Review

Machine Learning Applied to Banking Supervision a Literature Review

Pedro Guerra [1,*,†] and Mauro Castelli [2]

[1] Prudencial Supervision Department, Banco de Portugal, Rua Castilho 24, 1269-179 Lisbon, Portugal
[2] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal; mcastelli@novaims.unl.pt
* Correspondence: paguerra@bportugal.pt
† This disclaimer informs readers that the views, thoughts, and opinions expressed in the text belong solely to the authors, and not necessarily to Banco de Portugal.

Abstract: Machine learning (ML) has revolutionised data analysis over the past decade. Like innumerous other industries heavily reliant on accurate information, banking supervision stands to benefit greatly from this technological advance. The objective of this review is to provide a comprehensive walk-through of how the most common ML techniques have been applied to risk assessment in banking, focusing on a supervisory perspective. We searched Google Scholar, Springer Link, and ScienceDirect databases for articles including the search terms "machine learning" and ("bank" or "banking" or "supervision"). No language, date, or Journal filter was applied. Papers were then screened and selected according to their relevance. The final article base consisted of 41 papers and 2 book chapters, 53% of which were published in the top quartile journals in their field. Results are presented in a timeline according to the publication date and categorised by time slots. Credit risk assessment and stress testing are highlighted topics as well as other risk perspectives, with some references to ML application surveys. The most relevant ML techniques encompass k-nearest neighbours (KNN), support vector machines (SVM), tree-based models, ensembles, boosting techniques, and artificial neural networks (ANN). Recent trends include developing early warning systems (EWS) for bankruptcy and refining stress testing. One limitation of this study is the paucity of contributions using supervisory data, which justifies the need for additional investigation in this field. However, there is increasing evidence that ML techniques can enhance data analysis and decision making in the banking industry.

Keywords: banking; supervision; risk assessment; machine learning; EWS

1. Introduction

Decision support systems had their genesis in the 1960s (Burstein et al. 2008). Perhaps because of the exposure risk and magnitude of revenues generated, the financial sector has been a particularly avid driver for developing these technologies.

Predicting how financial institutions will perform and whether they will create value is key for every contender in this field—financial institutions, central banks, consultancy companies, and academia. Consequently, the use of new technology and methods to support risk assessment tasks (fin-tech) is a rising trend in this sector (Milian et al. 2019). In recent years, machine learning (ML) methods and, to some extent, deep learning (DL), have been used for the assessment of credit risk, and more broadly, predicting bank failures. Currently, traditional statistical methods are still commonly used for this purpose. Nevertheless, machine learning techniques are overcoming traditional approaches by allowing practitioners to module past decisions, exploit them for other scenarios, and predict future chaotic phenomena.

This review intends to provide a comprehensive picture of how machine learning techniques have been used so far in risk assessment from a central bank's perspective.

Thus, the scope of this work encompasses credit institutions and investment firms since those are the ones the European Banking Authority (EBA) regulation focuses on (European Banking Authority 2013). Henceforth, the term *institutions* will be used to refer to both.

The above-mentioned regulation establishes the standardisation of reporting requirements under the Single Supervisory Mechanism (SSM) (European Commission 2015). As a consequence, this study focuses on the European banking sector. Although we are aware of the importance of insurance, pension funds, securities, and markets in the financial sector, these are subject to different regulations and would benefit from a dedicated study. This work intends to contribute to several stakeholders in the supervisory landscape:

1. Institutions can have a comprehensive perspective on which risk assessment approaches are available and how they can evaluate their own exposures.
2. Central banks can acquire an integrated view of several validated methodologies for risk assessment. These can be the pillars of their next decision support systems by laying down the technologies supporting risk assessment processes. Furthermore, this work can also incite surveys and case studies on the use and adoption of ML at central banks.
3. Consultancy companies will benefit from a compendium of ML techniques and risk measures, to better support their clients.
4. Academia receives an important contribution that gathers an extensive number of papers on risk assessment and collates the identified methodologies from a supervisory perspective. This will hopefully serve as a stepping stone for future developments in this area, and provide a baseline for testing new methodologies.

This paper is organised as follows: it starts by justifying the methodology and describing how the references were selected. The results section gathers similarities among published scientific knowledge and presents the most relevant works that influence this field. The last section provides a space for discussing lessons learned and future work.

2. Methodology

This research was conducted through a series of exploratory steps on the topics of machine learning, banking, risk assessment, and banking supervision. The initial objective was to evaluate how machine learning techniques were being used at central banks. Additionally, we intended to analyse how these methods were informing the analytical capabilities of supervisors. We then refined a search query broad enough to return a set of articles we could work on. The following subsections describe a step-by-step guide for the reference search and selection.

2.1. Engines

This literature review relies on three search engines: Springer Link, ScienceDirect, and Google Scholar, queried until June 2021. The first and second search engines are extensively renowned for their trustworthiness and for selecting top journals for their results. The last one provides an extensive overview of all articles published in English (Gusenbauer 2019).

2.2. Query

Through extensive addition and diversification of search terms, we refined the search query to the following: *"machine learning" and ("bank" or "banking" or "supervision")*.

The underlying reasoning is that machine learning techniques are the focal point of this review article. The added value comes from analysing their potential applications to the banking sector, specifically banking supervision. No limitation concerning the year of publication was applied. Overlapping results are addressed in our secondary analysis. Furthermore, no filter regarding type or place of publication was applied, since the included papers' journals of publication were evaluated and classified after screening. Additionally, to keep up with new publications, we defined an alert in *Google Scholar* with this query. Finally, we pay close attention to Mendeley's alerts for articles related to the set gathered in this review.

2.3. Steps

The following subsections detail every step of the selection process summarise in the following PRISMA diagram Figure 1.
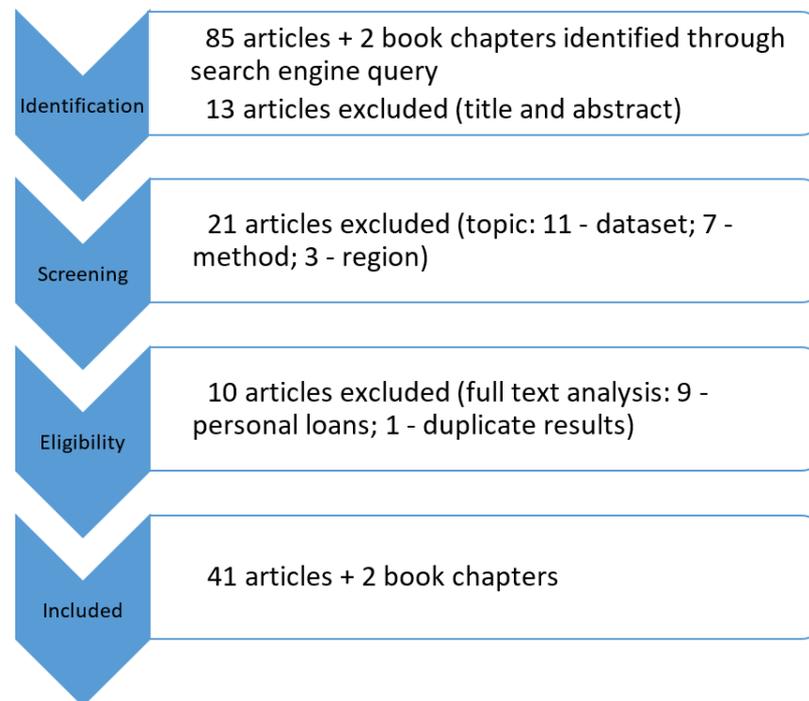
Figure 1. PRISMA diagram detailing the selection process of the identified articles.

2.3.1. Identification

The research query identified 85 articles and two books, from the three search engines. All the papers were published in English, in several different journals, and spanned from 2000 to 2021. This first step involved title and abstract analysis, and excluded 14 articles for lack of relevance.

2.3.2. Screening

In this phase, the main topics of each article were analysed, resulting in the exclusion of 21 papers, based on the following criteria:

- Dataset: when the analysed paper used data other than the banking sector, it was discarded. We are aware that applications of ML to the stock market are a trendy topic in the literature, and that the insurance and pension funds sector is of great importance in the Eurozone. Nevertheless, the regulation is substantially different, and they would merit from a different study and approach;
- Methodology: risk assessment exercises are historically based on quantitative data, combined with expert judgment. Furthermore, it is the quantitative data that holds the largest amount of information regarding risk exposure practices. Therefore, we focus our analysis on quantitative methods, for which a risk assessment classification has already been assigned (leveraging on previous knowledge through supervised learning). We thus excluded works concerning unsupervised learning methods, or sentiment analysis (qualitative);
- Region: this criterion is closely related to the first, since regulation changes according to geography. We chose to focus mainly on works based upon institutions operating in the Eurozone. Nonetheless, relevant works by other central banks were considered eligible.

2.3.3. Eligibility

The next step required a thorough analysis of each paper, to verify its sources and classify the journal it was published in (quartile of impact). Papers were analysed from 2021 backward to identify any overlapping results or new or improved methodologies, resulting in the exclusion of ten more articles: nine being personal loans related and one duplicate result.

The scope of this review is the application of ML techniques to risk assessment from a supervisory perspective, which includes at best how institutions are addressing their risk assessment exercises. The data and predictors used to evaluate an individual credit application (personal loan) differ substantially from the data used by banks from a corporate perspective, and even more from the data collected in the regulatory context. As such, works regarding credit risk for individual applicants were also excluded.

2.3.4. Considered Papers

The final article base consists of 41 papers and two books, published from 2000 until 2021, selected through the steps mentioned. In the next section, we will describe the similarities among the papers, as well as the methods applied and respective banking areas. Table 1 lists the selected papers, providing a single-sentence summary of their content.

3. Results

3.1. Distribution

Based on the reviewed works from the previous section, the following paragraphs describe how machine learning techniques have been used in the banking sector. Our research intends to provide a future reference on how these technologies address and support the risk assessment process, in particular from a central bank's perspective. These results solely reflect the analysis of the papers selected for this review. They represent neither the total of publications throughout these years nor the distribution of topics for all publications.

Table 2 summarises the selected articles, referenced by author, year of publication, affiliation and number of citations. Additionally, Table 3 lists the journals from the selected articles.

Table 1. Short summary of each analysed paper, referenced by authors and year.

Authors	Year	Summary Sentence
Galindo et al.	2000	CART decision-trees out-perform statistics for credit risk assessment, using a commercial bank loans dataset
Hillegeist et al.	2004	Black–Scholes–Merton option-pricing model is a better indicator of bankruptcy probability than Z-Score and O-Score.
Min et al.	2005	Motivated by the increasing use of machine learning techniques, this paper aims to outperform classical statistics in bankruptcy prediction. An optimised SVM model performs better than MDA, logit and BPN for bankruptcy prediction.
Angelini et al.	2008	Regulation-imposed capital requirements increase the need for precise credit risk assessment systems. This paper shows ANNs' very good results predicting the default tendency of a borrower.
Boyacioglu et al.	2009	Multi-layer perceptrons and learning vector quantization are the most successful models predicting bank failure as a classification problem, in a Turkish case.
Chaudhuri et al.	2011	Fuzzy-SVM satisfies Basel II demands for detecting bankruptcy probability, outperforming other approaches. This algorithm also proved to have more clustering capabilities than PNN.

Table 1. *Cont.*

Authors	Year	Summary Sentence
Hammer et al.	2012	The logical analysis of data (LAD) is able to reverse-engineer Fitch risk ratings of bank, showing better results than support-vector machines and logistic regression when evaluating the creditworthiness of banks.
Ribeiro et al.	2012	This study establishes the limitations of using exclusively quantitative financial data when developing default risk models. The authors propose a new approach that includes contextual knowledge in an SVM model, showing better predictability performance t
Lopez Iturriaga et al.	2015	Profiling distressed banks using self-organising maps and modelling failure detection with multi-layer perceptron outperforms traditional models of bankruptcy prediction. The resulting model detects 96% of failures, up to 3 years before the bankruptcy ev
Ala'raj et al.	2016	The proposed hybrid ensemble model improves predicting capability compared to base classifiers, using 7 real-world datasets. It uses a classifier consensus system to compare this new approach with the traditional combination methods.
Abellan et al.	2017	Selection of the best base classifier in ensemble methods for credit scoring problems. The individual performance of classifiers is not the only criteria for ensemble schemes.
Chakraborty et al.	2017	An overview of the applications of machine learning to financial problems, the most popular modelling approaches, and three case studies of relevant works for central banks. This study also establishes that machine learning models usually outperform tradi
Pompella et al.	2017	An EWS is proposed to detect likely-to-fail banks. This method is compared with risk agencies' rating and detects possibly wrongly rated banks. The authors suggest the adoption of this EWS by regulators.
Xia et al.	2017	The credit scoring problem is addressed using a XGBoost model with Bayesian hyper-parameter optimisation, not only obtaining better accuracy than baseline models, but also providing feature importance and a decision chart for interpretability.
Alessi et al.	2018	The use of random forest to predict banking crises secondary to excessive credit growth, using credit and real estate predictors.
Broeders et al.	2018	A survey on the use of innovative technologies in financial supervision, the challenges faced by supervisory agencies and the need for a clear suptech strategy. Additionally, the experience of early adopters is described.
Chang et al.	2018	The development of a credit risk model using XGBoost classifier to address the heterogeneous nature of financial data. An under-sampling method is applied to deal with the imbalanced data.
Gogas et al.	2018	Outperforming the Ohlson's score with stress-testing tool based on a support-vector machine model to forecast bank failures. The adopted methodology defines a clear boundary between solvent and insolvent banks.
Jagtiani et al.	2018	The impact of machine learning in banking supervision in terms of new possible analytical solutions and risks involved in those new approaches.
Kupiec et al.	2018	Addressing the need for validation of bank stress test models, by emphasising model forecast accuracy. A Lasso model shows the best forecasting capabilities for determining capital requirements in stressful conditions.

Table 1. *Cont.*

Authors	Year	Summary Sentence
Le et al.	2018	Artificial neural networks and k-nearest neighbour methods are more accurate for predicting bank failure than traditional statistics.
Petropoulos et al.	2018	Predicting the probability of default of Greek banks using data mining techniques to reduce dimensionality, with XGBoost emerging as the best model. The authors aim to fully capture the information within these large datasets to better support the overall
Tavana et al.	2018	Addressing liquidity risk assessment through a model that uses neural networks and Bayesian networks. The models were capable of distinguishing the most critical factors in liquidity risk measurement.
Climent et al.	2019	Using XGBoost to identify the best predictors of bank failure and develop a classification model to label failed and non-failed banks in the Eurozone. The data used in this study is composed of 25 annual financial ratios for commercial banks in the Eurozo
Dwivedi et al.	2019	Expert contributors identify and compile a series of opportunities, impacts and research topics raised by the rapid adoption of AI. The financial sector shows enormous potential in robot advisory and automation, and bankruptcy prediction.
Hohl et al.	2019	A survey of activities within the scope of suptech, classifying the degree of technological development, and the strategies in place to implement them, highlighting the experimental nature of these initiatives and the need for international coordination.
Kolari et al.	2019	Successfully undergoing European bank stress-tests depends largely on the risks a bank is exposed to, as opposed to being prepared for specific adverse scenarios. Using Bankscope data, the developed model accurately predicts 90% of the failing banks.
Kou et al.	2019	A survey depicting the most common methodologies to assess systemic risk in the financial system, using machine learning, big data analysis, network analysis and sentiment analysis. The paper showcases current researches on the use of machine learning in
Leo et al.	2019	A literature review evidencing machine learning use for risk management purposes in the banking industry, while also noting the experimental nature of most approaches.
Milian et al.	2019	A literature review aiming to find consensus on a fintech definition, showing how banks and supervisory agencies are using these innovative technologies and dealing with the risks involved.
Soui et al.	2019	Using evolutionary algorithms to address credit risk assessment by considering it as an optimisation (rule-based) search problem: minimise complexity, maximise accuracy and weight (rules importance).
Alonso et al.	2020	Comparing machine learning models from credit default prediction. Necessity for a structured strategy for assessing ML models to increase transparency in the use of these technologies, and promote innovation in the financial industry.
Dastile et al.	2020	A systematic literature review on how statistic and machine learning techniques have been used to address the credit scoring problem. Although machine learning is often incapable of explaining predictions, these models consistently outperform the classic
Filippopoulou et al.	2020	Developing an EWS to detect systemic banking crisis based on the ECB Macroprudential database. Most of the risk indicators used in the dataset are key to forecast a systemic risk crisis 1 to 4 years before the event.

Table 1. *Cont.*

Authors	Year	Summary Sentence
Giudice et al.	2020	Developing an automatic classification system for the sector of economic activity for Italian companies, using a multi-step classifier with gradient boosting and support-vector machine models. The developed model is already being used in a production envi
Lee et al.	2020	A study on types of machine learning applications, exploring the accuracy-interpretability trade-off, and three use cases in financial industry.
Alonso et al.	2021	Predicting credit default probability with machine learning surpasses traditional statistic methods, potentially leading to savings of up to 17% in regulatory capital requirements.
Antunes	2021	Establishing the need for supervisory on-site inspection by comparing the results of two machine learning models, one based on the banks' own risk assessment and the other based on the findings from previous on-site inspections.
Doerr et al.	2021	Policy brief showing central banks are relying on big data for daily tasks, and identifying a clear need for specialised knowledge on how to adequately use machine learning, and extract greater value from that data.
Huang et al.	2021	This study is developed under the assumption that the intricate nature of financial data cannot be properly explored through traditional methods. An advanced deep learning model to address the complex and hierarchical features of financial data, that outperforms traditional methods and other advanced approaches.
Wang et al.	2021	Random forest based EWS outperforms the classic logit approach as the predictive tool to prevent systemic banking crises. This paper shows an expert voting approach to model the multivariate nature of systemic risk assessment data.

Table 2. List of papers collected through the research query, referenced by author, year of publication, affiliation, and number of citations.

Authors	Year	Affiliation	Title	Citations
Abellan et al.	2017	academia	A comparative study on base classifiers in ensemble methods for credit scoring	88
Ala'raj et al.	2016	academia	A new hybrid ensemble credit scoring model based on classifiers consensus system approach	66
Alessi et al.	2018	central bank	Identifying excessive credit growth and leverage	135
Alonso et al.	2020	central bank	Machine Learning in Credit Risk: Measuring the Dilemma Between Prediction and Supervisory Cost	1
	2021	central bank	Understanding the Performance of Machine Learning Models to Predict Credit Default: A Novel Approach for Supervisory Evaluation	0
Angelini et al.	2008	academia	A neural network approach for credit risk evaluation	305

Table 2. *Cont.*

Authors	Year	Affiliation	Title	Citations
Antunes	2021	central bank	To supervise or to self-supervise: A machine learning based comparison on credit supervision	0
Boyacioglu et al.	2009	academia	Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey	272
Broeders et al.	2018	industry	FSI Insights Innovative technology in financial supervision	23
Chakraborty et al.	2017	central bank	Machine Learning at Central Banks	62
Chang et al.	2018	academia	Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions	17
Chaudhuri et al.	2011	academia	Fuzzy Support Vector Machine for bankruptcy prediction	155
Climent et al.	2019	academia	Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach	10
Dastile et al.	2020	academia	Statistical and machine learning models in credit scoring: A systematic literature survey	0
Doerr et al.	2021	industry	How do central banks use big data and machine learning?	0
Dwivedi et al.	2019	academia	Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy	39
Filippopoulou et al.	2020	academia	An early warning system for predicting systemic banking crises in the Eurozone: A logit regression approach	1
Galindo et al.	2000	academia	Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications	213
Giudice et al.	2020	central bank	Institutional Sector Classifier, a Machine Learning Approach	0
Gogas et al.	2018	academia	Forecasting bank failures and stress testing: A machine learning approach	20
Hammer et al.	2012	academia	A logical analysis of banks' financial strength ratings	49
Hillegeist et al.	2004	academia	Assessing the probability of bankruptcy	1393
Hohl et al.	2019	industry	FSI Insights on policy implementation The suptech generations	3
Huang et al.	2021	academia	Intelligent FinTech Data Mining by Advanced Deep Learning Approaches	0
Jagtiani et al.	2018	central bank	The Roles of Big Data and Machine Learning in Bank Supervision	4

Table 2. *Cont.*

Authors	Year	Affiliation	Title	Citations
Kolari et al.	2019	academia	Predicting European bank stress tests: Survival of the fittest	4
Kou et al.	2019	academia	Machine learning methods for systemic risk analysis in financial sectors	47
Kupiec et al.	2018	industry	On the accuracy of alternative approaches for calibrating bank stress test models	5
Le et al.	2018	academia	Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios	24
Lee et al.	2020	academia	Machine learning for enterprises: Applications, algorithm selection, and challenges	7
Leo et al.	2019	(blank)	Machine learning in banking risk management: A literature review	11
Lopez Iturriaga et al.	2015	academia	Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks	129
Milian et al.	2019	academia	Fintechs: A literature review and research agenda	31
Min et al.	2005	academia	Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters	866
Petropoulos et al.	2018	central bank	A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting	6
Pompella et al.	2017	academia	Ratings based Inference and Credit Risk: Detecting likely-to-fail Banks with the PC-Mahalanobis Method	5
Ribeiro et al.	2012	academia	Enhanced default risk models with SVM+	57
Soui et al.	2019	academia	Rule-based credit risk assessment model using multi-objective evolutionary algorithms	3
Tavana et al.	2018	academia	An Artificial Neural Network and Bayesian Network model for liquidity risk assessment in banking	30
Wang et al.	2021	academia	A machine learning-based early warning system for systemic banking crises	2
Xia et al.	2017	academia	A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring	158

The most common topic on these papers is credit risk related (nearly 34% of references), as shown in Figure 2.

The second major category relates to "ML application" (surveys, fin-tech and sup-tech, as per the division suggested by Broeders and Prenio (2018), the use of innovative technologies by supervisory agencies to support their processes) along with "stress tests". The remainder of the results focuses either on "bank risk" more broadly, or on specific topics for supervision such as liquidity risk and other banking risk perspectives. Another relevant aspect is the publication date of these articles, ranging from 2000 to 2021 and distributed as shown in Figure 3.

Table 3. Journals of the selected articles and their quartile. Where the journal is not indexed, the entity responsible for the publishing was included.

Quartile/Origin	Journal	Number of Papers
A (ERA)	Advances in Neural Information Processing Systems	1
Banca d'Italia	Questioni di Economia e Finanza	1
Banco de España	SSRN Electronic Journal	2
Bank for International Settlements	FSI Insights on policy implementation	2
Bank of England	Bank of England	1
Bank of Greece	Ninth IFC Conference on "Are post-crisis statistical initiatives completed?"	1
Federal Reserve	Banking Perspectives, Forthcoming	1
Q1	Applied Soft Computing Journal	3
	Business Horizons	1
	Electronic Commerce Research and Applications	1
	Expert Systems with Applications	9
	International Journal of Forecasting	1
	International Journal of Information Management	1
	Journal of Business Research	1
	Journal of Economic Behavior and Organization	1
	Journal of Financial Stability	2
	Neurocomputing	1
	Research in International Business and Finance	1
	Review of Accounting Studies	1
	Technological and Economic Development of Economy	1
Q2	Applied Economics	1
	Computational Economics	2
	Economic Modelling	1
	Financial Innovation	1
	Global Finance Journal	1
	Quarterly Review of Economics and Finance	1
	Risks	1
SUERF	SUERF—The European Money and Finance Forum	1
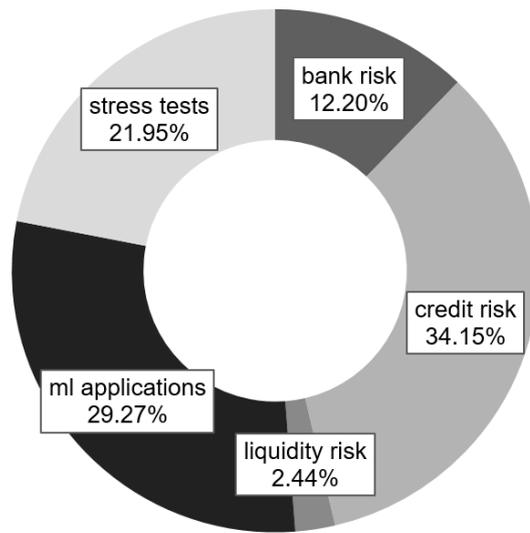
Figure 2. Distribution of articles according to main topic.
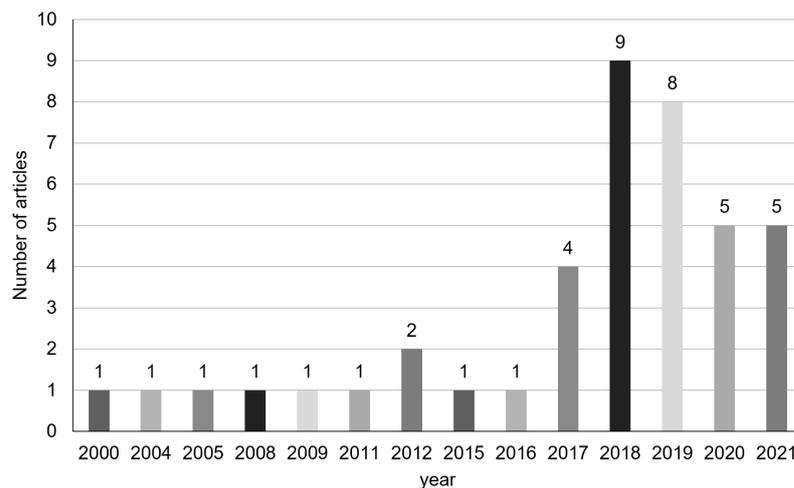
Figure 3. References according to year of publication.

Importantly, although ML applied to the financial sector has been present since 2000, by 2015 the intersection of these knowledge areas gained a huge interest. This translated to increasing numbers of publications in this field, with the majority of relevant articles in this study being published from 2017 onward. Table 4 lists the machine learning methods applied by each author as well as the datasets that supported each research.

Table 4. Machine learning methods applied in each paper and the respective dataset, referenced by authors.

Authors	ML Methods	Dataset
Abellan et al.	ada-boosting, bagging, random subspace, DECORATE, rotation forest	public: Australian, German, and Japanese datasets obtained from UCI repository of machine learning; Iranian dataset from "A comparison between statistical and data mining methods for credit scoring in case of limited available data. (2007)"; Polish datase

Table 4. *Cont.*

Authors	ML Methods	Dataset
Ala'raj et al.	neural networks, support vector machines, random forests, decision trees, Naive Bayes	public: Australian, German, and Japanese datasets obtained from UCI repository of machine learning; Iranian dataset from "A comparison between statistical and data mining methods for credit scoring in case of limited available data. (2007)"; Polish datase
Alessi et al.	logit, decision trees, random forest	public: crisis dataset by Detken et al. 2014, capturing systemic banking crises related to domestic credit cycle
Alonso et al.	logit, lasso, CART, random forest, xgboost, deep learning	private: anonymized dataset from Banco Santander, containing more than 75,000 credit operations
Alonso et al.	logit, lasso, CART, random forest, xgboost, deep learning, RL & ensemble methods	public: kaggle.com "Give me some credit" dataset
Angelini et al.	ann	private: SME loans from na Italian bank
Antunes	random forest	public: Central Bank of Brazil financial series repository
Boyacioglu et al.	Multi-layer perceptron, Competitive learning, Self-organizing map, Learning vector quantization, Support vector machines, Multivariate discriminant analysis, K-means cluster analysis, Logistic regression analysis	public: financial ratios using CAMELS system; annual publication "Banks Association of Turkey"
Broeders et al.	NA	NA
Chakraborty et al.	ann, dt, svm, clustering	NA
Chang et al.	logit, gmdh, svm, xgboost	private: credit data from a financial institution in Taiwan (2009–2016)
Chaudhuri et al.	logit, ann, svm, ga-svm, fuzzy-svm	private: dataset comprising American organizations with capitalization greater than $1 billion that filed for protection (2001–2002).
Climent et al.	xgboost	public: Orbis database (2006–2016)
Dastile et al.	LR (Logistic Regression), NB (Naïve Bayes), LDA (Linear Discriminant Analysis), XGB (XGBoost), EML (Extreme Learning Machines), k-NN (k-Nearest Neighbor), SVM (Support Vector Machine), ANN (Artificial Neural Network), BA (Bagging), BO (Boosting), RF (Rand)	NA
Doerr et al.	NA	NA
Dwivedi et al.	evolution	NA
Filippopoulou et al.	logit, ewm	public: Macroprudential Database by the ECB
Galindo et al.	probit, knn, dt, CART	private: loans from a commercial bank provided by Comision Nacional Bancaria y de Valores (Mexico's security exchange and banking commission)
Giudice et al.	svm, xgboost	private: Bank of Italy Entities Register
Gogas et al.	O-score, svm	public: US banks (2007–2013); 481 failed and 962 solvent banks (1443 in total).

Table 4. *Cont.*

Authors	ML Methods	Dataset
Hammer et al.	logit, svm, lad	public: 800 banks rated by Fitch along with 24 explanatory variables (2001).
Hillegeist et al.	logit, classic statistics	public: Moody's Default Risk Services' Corporate Default database and SDC Platinum Corporate Restructurings database (1980–2000)
Hohl et al.	evolution	NA
Huang et al.	deep CCAE, fuzzy rules, fuzzy rough nn, fuzzy nn, random tree, random forest	public: enterprise financial statement information from Taiwan securities market—Taiwan Economic Journal (2008–2013)
Jagtiani et al.	evolution, big data, ml	NA
Kolari et al.	AdaBoost, logit, ann, random forest, svm radial, svm linear	public: Bankscope database (2010, 2011 and 2014); 273 banks where 29 failed at least one stress test
Kou et al.	comparison	NA
Kupiec et al.	comparison; classic methods severely underestimate stress tests	public: quarterly financial data (balance sheet, income statements, etc.) from Federal Reserve Bank of St. Louis FRED economic database (1993–2011)
Le et al.	svm, ann, k-NN, linear discriminant analysis, logit	public: Bankscope database (2010–2016); 3000 US banks, 1438 failed, 1562 active. 31 ratios based on financial statements
Lee et al.	evolution	NA
Leo et al.	evolution	NA
Lopez Iturriaga et al.	mlp, som	public: 32 indicators extracted from financial statements—Federal Deposit Insurance Corporation between 2002 and 2012
Milian et al.	evolution	NA
Min et al.	mda, logit, svm, ann backpropagation	private: a Korean credit guarantee organization (2000–2002); 1888 institutions, 944 failed and 944 non-failed.
Petropoulos et al.	logit, LDA, XGBoost MXNET	private: Bank of Greece corporate loans database (2005–2015).
Pompella et al.	ewm	public: Bloomberg indicators extracted from balance sheet, income statement and others (solvency, performance, etc.) (2005 to 2014); 482 banks
Ribeiro et al.	svm, svm+, svm+mtl	public: Diane database by COFACE; financial statements of French companies from 2002 to 2006.
Soui et al.	"Non-dominated" Sorting Genetic Algorithm (NSGAII), multi-objective evolutionary algorithm based on decomposition (MOEA/D), multi-objective particle swarm optimisation (SMOPSO), Strength Pareto Evolutionary Algorithm (SPEA2)	public: German (1000 observations, 70% good applicants, 30% bad applicants, 20 features) and Australian (690 observations, 383 good applicants, 307 bad applicants, 14 features) datasets from University of California, Irvine;

Table 4. *Cont.*

Authors	ML Methods	Dataset
Tavana et al.	ANN, bayes	public: monthly reports on loan data provided by a large US bank (2005–2011); 353 observations, 10 features; balance sheet ratios
Wang et al.	logit, svm, adaboost, ann, random forest	public: yearly data for 95 economies with crisis data (1981–2017); 1690 observation of which 210 are crises; 11 features; dataset from Laeven and Valencia 2018, Global Financial Database and IMF International Financial Statistics
Xia et al.	AdaBoost, AdaBoost-NN, Bagging-DT, Bagging-NN, DT, LR, NN, RF, SVM, GBDT, XGBoost-MS, XGBoost-GS, XGBoost-RS, XGBoost-TPE	public: three datasets from UCI machine learning repository (German, Australian and Taiwan); two datasets from P2P lending platforms (Lending Club from the US and We.com from China)

3.2. Evolution

The selected papers were organised by date of publication. Publication intervals were defined based on relevant events in the banking sector, technological evolution, and the number of papers per interval. The first slot ranging from 2000 to 2011 encompasses the effects of the financial crisis of 1999 and 2008. The second range (from 2012 to 2016) still reflects several studies based on the 2008 crisis, but with a more mature insight. In this period there is also a trending increase of ANN models. The third slot encompasses the years of 2017–2018, which show a significant increase in publications intersecting ML and the banking sector.

The final interval (2019 to the current date) depicts important ML applications to the financial market in general. Studies in this period reveal an increased ponderation of the uses and impacts of machine learning in banking supervision, with several publications from banking authorities.

3.2.1. 2000–2011

Six papers were identified from this period. They mostly focus on stress tests although three of them engage on the topic of credit risk and default risk.

Early in this period, Galindo and Tamayo (2000) identified the risk assessment task as crucial for an efficient use of resources. They used an error curve methodology to compare model precision and concluded that tree-based models outperform ANNs, KNN and probit. This sets forward the finding that tree-based models are more appropriate to structured data, as opposed to ANNs.

Hillegeist et al. (2004) proposed a new method for assessing bankruptcy probability. Based on the Black–Scholes–Merton option-pricing model, this method was compared to the well-known Z-score (Altman 1968) and O-score (Ohlson 1980), obtaining superior results. These authors stressed the need for a standardised risk assessment measure mainly for comparability purposes.

Min and Lee (2005) presented a paper that compares statistical and artificial intelligence methods, with the latter outperforming the former in the classification of bankruptcy. Although this study focuses on credit risk assessment for heavy industry firms in Korea, we included it in our sample for a compelling reason. It is a clear example of machine learning methods outperforming conventional statistics and it uses a set of predictors (financial ratios) easily mapped to regulatory financial reporting since they are based on balance sheet entries. Angelini et al. (2008) based their work on the Basel II capital requirements and the need for a system to assess credit risk. The main objective of this work is to evaluate the possibility of using neural networks to estimate the probability of default of a

borrower (Italian small companies). In spite of some ANNs being used, the comparison of classic machine learning models to conventional statistical methods was the more recurrent approach. Furthermore, the risk definition used to evaluate the data sets was based on the probability of default. This is explained by the fact that the datasets are mostly from loan applications, either from small and medium enterprises or personal loans (housing included). These findings contradict Galindo and Tamayo (2000) as well as more recent developments in this area. ANNs have been proved to excel in time-series, image, and voice recognition, as opposed to their performance using structured data.

Additionally, some articles used financial ratios and CAMELS rating model (an international rating system used by regulatory banking authorities to rate financial institutions) to assess an institution's performance (stress testing and bankruptcy prediction). Assessing the health of a bank is crucial to prevent its failure and contain the systemic risk its failure or losses represent. The work of Boyacioglu et al. (2009) identifies this assessment as an original classification problem. The authors use the CAMELS method to select the most relevant predictors. Using this method, neural networks were shown to outperform multivariate statistical methods for a Turkish banking sector use case.

Chaudhuri and De (2011) considers Basel II definition of risk to select features for the models. In this case, ANNs are not as frequently used as other conventional ML techniques, such as support vector machines and k-nearest neighbours. As a consequence, the authors focus on the optimisation of those models to the problem at hand (i.e. nature of the dataset).

3.2.2. 2012–2016

In this period, articles mostly reflect the first insights gained from the 2008 financial crisis.

Having identified the lack of a comprehensive method to incorporate circumstantial aspects into the banking default risk predictive models, Ribeiro et al. (2012) reported that SVM+ outperformed other methods that did not include non-financial information. Hammer et al. (2012) showed that Logical Analysis of Data (LAD) is an accurate method by reverse-engineering Fitch risk ratings. The authors stated that LAD can be used as an internal rating system that is Basel compliant.

López Iturriaga and Sanz (2015) took a different approach to this matter. First, they used self-organising maps (SOM) to profile distressed banks. This unsupervised learning method is competitive so it thrives to reach the right pattern, the representation of bankruptcy for a bank. Afterward, the authors applied multi-layer perceptrons to assess a bank's risk in several time frames, obtaining very promising results predicting bankruptcy for commercial banks. This two-step approach is the first in this selection of papers to recognise the benefits of a pre-processing phase to map the bankruptcy layout of a bank. Although previous research has shown better results using conventional ML, the success shown by this perceptron model suggests it is adequate to model the time evolution of quantitative data.

A new approach to credit scoring using an ensemble model was proposed by Ala'raj and Abbod (2016). These authors combine several data filtering and feature selection methods before evaluating model performance, and compare the most traditional classifiers with their method. The results are validated on several public datasets and their accuracy assessed under several measures: average accuracy, area under the curve (AUC), H-measure, and Brier Score. This is the first paper in our sample showing that ensembles outperform single models for classification problems.

3.2.3. 2017–2018

These two years showed a more than 60% increase in publications in the intersection of ML and banking sector. As highlighted by Strydom and Buckley (2019), the technological evolution allowed for the development of deep learning (DL) models, as well as new ensemble methods like extreme gradient boosting (XGBoost). Although the DL's first

reappearance happened in 2012 (Zhang et al. 2020), its application to financial risk only came to light in 2016–2017.

Traditional ML and classical statistical approaches are still the cornerstones of most of these articles. However, an increasing trend is noticeable in the use of ANN-based models mainly due to bigger datasets and enhanced computing power.

Abellán and Castellano (2017) build on their previous work showing how ensembles achieve better results in credit risk assessment than single models, validating the findings of Ala'raj and Abbod (2016). The authors stress the importance of individual model performance as a criterion for ensemble selection. Although the authors emphasize their own tree-based model (Credal Decision Tree, CDT), the main finding of their work is the corroboration of the hypothesis that ensembles outperform single classifiers.

Prompted by the 2008 Global Financial Crisis and the need to foresee signals of financial instability, Italian authors Pompella and Dicanio (2017) developed an Early Warning System (EWS) to help uncover distress signs for banks. This credit risk model allows users to discriminate stable from likely-to-fail banks and might be useful in adjusting rating assignments by Rating Agencies. The authors suggest its implementation in regulators to support the supervisory process.

Xia et al. (2017) present an extreme gradient boosting model (XGBoost by Chen and Guestrin (2016)) that consistently outperforms baseline models. The authors stress the importance of model-based feature selection as well as the use of Bayesian hyper-parameter optimisation to achieve better predictive results. Although personal credit risk is not the main topic of interest in this review, this study shows the advantages of boosting techniques and the importance of an interpretable model for decision making. This type of models have won several Kaggle competitions and are consistently showing excellent results with structured data.

Chakraborty and Joseph (2017) from the Bank of England introduce a central bank perspective on machine learning and its applications. The authors provide an overview of machine learning models and model validation to support the presentation of three case studies. As a final note, this work acknowledges the amount of available data as an important vector in decision support systems based on machine learning at central banks and other offices. As previously stated, agency papers as this one are paramount in understanding the use of machine learning in these contexts, providing use cases and areas of interest for future work.

Alessi and Detken (2018) contribute with another EWS to detect excessive credit growth. This phenomenon is usually at the root of systemic risk to financial stability and its early detection can help avoid cases of bankruptcy. The authors use Random Forest classifier model with credit and real estate predictors. Their work pioneers in the domain of risk assessment from the perspective of central banks, thus setting peer practitioners in their future path. Moreover, the work reinforces that ensembles consistently outperform single models. Other authors successfully use extreme gradient boosting to develop a credit risk model for financial institutions (Chang et al. 2018). Those tools promise significant support (i.e. low error rate) for risk assessment in loans.

The Central Bank of Greece also provides a thorough analysis based on post-2008 crisis loan data from Greek banks, by Petropoulos et al. (2018). This study sets a milestone for the use of advanced ML techniques from a supervisory perspective. Furthermore, it leverages the resulting model to create an EWS that will support subsequent decisions in loan approval. Similar to what López Iturriaga and Sanz (2015) have shown, modeling a timeline evolution is where neural networks (in this case deep neural networks, DNN's) excel. Another important result is that DNNs can perform just as well as XGBoost, showcasing how precisely deep learning models adapt to structured data.

Tavana et al. (2018) present a study that directly addresses liquidity risk, which is the most rapidly devastating risk a bank is exposed to. In this paper, the authors present an artificial neural network model combined with a Bayesian network (BN) to assess liquidity risk using solvency as a proxy. This combined approach models the liquidity risk indicator

through the ANN and the probability of occurrence through the BN. The results show this approach distinguishes the most critical factors for liquidity in this dataset.

Broeders and Prenio (2018) conduct a study that compiles the experience of early users of innovative technology in financial supervision (sup-tech). The authors structure a definition of sup-tech and show how it is used for data collection and analytics. These two applications have different initiators in supervisory agencies. Data collection tends to be initiated by management decisions and projects whereas analytics usually start out as research questions or analysis queries from supervision units. A conductive thread of all use cases is the sharing of the experience of some early adopters and the impact those technologies are having on the organisation. Similar studies, such as the one conducted by Chakraborty and Joseph (2017) are essential for compiling, sharing, contrasting the several approaches throughout central banks and other agencies.

The Federal Reserve provides a broader perspective, analysing how the use of machine learning and big data will impact compliance aspects (Jagtiani et al. 2018). The authors also stress the need to identify the risks that these technologies carry when applied to the financial market.

Gogas et al. (2018) propose a methodology that separates solvent and failed banks, using machine learning models. The authors present an alternative tool for stress-testing that outperforms the O-score. Their approach is based on a support vector machine model that helps to define a boundary between solvent and insolvent banks, converting this issue into a classification problem. Kupiec (2018) presents a related study that stresses the need for new methodologies to validate conventional bank stress tests.

As a final reference for this period, Le and Viviani (2018) also tackle the problem of bank failure prediction using machine learning and classical financial ratios. One important aspect of this work is that the authors use ratios from 5 different risk perspectives: Loan quality, Capital quality, Operations efficiency, Profitability, and Liquidity. This work validates yet again that machine learning methods outperform traditional statistics. However, these authors do not explore the possibility of using ensembles, which have already been proven to be top performers in classification problems.

3.2.4. 2019–2021

Credit and banking risks are essential for a balanced economy; trying to prevent systemic repercussions stemming from them is considered of the utmost importance. Similarly to earlier periods, these risks maintain a privileged spot in research. Still, it was on ML application we saw the most significant increase in publications. This suggests the demand for coordination and a global perspective on the developments conquered so far in this area.

Leo et al. (2019) produce a thorough review on how machine learning has been used at banks for risk assessment. This paper offsets the industrial and academic claim for ML application *versus* real-life practices, highlighting a series of perspectives where risk management has been poorly applied. Climent et al. (2019) develop an insightful study that aims to identify a set of financial predictors that best model a bank's financial distress. To this end, the authors apply an XGBoost based model to a set of indicators that might predict a bank failure in the Eurozone. The set of selected indicators (Total assets, Loan loss provisions/net interest revenue, Equity/net loans and Interbank ratio) are shown to best help regulators monitor financial distress for those banks. From a technical perspective, this work reinforces the choice of XGBoost for classification problems using structured data. A recent study by Wang et al. (2021) deconstructs the use of logit as the base classifier for EWS developed to predict banking crisis. In fact, the authors use random forest classifier to simulate expert decision, obtaining a generalisation capability above 80% area under the curve (AUC).

Kou et al. (2019) compare several ongoing researches concerning the applications of machine learning methods to the detection of systemic risk events, that is, financial distress

phenomena that affect several markets or geographic regions. They also propose the use of big-data analysis to assess systemic risk.

Soui et al. (2019) address the issue of comprehensibility of machine learning models for credit risk assessment. Interestingly, in this study, interpretability was mentioned as one of the barriers for adopting ML models in day-to-day decision making. In an attempt to circumvent this problem, the authors proceeded to develop an evolutionary algorithm to approach credit risk assessment as an optimisation problem: minimising complexity while maximising accuracy.

A recent review by Dastile et al. (2020) comparing statistical and ML learning models for credit scoring showed that ensembles outperform single classifiers, confirming the results of previously mentioned works. The authors identify model explainability and the ability to deal with imbalanced datasets, as the main issues to deal with when modelling credit risk. Deep learning models also show promising results, although they have not been extensively explored for credit risk assessment. The authors identify the lack of interpretability as the main barrier for adopting deep learning for credit risk assessment.

Banco de España (Alonso and Carbo 2021) published a comparison of several well-known machine learning algorithms for credit default prediction, showing significant improvements over logit. The authors estimate that implementing XGBoost-mediated assessment could lead to savings of up to 17% of capital requirements under current ECB regulation. Antunes (2021) from the Central Bank of Brazil presents a solid argument to maintain supervisory on-site inspections. The author compares two machine learning models, one trained with portfolio ratings assessed by the banks themselves, and the other based on past ratings obtained through on-site inspections. The results show that the overall performance is consistently higher when using data retrieved through inspections.

This is the period with the most ML applications papers identified (with a total of 9 out of 13). They span from insights on how AI will continue to revolutionise industries and change social behaviour (Dwivedi et al. 2021), to more practical approaches on how to incorporate ML in financial services (Lee and Shin 2020). Milian et al. (2019) also provide a list comparing fin-tech definitions, how it is supported by digital transformation, and the financial risks associated with the use of ML.

A comprehensive study from 2019 by di Castri et al. (2019) focuses on the definition of sup-tech and highlights the need for a more precise notion of what to include as "innovative technology" at the service of a financial authority. It presents several use cases and classifies the technologies onto maturity levels (named in the paper as "generations"), concluding that the identified initiatives (applications of innovative technologies to support the activities carried out by financial regulators and authorities) are mostly experimental. The authors suggest an international coordination effort and alignment to create synergies that leverage sup-tech development.

The Bank of Italy presented a use case for a classification problem (deducing the institutional sector code of a company based on its characteristics) (Massaro et al. 2020). Although this work is not related to risk assessment, it provides an excellent example of a production-ready application of ML to supervisory tasks.

Alonso and Carbo (2020) from Banco de España stress the need for a joint strategy to assess ML models to increase transparency and promote adherence to this technology. The authors conclude ML models increase the predictive capability of a credit default classifier by 20%. The study also identifies factors in credit risk management that might increase supervisory costs.

Driven by the recent progress in financial technology, Huang et al. (2021) acknowledge the complex and hierarchical nature of financial data and the technological barriers found when using statistics and classic ML. The authors then proceed to apply advanced deep learning methods and make use of several graphic processors to improve computation.

As a final remark regarding ML applications, Doerr et al. (2021), from the Bank of International Settlements, presented a policy briefing on the European Money and Finance Forum, evaluating to what extent central banks are making use of ML and big

data. The authors conclude that although central banks are acquainted with big data, there exists a persistent need for specialised knowledge on how to use ML throughout these organisations.

Stress tests are also referenced in these years. In a 2019 study, Kolari et al. (2019) hypothesise that stress tests themselves are more of an assessment of a bank's ability to deal with the risks it is exposed to. This statement challenges the common conception of stress tests as a marker of a bank's resilience to adverse alternative macroeconomic scenarios. For this purpose, the authors develop an early warning system to assess how European banks will perform on stress tests. These authors suggest surviving stress tests depends largely on the underlying risk dimensions of individual banks. Moreover, this paper reaffirms boosting techniques as winning solutions, not only for this sort of classification problems but also when applied to structured data. As a future work, the authors recommend a similar approach using regulatory data.

In the same line of investigation, an EWS was developed by Filippopoulou et al. (2020) to predict bank systemic risks in the Eurozone. This study starts by analysing the importance of the indicators that are usually applied and presents a model that detects a systemic crisis one to four years beforehand. In spite of using a classic multivariate binary logistic regression model, the methodology adopted for this EWS shows promising results and can be a reference for future developments in this area.

3.3. Datasets

Most central banks and supervisory agencies do not make their datasets available for confidentiality reasons. This is true for several types of data, such as credit responsibilities and supervisory data (European Banking Authority 2013).

As depicted in Figure 4, regardless of the research topic, most datasets used in these papers are public. The main reason for this is that most researchers cannot gain access to validated supervisory data. Another relevant aspect is that central banks and supervisory agencies have just begun to engage in programs where ML development strategies were in place. These developments are starting to appear, as can be seen by the growing number of titles under the "ML applications" topic. Table 4 lists the datasets used in each paper.
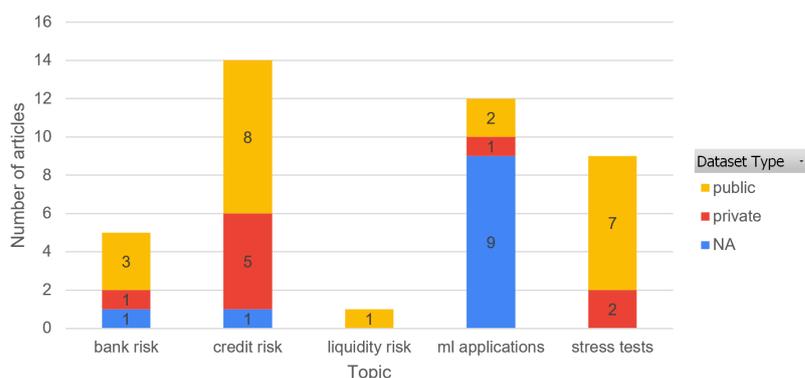
Figure 4. Dataset types by research topic (*NA*—not applicable).

Some rating agencies, central banks and other institutions provide datasets to support research projects. A good example is *Banco de Portugal* BPLIM (de Portugal 2021), a micro-data research laboratory that provides up-to-date anonymised datasets available for national and international researchers. Another example is Moody's DataHub (Moody's 2021), that provides a cloud-based platform containing eligible data alongside affiliated third-party participants.

3.4. Related Work

In this research, we have found few papers strictly addressing the use of machine learning techniques for supervisory risk assessment. As a consequence, we have broadened

our research question to include banking risk assessment and machine learning in the financial sector. This reasoning is thoroughly presented in Section 2.2.

Nonetheless, we found some works that support the purpose of this review. di Castri et al. (2019) is a survey that summarises the activities that can be considered as an application of innovative technology to supervisory purposes. The authors also present a series of use cases, mostly experimental and originated by supervisory agencies. Kou et al. (2019) list the most common methodologies—ML, big data analysis and sentiment analysis—to address systemic risk in the banking sector. Last, and closest to this research, Leo et al. (2019) contribute with a literature review that brings to light how machine learning is currently being used in the banking sector. The authors stress that contrarily to what might be expected due to the magnitude of financial consequences involved, the real-life use of these sophisticated technologies is in fact under-used and poorly developed.

The authors' specific knowledge of banking context, namely projects within *Banco de Portugal* and European Central Bank, allowed them to propose a reliable proxy for the scarcity of published works on this topic. To establish the ideal perspective, we evaluated how risk assessment is carried out in the banking industry, and central banks in the SSM. On the other hand, we investigated how ML is being used for risk assessment in banks. Additionally, we referenced various surveys from central banks to depict and support our statements regarding the use of innovative technologies for supervisory purposes.

In this sense, although this review is sustained by a proxy and there is a paucity of related works from a central bank perspective, the authors propose this review as a starting point for researchers and industry stakeholders. We aggregate relevant contributions to support and ignite the use of ML in risk assessment exercises, from a central bank or supervisory agency perspective.

3.5. Global Analysis

The set of papers identified in this review includes diverse approaches to risk assessment. We have selected some works that use a specific bankruptcy indicator (such as the Altman score or the O-score). However, most of the authors set forth from a set of financial ratios and, knowing the final result, try to model that knowledge through supervised learning. Most of these approaches convert the problem at hand to a classification task, for example, "failure" or "no failure" of a bank.

Another interesting aspect is how the datasets are designed. Most of these works use public datasets to validate a certain approach, even though some of these datasets are specifically collected to depict financial crises. The set of features available in these datasets often reflect a certain industry perspective of risk assessment. For instance, many datasets focus on credit and profitability ratios, since both are two crucial vectors for the industry: how a bank performs and how it is exposed to its main business model.

As a final remark, although most of the selected works come from the academia, we would like to mention the five papers published from 2017 until now by central banks. Alessi and Detken (2018), from the European Central Bank (ECB) and European Commission, have a significant number of citations (135 by the end of 2020) and present an important EWS that can support everyday processes. Also, Chakraborty and Joseph (2017) from the Bank of England give a great contribution with a broad view of what is being done with ML in this context. By presenting some use cases, they also turn the spotlight on the successes of these approaches. The Bank of Greece presents an insightful use case by Petropoulos et al. (2018) for credit risk analysis.

From a more strategic point of view, Jagtiani et al. (2018) from Federal Reserve Banks depict the impacts, roles and possible risks of using ML at central banks.

Although not related to risk assessment, a recent study Massaro et al. (2020) from Bank of Italy presents a production-ready solution of the application of ML techniques to everyday central bank tasks. This is one of the most recently works, showing how ML can make a difference in day to day tasks.

4. Conclusions

This review provides a comprehensive picture of how machine learning techniques have been used so far in risk assessment from a central bank's perspective. It is organised by timeline and topic. All of the presented topics relate to some extent to the supervisory activity and to dimensions of analysis that are part of the day-to-day processes. As a consequence of the SSM legislation and the EBA reporting requirements, this work focused on the European banking sector.

The majority of the selected papers reflect upon the credit scoring problem. This stems largely from the fact that granting loans is the core business of most of the commercial banking sector. Stress testing in the form of bankruptcy prediction is also in the spotlight since it is strongly connected with regulators' compliance. There are several other risks a bank is exposed to that require their own studies, such as liquidity or operational risk. However, focusing on those risks is more of a compliance issue, rather than a business model perspective.

Some studies benefited from more structure and clarity, which is useful for comparability purposes. The more structured studies answer the questions of which problem they are addressing (a measure of risk and its perspective, a stock index, portfolio pricing, etc.), ML techniques that were applied, and variables considered. They also offer insight into the datasets they were based upon, and clarify the methods used to assess the models' precision and prediction capability. The lack of this organised approach evidenced in some articles made it more difficult to review and condense the information published across the broad spectrum of expertise found. As a consequence, interpreting data originating in different geographies and diverse banks' business models proves to be a challenging task. International consensus must be established regarding terminology, analysis methods and result reporting, as pertaining to this field. The authors advocate for a universal risk assessment methodology, classifying bank risk according to preset parameters and based on the same data, regardless of their location or business model. To this end and taking advantage of the central bank's perspective, the authors suggest the use of the Supervisory Review and Evaluation Process (SREP), namely, one of its pillars, the Risk Assessment System (RAS). This methodology is used by the ECB and applied, to some extent, to every institution in the SSM. Through the application of such a broad methodology, results of analysis and ML application are more comparable to an already established practice.

Another relevant aspect is the paucity of data published from a supervisory perspective. The reviewed papers mainly focus on credit risk and stress tests using public data. Despite being useful in assessing the financial health of a credit institution, they seldom use data collected through supervisory directives. Scenario testing, sometimes used as a synonym for stress testing, is another decision support system that greatly increases the analytical capabilities of supervisors. The authors emphasise the importance of landmark publications such as the EWS proposed by Filippopoulou et al. (2020), using data gathered in the aftermath of the 2008 economic collapse (European Central Bank Macroprudential Database). These systems are especially relevant since they function as a daily tool for analysts, and strongly benefit from supervisory data. The EWS developed by Alessi and Detken (2018) has also had an enormous impact in the literature by presenting a solution for anticipating banking crisis, using random forests.

As a final remark, we point out that many of these studies rely on public datasets. This often implies they are not as recent as desired since the data might not include the more recent events. For instance, a dataset from 2005 to 2011 captures the market behaviour before the crisis, the crisis itself, and a fraction of the decline of the market. It would be useful to model the behaviour of the institutions with the new regulation as well as the economic recovery seen later until 2019.

Limitations and Future Work

This study proposed to select and review the literature regarding the applications of machine learning to banking supervision. However, since this is a rather specific topic and

the regulation has suffered a thorough revision after the 2008 financial crisis, our review falls short on papers that address solely this issue. There is some literature published by central banks and other agencies, but these works are mostly surveys, assessments of adoption, or definition of new concepts. As a consequence, the research query was broadened to include works from other perspectives:

- Assessment of credit defaults (the topic most explored in the reviewed literature);
- New stress test methodologies;
- Systemic risk detection;
- Other surveys regarding fin-tech and sup-tech.

All these topics are pillars of financial analysis and as such, they relate in a direct and crucial manner to proper supervision. Nevertheless, they are all collateral aspects and do not correspond to the core of the supervisory process itself.

Another aspect worth mentioning is the fact that our work is not a detailed review of the literature cited within it. Due to the heterogeneous structure of the included literature, we opted for a broader approach when comparing them. Each topic would merit an individual in-depth analysis and review, which was not warranted in the scope of this article. The authors believe this review will provide a stepping stone for supervisors, analysts, consultants, or academics that desire to further explore machine learning as a tool for banking risk assessment.

References

Abellán, Joaquín, and Javier G. Castellano. 2017. A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Systems with Applications* 73: 1–10. [CrossRef]

Ala'raj, Maher, and Maysam F. Abbod. 2016. A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Systems with Applications* 64: 36–55. [CrossRef]

Alessi, Lucia, and Carsten Detken. 2018. Identifying excessive credit growth and leverage. *Journal of Financial Stability* 35: 215–25. [CrossRef]

Alonso, Andrés, and Jose Manuel Carbo. 2020. Machine Learning in Credit Risk: Measuring the Dilemma Between Prediction and Supervisory Cost. *SSRN Electronic Journal*. [CrossRef]

Alonso, Andrés, and Jose Manuel Carbo. 2021. Understanding the Performance of Machine Learning Models to Predict Credit Default: A Novel Approach for Supervisory Evaluation. *SSRN Electronic Journal*. [CrossRef]

Altman, Edward. 1968. Financial Ratios, Discriminant Analysis and The Prediction of Corpporate Bankruptcy. *The Journal of Finance* XXIII: 589–609.

Angelini, Eliana, Giacomo di Tollo, and Andrea Roli. 2008. A neural network approach for credit risk evaluation. *Quarterly Review of Economics and Finance* 48: 733–55. [CrossRef]

Antunes, José Américo Pereira. 2021. To supervise or to self-supervise: A machine learning based comparison on credit supervision. *Financial Innovation* 7. [CrossRef]

Boyacioglu, Melek Acar, Yakup Kara, and Ömer Kaan Baykan. 2009. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications* 36 Pt 2: 3355–66. [CrossRef]

Broeders, Dirk, and Jeremy Prenio. 2018. FSI Insights Innovative technology in financial supervision. *FSI Insights on Policy Implementation* 2018: 29.

Burstein, Frada, Clyde W. Holsapple, and Daniel J. Power. 2008. *Decision Support Systems: A Historical Overview*. Berlin/Heidelberg: Springer. [CrossRef]

Chakraborty, Chiranjit, and Andreas Joseph. 2017. Machine Learning at Central Banks. *SSRN Electronic Journal*. [CrossRef]

Chang, Yung Chia, Kuei Hu Chang, and Guan Jhih Wu. 2018. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing Journal* 73: 914–20. [CrossRef]

Chaudhuri, Arindam, and Kajal De. 2011. Fuzzy Support Vector Machine for bankruptcy prediction. *Applied Soft Computing Journal* 11: 2472–86. [CrossRef]

Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. Paper presented at ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17; pp. 785–94. [CrossRef]

Climent, Francisco, Alexandre Momparler, and Pedro Carmona. 2019. Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research* 101: 885–96. [CrossRef]

Dastile, Xolani, Turgay Celik, and Moshe Potsane. 2020. Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing Journal* 91: 106263. [CrossRef]

de Portugal, Banco. 2021. Banco de Portugal Microdata Research Laboratory. Available online: https://bplim.bportugal.pt (accessed on 4 May 2021).

di Castri, Simone, Stefan Hohl, and Arend Kulenkampff. 2019. FSI Insights on policy implementation No. 19: The suptech generations. *Financial Stability Institute* 19: 19.

Doerr, By Sebastian, Leonardo Gambacorta, and Jose Maria Serena. 2021. How do central banks use big data and machine learning? *The European Money and Finance Forum* 67: 1–6. Available online: https://www.bis.org/publ/work930.pdf (accessed on 20 May 2021)

Dwivedi, Yogesh K., Laurie Hughes, Elvira Ismagilova, Gert Aarts, Crispin Coombs, Tom Crick, Yanqing Duan, Rohita Dwivedi, John Edwards, Aled Eirug, and et al. 2021. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 57. [CrossRef]

European Banking Authority. 2013. EBA Implementing Technical Standards (ITS). Available online: http://www.eba.europa.eu/documents/10180/532570/EBA-ITS-2013-12+(Final+draft+ITS+on+Hypothetical+Capital+of+a+CCP).pdf (accessed on 20 May 2021).

European Commission. 2015. *Single Supervisory Mechanism*. Brussels: European Commission.

Filippopoulou, Chryssanthi, Emilios Galariotis, and Spyros Spyrou. 2020. An early warning system for predicting systemic banking crises in the Eurozone: A logit regression approach. *Journal of Economic Behavior and Organization* 172: 344–63. [CrossRef]

Galindo, Jua , and Pablo Tamayo. 2000. Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. *Computational Economics* 15: 107–43. [CrossRef]

Gogas, Periklis, Theophilos Papadimitriou, and Anna Agrapetidou. 2018. Forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting* 34: 440–55. [CrossRef]

Gusenbauer, Michael. 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics* 118: 177–214. [CrossRef]

Hammer, Peter L., Alexander Kogan, and Miguel A. Lejeune. 2012. A logical analysis of banks' financial strength ratings. *Expert Systems with Applications* 39: 7808–21. [CrossRef]

Hillegeist, Stephen A., Elizabeth K. Keating, Donald P. Cram, and Kyle G. Lundstedt. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9: 5–34. [CrossRef]

Huang, Shian Chang, Cheng Feng Wu, Chei Chang Chiou, and Meng Chen Lin. 2021. Intelligent FinTech Data Mining by Advanced Deep Learning Approaches. *Computational Economics*. [CrossRef]

Jagtiani, Julapa, Larry Wall, and Todd Vermilyea. 2018. The Roles of Big Data and Machine Learning in Bank Supervision. *Banking Perspectives, Forthcoming* 1–11.

Kolari, James W., Félix J. López-Iturriaga, and Ivan Pastor Sanz. 2019. Predicting European bank stress tests: Survival of the fittest. *Global Finance Journal* 39: 44–57. [CrossRef]

Kou, Gang, Xiangrui Chao, Yi Peng, Fawaz E. Alsaadi, and Enrique Herrera-Viedma. 2019. Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy* 25: 716–42. [CrossRef]

Kupiec, Paul H. 2018. On the accuracy of alternative approaches for calibrating bank stress test models. *Journal of Financial Stability* 38: 132–46. [CrossRef]

Le, Hong Hanh, and Jean Laurent Viviani. 2018. Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. *Research in International Business and Finance* 44: 16–25. [CrossRef]

Lee, In, and Yong Jae Shin. 2020. Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons* 63: 157–70. [CrossRef]

Leo, Martin, Suneel Sharma, and Koilakuntla Maddulety. 2019. Machine learning in banking risk management: A literature review. *Risks* 7. [CrossRef]

López Iturriaga, Félix J., and Iván Pastor Sanz. 2015. Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications* 42: 2857–69. [CrossRef]

Massaro, Paolo, Ilaria Vannini, and Oliver Giudice. 2020. Institutional Sector Cassifier, a Machine Learning Approach. *SSRN Electronic Journal* 548. [CrossRef]

Milian, Eduardo Z., Mauro de M. Spinola, and Marly M. de Carvalho. 2019. Fintechs: A literature review and research agenda. *Electronic Commerce Research and Applications* 34. [CrossRef]

Min, Jae H., and Young Chan Lee. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications* 28: 603–14. [CrossRef]

Moody's. 2021. Moody's DataHub. Available online: https://datahub.moodys.io/ (accessed on 21 June 2021).

Ohlson, James A. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* 18: 109. [CrossRef]

Petropoulos, Anastasios, Vasilis Siakoulis, Evaggelos Stavroulakis, and Aristotelis Klamargias. 2018. A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting. *The Use of Big Data Analytics and Artificial Intelligence in Central Banking* 50: 30–31.

Pompella, Maurizio, and Antonio Dicanio. 2017. Ratings based Inference and Credit Risk: Detecting likely-to-fail Banks with the PC-Mahalanobis Method. *Economic Modelling* 67: 34–44. [CrossRef]

Ribeiro, Bernardete, Catarina Silva, Ning Chen, Armando Vieira, and João Carvalho Das Neves. 2012. Enhanced default risk models with SVM+. *Expert Systems with Applications* 39: 10140–52. [CrossRef]

Soui, Makram, Ines Gasmi, Salima Smiti, and Khaled Ghédira. 2019. Rule-based credit risk assessment model using multi-objective evolutionary algorithms. *Expert Systems with Applications* 126: 144–57. [CrossRef]

Strydom, Moses, and Sheryl Buckley. 2019. *AI and Big Data's Potential for Disruptive Innovation*, 1st ed. Engineering Science Reference. Hershey: IGI Global. [CrossRef]

Tavana, Madjid, Amir Reza Abtahi, Debora Di Caprio, and Maryam Poortarigh. 2018. An Artificial Neural Network and Bayesian Network model for liquidity risk assessment in banking. *Neurocomputing* 275: 2525–54. [CrossRef]

Wang, Tongyu, Shangmei Zhao, Guangxiang Zhu, and Haitao Zheng. 2021. A machine learning-based early warning system for systemic banking crises. *Applied Economics* 53 : 1–19. [CrossRef]

Xia, Yufei, Chuanzhe Liu, Yu Ying Li, and Nana Liu. 2017. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications* 78: 225–41. [CrossRef]

Zhang, Yicheng, Jipeng Gao, and Haolin Zhou. 2020. ImageNet Classification with Deep Convolutional Neural Networks. In *ACM International Conference Proceeding Series*. New York: Association for Computing Machiner; vol. 2, pp. 145–151. [CrossRef]