-based ensemble approach for credit scoring. *IEEE Access* **2020**, *8*, 222449–222465. [\[CrossRef\]](#)
38. Dumitrescu, E.; Hué, S.; Hurlin, C.; Tokpavi, S. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *Eur. J. Oper. Res.* **2022**, *297*, 1178–1192. [\[CrossRef\]](#)
39. Li, D.; Li, L. Research on Efficiency in Credit Risk Prediction Using Logistic-SBM Model. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5986295. [\[CrossRef\]](#)
40. Ma, X.; Lv, S. Financial credit risk prediction in internet finance driven by machine learning. *Neural Comput. Appl.* **2019**, *31*, 8359–8367. [\[CrossRef\]](#)
41. Karn, A.L.; Sachin, V.; Sengan, S.; Gandhi, I.; Ravi, L.; Sharma, D.K.; Subramaniaswamy, V. Designing a Deep Learning-Based Financial Decision Support System for Fintech to Support Corporate Customer's Credit Extension. *Malays. J. Comput. Sci.* **2022**, *2022*, 116–131. [\[CrossRef\]](#)
42. Zheng, B. Financial default payment predictions using a hybrid of simulated annealing heuristics and extreme gradient boosting machines. *Int. J. Internet Technol. Secur. Trans.* **2019**, *9*, 404–425. [\[CrossRef\]](#)
43. Mancisidor, R.A.; Kampffmeyer, M.; Aas, K.; Jenssen, R. Learning latent representations of bank customers with the variational autoencoder. *Expert Syst. Appl.* **2021**, *164*, 114020. [\[CrossRef\]](#)
44. Wang, T.; Liu, R.; Qi, G. Multi-classification assessment of bank personal credit risk based on multi-source information fusion. *Expert Syst. Appl.* **2022**, *191*, 116236. [\[CrossRef\]](#)
45. Liu, W.; Fan, H.; Xia, M.; Pang, C. Predicting and interpreting financial distress using a weighted boosted tree-based tree. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105466. [\[CrossRef\]](#)
46. Andrade Mancisidor, R.; Kampffmeyer, M.; Aas, K.; Jenssen, R. Deep generative models for reject inference in credit scoring. *Knowl.-Based Syst.* **2020**, *196*, 105758. [\[CrossRef\]](#)
47. Wu, Z. Using machine learning approach to evaluate the excessive financialization risks of trading enterprises. *Comput. Econ.* **2021**, *59*, 1607–1625. [\[CrossRef\]](#)
48. Liu, J.; Zhang, S.; Fan, H. A two-stage hybrid credit risk prediction model based on XGBoost and graph-based deep neural network. *Expert Syst. Appl.* **2022**, *195*, 116624. [\[CrossRef\]](#)
49. Shu, R. *Deep Representations with Learned Constraints*; Stanford University: Stanford, CA, USA, 2022.
50. Tripathi, D.; Edla, D.R.; Kuppili, V.; Bablani, A. Evolutionary extreme learning machine with novel activation function for credit scoring. *Eng. Appl. Artif. Intell.* **2020**, *96*, 103980. [\[CrossRef\]](#)
51. Uj, A.; Nmb, E.; Ks, C.; Skl, D. Financial crisis prediction model using ant colony optimization-ScienceDirect. *Int. J. Inf. Manag.* **2020**, *50*, 538–556.
52. Feng, Y. Bank Green Credit Risk Assessment and Management by Mobile Computing and Machine Learning Neural Network under the Efficient Wireless Communication. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 3444317. [\[CrossRef\]](#)
53. Tian, J.; Li, L. Digital universal financial credit risk analysis using particle swarm optimization algorithm with structure decision tree learning-based evaluation model. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 4060256. [\[CrossRef\]](#)
54. Chrościcki, D.; Chlebus, M. The Advantage of Case-Tailored Information Metrics for the Development of Predictive Models, Calculated Profit in Credit Scoring. *Entropy* **2022**, *24*, 1218. [\[CrossRef\]](#)
55. de Castro Vieira, J.R.; Barboza, F.; Sobreiro, V.A.; Kimura, H. Machine learning models for credit analysis improvements: Predicting low-income families' default. *Appl. Soft Comput.* **2019**, *83*, 105640. [\[CrossRef\]](#)

56. Li, Z.; Zhang, J.; Yao, X.; Kou, G. How to identify early defaults in online lending: A cost-sensitive multi-layer learning framework. *Knowl.-Based Syst.* **2021**, *221*, 106963. [\[CrossRef\]](#)
57. Koç, O.; Başer, F.; Kestel, S.A. Credit Risk Evaluation Using Clustering Based Fuzzy Classification Method. *Expert Syst. Appl.* **2023**, *223*, 119882.
58. Rishchchi Fayyaz, M.; Rasouli, M.R.; Amiri, B. A data-driven and network-aware approach for credit risk prediction in supply chain finance. *Ind. Manag. Data Syst.* **2021**, *121*, 785–808. [\[CrossRef\]](#)
59. Muñoz-Cancino, R.; Bravo, C.; Ríos, S.A.; Graña, M. On the combination of graph data for assessing thin-file borrowers' creditworthiness. *Expert Syst. Appl.* **2023**, *213*, 118809. [\[CrossRef\]](#)
60. Li, Y.; Stasinakis, C.; Yeo, W.M. A hybrid XGBoost-MLP model for credit risk assessment on digital supply chain finance. *Forecasting* **2022**, *4*, 184–207. [\[CrossRef\]](#)
61. Haro, B.; Ortiz, C.; Armas, J. Predictive Model for the Evaluation of Credit Risk in Banking Entities Based on Machine Learning. In *Brazilian Technology Symposium*; Springer: Dordrecht, The Netherlands, 2018; pp. 605–612.
62. Qian, H.; Wang, B.; Yuan, M.; Gao, S.; Song, Y. Financial distress prediction using a corrected feature selection measure and gradient boosted decision tree. *Expert Syst. Appl.* **2022**, *190*, 116202. [\[CrossRef\]](#)
63. Alam, T.M.; Shaukat, K.; Hameed, I.A.; Luo, S.; Sarwar, M.U.; Shabbir, S.; Li, J.; Khushi, M. An investigation of credit card default prediction in the imbalanced datasets. *IEEE Access* **2020**, *8*, 201173–201198. [\[CrossRef\]](#)
64. Song, Y.; Peng, Y. A MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction. *IEEE Access* **2019**, *7*, 84897–84906. [\[CrossRef\]](#)
65. Biswas, N.; Mondal, A.S.; Kusumastuti, A.; Saha, S.; Mondal, K.C. Automated credit assessment framework using ETL process and machine learning. *Innov. Syst. Softw. Eng.* **2022**, 1–14. [\[CrossRef\]](#)
66. Wang, Y. Research on supply chain financial risk assessment based on blockchain and fuzzy neural networks. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5565980. [\[CrossRef\]](#)
67. Machado, M.R.; Karray, S. Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Syst. Appl.* **2022**, *200*, 116889. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.