Research Article / Industrial engineering http://dx.doi.org/10.15446/ing.investig.v40n2.78649

Ten-year evolution on credit risk research: a Systematic Literature Review approach and discussion

Diez años de evolución en la investigación de riesgo de crédito: un enfoque y discusión de revisión sistemática de literatura

Fernanda Medeiros Assef[1] and Maria Teresinha Arns Steiner[2]

ABSTRACT

Given its importance in financial risk management, credit risk analysis, since its introduction in 1950, has been a major influence both in academic research and in practical situations. In this work, a systematic literature review is proposed which considers both "Credit Risk" and "Credit risk" as search parameters to answer two main research questions: are machine learning techniques being effectively applied in research about credit risk evaluation? Furthermore, which of these quantitative techniques have been mostly applied over the last ten years of research? Different steps were followed to select the papers for the analysis, as well as the exclusion criteria, in order to verify only papers with Machine Learning approaches. Among the results, it was found that machine learning is being extensively applied in Credit Risk Assessment, where applications of Artificial Intelligence (AI) were mostly found, more specifically Artificial Neural Networks (ANN). After the explanation of each answer, a discussion of the results is presented.

Keywords: credit risk assessment, machine learning, systematic literature review.

RESUMEN

Dada su importancia en la gestión del riesgo financiero, el análisis del riesgo crediticio, desde su introducción en 1950, ha tenido una gran influencia tanto en investigaciones académicas como en situaciones prácticas. En este trabajo se propone una revisión bibliográfica sistemática que considere "Credit Risk" y "Credit risk" como parámetros de búsqueda para responder dos preguntas de investigación principales: ¿se están aplicando efectivamente las técnicas de aprendizaje automático en las investigaciones sobre la evaluación del riesgo de crédito? Incluso, ¿cuáles de estas técnicas cuantitativas se han aplicado mayoritariamente en los últimos diez años de investigación? Se siguieron diferentes pasos para seleccionar los artículos para el análisis, así como los criterios de exclusión para verificar solo los artículos con enfoques de aprendizaje automático. Entre los resultados, se encontró que el aprendizaje automático se está aplicando ampliamente en la Evaluación de Riesgo de Crédito, donde en su mayoría se encontraron aplicaciones de Inteligencia Artificial (AI), más específicamente, de Redes Neuronales Artificiales (ANN). Después de la explicación de cada respuesta, se presenta una discusión sobre los resultados.

Palabras clave: evaluación de riesgo de crédito, aprendizaje automático, revisión sistemática de literatura.

Received: March 24th, 2019
Accepted: April 23th, 2020

Introduction

Credit risk analysis is an active research area in financial risk management, and credit scoring is one of the key analytical techniques in credit risk evaluation (Yu, Wang, and Lai, 2009; Steiner, Nievola, Soma, Shimizu, and Steiner Neto, 2007). With the fast development of financial products and services, bank credit departments have collected large amounts of data, which risk analysts use to build appropriate credit risk models to accurately evaluate an applicant's credit risk (Zhang, Gao, and Shi, 2014).

Credit risk evaluation is a data mining research problem, both challenging and important in the field of financial analysis. This assessment is used in predicting whether or not there is a possibility for credit concession. Since its introduction in 1950, it has been extensively applied and, more recently, it has been performed in lending concessions, credit card analysis, and its natural application, credit concession (Luo, Kong, and Nie, 2016).

According to the work of Zhang, Gao, and Shi (2014), there is a wide range of methodologies for solving credit risk classification problems. These methods include mainly logistic regression, probit regression, nearest neighbor analysis, Bayesian networks, Artificial Neural Networks (ANN), decision trees, genetic algorithms (GA), multiple criteria decision making (MCDM), support vector machines (SVM),

[1]Industrial Engineer, Federal University of Paraná (UFPR), Curitiba, Brasil. M.Sc. Industrial Engineering, Federal University of Paraná, Brazil. Affiliation: Doctoral Student at Pontifical Catholic University of Paraná (PUCPR). Brazil.
E-mail: fermassef@gmail.com
[2]Mathematics and Civil Engineer, UFPR, Brasil. M.Sc. and D.Eng. Industrial Engineering, UFSC, Brasil. Pos-Doc Industrial Engineering, ITA, Brasiland IST de Lisboa, Portugal. Affiliation: Titular Professor, Pontifícia Universidade Católica do Paraná (PUC-PR), Brasil. E-mail: maria.steiner@pucpr.br

How to cite: Assef, F. M. and Steiner, M. T. A. (2020). Ten-year evolution on credit risk research: a systematic literature review approach and discussion. *Ingeniería e Investigación*, *40*(2), 50-71. 10.15446/ing.investig.v40n2.78649

among many others. ANN credit assessment models are highly accurate, but some modeling skills are needed, for example, to design appropriate network topologies. On the other hand, models based on SVM have indicated promising results in credit risk assessment, but they need to solve a convex quadratic programming problem which, computationally, is very expensive in real-world applications.

Looking at the potential benefits that can be achieved through the deployment of research surrounding credit risk, as well as its different methodologies, some questions arise:

Q1. Are machine learning techniques being effectively applied in research about credit risk evaluation?

Q2. Which of these quantitative techniques have been mostly applied over the last ten years of research?

With the objective of seeking answers for these two questions through an extensive search in the available literature, a systematic literature review (SLR) is proposed, as well as a discussion about the obtained results in an attempt to understand the current research landscape and how future works may be steered. The use of SLR does not only contribute to more robust research findings but also enables reproduction and updates of a given review by members of the scientific community. The importance of this work lies in clarifying the current role being played by quantitative methods in Credit Risk Evaluations.

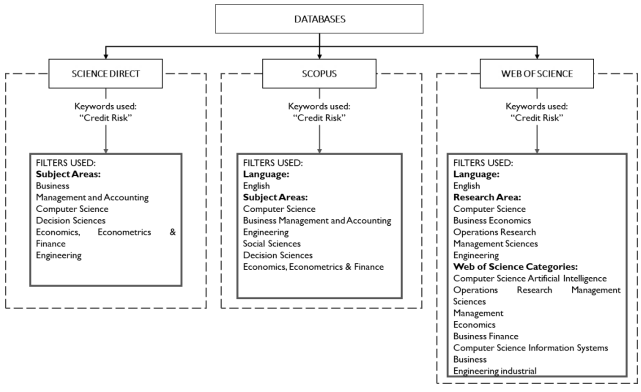

Figure 1. Search filters for each paper database.
Source: Authors

Review Protocol

Having defined the questions, we chose the Web of Science (WoS), Science Direct, and also Scopus databases due not only to their both dynamic and simple interface, but also due to the possibility of obtaining different kinds of analysis from the search.

Systematic literature review method

The search parameters for this research were "Credit risk" and "Credit Risk," both used for this type of research. These keywords were used in the three above-mentioned paper databases.

Initially, 285 documents were found in the Scopus database, 227 in WoS, and 502 in Science Direct. For each database, a few other filters were applied to best select the cut from the total of papers on which we desired to develop our research, these filters and the databases on which they were applied can be found in Figure 1.

At the end of this step, the documents were exported in order to assess their information both in a bibliometric way, as well as through a content analysis, aiming to answer the previous research questions.

Besides the fact that we chose to use three different databases for our paper selection, the originality of our research lies in the types of assessment the authors present in the sections below. We chose to differentiate our bibliometric analysis by presenting the assessment of journals, the number of citations, and a Pareto analysis of each paper's citation. As for the content analysis portion, we present a summary in the form of a table for each paper, as well as a brief analysis.

Credit Risk Assessment Research: the past ten years of research

According to the exclusion criteria shown in Figure 1, 374 documents from the initial amount were approved for the next step of our review: both the bibliometric review and the qualitative content analysis.

As said before, a few premises were considered before starting the content analysis. Since the number of papers found might be too granular, and some papers were not as influential in research as others, we filtered the papers according to their citations (from most to least cited paper). After this step, we considered the proportion that each article had in comparison with the sum of citations from every single one of the collected papers. An example of how this procedure was made is shown below in Table 1.

Table 1. Citation percentage for each paper from the Web of Science

#	Authors	Cited By	%	% Acc
1	Louzis et al. (2011)	987	7,1%	7,1%
2	Cornett et al. (2011)	930	6,7%	13,7%
3	Puri et al. (2011)	591	4,2%	17,9%
4	Firth et al. (2009)	403	2,9%	20,8%
5	Yeh and Lien (2009)	390	2,8%	23,6%
6	Lessmann et al. (2015)	387	2,8%	26,4%
7	Wang et al. (2011)	335	2,4%	28,8%
8	Lee et al. (2015)	303	2,2%	31,0%
9	Khandani et al. (2010)	288	2,1%	33,0%
10	Bellotti and Crook (2009).	287	2,1%	35,1%
...	...	...	...	...

Source: Authors

After doing so, a Pareto analysis was performed in order to find how many papers were responsible for at least 80% of the overall citation found in the search. We chose this amount according to Pareto's Principle, or the 80/20 rule; we brought this management principle to our bibliometric analysis. By

observing the citation amount for each database, we were able to find that 27 papers happened to be responsible for 80% of the sum of citations, which represented 20% of the total of papers in the WoS database –thus confirming the possibility of using the above-mentioned rule.

The same procedure was applied for Scopus (38 papers were selected) and Science Direct (112 papers). Adding WoS 27 papers, Scopus' 38 and Science Direct 112, the 177 selected documents were put together, and the duplicated ones were excluded in order to present a clean-cut from all selected papers. After that, the next step for the proposed review was to apply several exclusion criteria. First, the papers which were not found were excluded; secondly, papers from conferences; after that, the ones without credit risk applications; then, papers before 2009 (they were excluded from the content analysis); and finally, the ones that had theoretical explanations (i.e., papers that did not apply data mining techniques, surveys, state of the art reviews, and theoretical frameworks).

The journals considered for this analysis can be found in Table 2, and their h-index was collected to illustrate their impact. Analyzing this table, we are able to observe that 12 out of 31 journals have an h-index over 100, and the average of the presented journals was around 90.

Table 2. The journals considered for the analyses

Journal	Count	H-Index
European Journal of Operational Research	12	211
Journal of Financial Economics	3	206
Research Policy	1	191
Information Sciences	2	154
Expert Systems with Applications	37	145
Journal of Business Research	1	144
Journal of Banking and Finance	9	126
Computers and Operations Research	1	124
Decision Support Systems	2	115
IEEE Transactions on Systems, Man and Cybernetics: Systems	1	111
Journal of Monetary Economics	1	107
Neurocomputing	1	100
Applied Soft Computing	6	97
Journal of the Operational Research Society	2	87
Computational Statistics and Data Analysis	1	85
Automation in Construction	1	83
Knowledge-Based Systems	3	82
Management Decision	1	77
International Journal of Forecasting	3	74
Journal of Financial Intermediation	3	63
Journal of Empirical Finance	1	63
International Journal of Neural Systems	1	50
Journal of Intelligent Information Systems	1	47
Journal of Applied Statistics	1	45
International Review of Financial Analysis	1	38
Review of Quantitative Finance and Accounting	1	36
Procedia Computer Science	1	34
International Journal of Finance and Economics	1	33
Journal of Financial Stability	1	32
Review of Development Finance	1	9
Intelligent Systems in Accounting, Finance, and Management	1	5

Continued on next column

Source: Authors

Another analysis we were able to obtain concerns the amount of citations each journal received considering the papers selected, as seen in Table 3.

Table 3. Journals and citations amounts

Journal	H-Index	Citation
Expert Systems with Applications	162	4 653
Journal of Banking & Finance	135	2 442
European Journal of Operational Research	226	1 833
Journal of Financial Economics	223	1 646
Applied Soft Computing	110	509
Journal of Financial Intermediation	67	447
International Journal of Forecasting	79	406
Information Sciences	154	389
Decision Support Systems	127	347
Research Policy	206	303
International Review of Financial Analysis	43	255
Journal of Financial Stability	38	237
Knowledge-Based Systems	94	234
Journal of Empirical Finance	66	168
International Journal of Finance & Economics	36	101
Computers & Operations Research	133	95

Continued on next column

Journal	H-Index	Citation
Review of Development Finance	13	95
Journal of the Operational Research Society	94	95
Neurocomputing	110	93
Journal of Monetary Economics	112	85
International Journal of Neural Systems	55	75
Journal of Business Research	158	71
Computational Statistics & Data Analysis	93	63
Procedia Computer Science	47	57
Journal of Intelligent Information Systems	49	55
Automation in Construction	95	49
Journal of Applied Statistics	48	45
Review of Quantitative Finance and Accounting	36	39
IEEE Transactions on Systems, Man and Cybernetics: Systems	111	37
Management Decision	82	32
Intelligent Systems in Accounting, Finance, and Management	5	9

Source: Authors

From Table 3, we were able to assume that journals with a higher h-index were not always cited more than the others. The seven first journals presented in this table represent 80% of the overall citation, being *Expert Systems with Applications, Journal of Banking & Finance, European Journal of Operational Research, Journal of Financial Economics, Applied Soft Computing, Journal of Financial Intermediation,* and the *International Journal of Forecasting*. After analyzing the information surrounding each paper, their content was reviewed, and the Table 4 below was built in order to summarize their information in chronological order. The best performance technique (where possible) is indicated in boldface.

Table 4. Summary of analyzed papers

#	Authors	Cited By	Techniques
1	Antonakis and Sfakianakis, 2009	33	NBR; LDA; LR; kNN; **CT**; **ANN**
3	Bose and Chen, 2009	109	Not Applied
4	Chen, Ma, and Ma, 2009	104	CART; **MARS**; **SVM**
5	Firth et al., 2009	326	Descriptive Statistics
6	Jiménez, Salas and Saurina, 2009	108	Descriptive Statistics
7	Khashman, 2009	63	**SHNN**; DHNN
8	Koopman et al., 2009	139	Proposed technique
9	Lessmann and Vob, 2009	63	RBF SVM; SVM; **LR**; CART
10	Lin, 2009	82	LR; **LLR**; **ANN**
11	Marinakis et al., 2009	83	**ACO**; **PSO**; TS; AG
12	Tsai et al., 2009	52	DA; LR; ANN; **DEA-DA**
13	Xu, Zhou, and Wang, 2009	82	SVM; **HARA**; HubAvgRA; ATkRA
14	Yeh and Lien, 2009	197	KNN; LR; DA; NB; **ANN**; CT
15	Yu, Wang, and Lai, 2009	187	**Fuzzy-GDM**; GDM; SVMR; RBF NN; BPNN; LR;LinR
16	Brown and Zehnder, 2010	73	Descriptive Statistics
17	Cardone-Riportella, Samaniego-Medina, and Trujillo-Ponce, 2010	120	UVA; MVA
18	Dong, Lai, and Yen, 2010	33	LRF; **LRR**
19	Khandani, Kim and Lo, 2010	145	**CART**
20	Khashman, 2010	159	Three architectures of ANN with nine learning methods
21	Malik and Thomas, 2010	53	CPH
22	Marinaki, Marinakis, and Zopounidis, 2010	69	**HBMO**; ACO; PSO; GA; TS
23	Paleologo, Elisseeff, and Antonini, 2010	122	SVMLin; ANN; **DT**; SVM-RBF; AdaBoost
24	Psillaki, Tsolas, and Margaritis, 2010	147	DEA; **LR**
25	Tsai and Chen, 2010	98	DT; **ANN**; NBC; **LR**; K-Means; EM
26	Twala, 2010	122	LD; KNN; NBC; ANN; **DT** (combined with whether Feature Selection, Boosting or Both)
27	Zhou, Lai, and Yu, 2010	125	LDA; DLDA; QDA; DQDA; LR; PR; ID3; CART; BPNN; ProbNN; BC; KNN; AdaBoost; **LSSVMRBF**; **LSSVMLin**; 1nSVMRBF

Continued on next column

#	Authors	Cited By	Techniques
28	Andrés et al., 2011	104	MARS; **Fuzzy C-Means**; ANN; DA
29	Cerqueiro, Degryse, and Ongena, 2011	86	Descriptive Statistics
30	Cornett et al., 2011	698	Descriptive Statistics; Regression Models (Not Specified)
31	Finlay, 2011	106	LR; LDA; CART; ANN; KNN; **ET Boost**
32	Hájek, 2011	85	ANN; RBF NN; **ProbNN**; CCNN; GMDH; SVM; MDA; LR; K-Means; CT
33	Huysmans et al., 2011	152	DT; **Dtab**
34	Khashman, 2011	65	12 Different architectures of EmBP and ANN (six of each)
35	Li et al., 2011	80	MDA; Logit Regression; CBR + kNN; CBR + DT; SPNIC-CBR; **SVM**
36	Louzis, Vouldis, and Metaxas, 2011	618	Descriptive Statistics; **GMM**
37	Magri and Pico, 2011	89	Descriptive Statistics
38	Peng et al., 2011	110	MCDM (TOPSIS, PROMETHEE, VIKOR); **NB**; LR; KNN; SVM; RBF NN;
39	Puri, Rocholl, and Steffen, 2011	500	Descriptive Statistics; Bivariate Tests
40	Tseng et al., 2011	37	**ESVM**; LR
41	Wang et al., 2011	210	**LRA**; **DT**; ANN; SVM; Bagging
42	Yap, Ong, Husain, 2011	121	Credit Scorecard Model; **LR**; DT
43	Zambaldi et al., 2011	54	PR
44	Zhou *et al.* 2011	25	SVM; **KASNP**
45	Derelioğlu and Gurgen 2011	42	**RFE-SVM**; CRED; DT; **MLP**; k-NN
	Wang and Ma, 2012	109	**RSB-SVM**
46	Koyuncugil and Ozgulbas, 2012	133	**CHAID**
47	Kwak *et al.* 2012	39	**MCLP**
48	Akkoç, 2012	127	LDA; LRA; ANN; **ANFIS**
48	Bellotti and Crook, 2012	87	**OLS**; DT
50	Bijak and Thomas, 2012	41	**CART**; **CHAID**; LOTUS; LR
51	Capotorti and Barbanera, 2012	56	Hybrid proposed by the authors

Continued on next column

#	Authors	Cited By	Techniques
52	Chang and Yeh, 2012	43	**AINE**; AIRS; SAIS
53	Chi and Hsu, 2012	61	HGADSM
54	Crone and Finlay, 2012	90	LR; LDA; CART; ANN
55	Hens and Tiwari, 2012	51	SVM; SVM + GA; BPNN; **GP**
56	Loterman et al., 2012	94	OLS; B-OLSBR; BC-OLS; RiR; RoR; RT; MARS; **LSSVM**; **ANN**; LR+OLS; LR+B-OLS; LR+BR; LR+BC-LS; LR+RiR; LR+RoR; LR+RT; LR+MARS; LR+LSSVM; LR+ANN; OLS+RT; OLS+MARS; OLS+LSSVM; OLS+ANN; OLS + MARS; OLS + LSSVM; OLS + ANN
57	Marqués, García, and Sánchez, 2012	58	ANN; LR; SVM; AdaBoost; **Bagging**; RSM; **RF**
58	Marqués, García and Sánchez, 2012b	56	ANN; NBC; LR; RBF NN; **SVM**; C4.5
59	Menkhoff, Neuberger and Rungruxsirivorn, 2012	87	Descriptive Statistics
60	Oreski, Oreski, and Oreski, 2012	117	**GANN**; FSNN; NNGM
61	Sánchez-Lasheras et al., 2012	47	**SOM**; **MARS**; BPNN
62	Tong, Mues and Thomas, 2012	67	MCM; CPH; **LR**
63	Van et al., 2012	80	LR
64	Vukovic et al., 2012	49	**CBR**; KNN; **GA**; **PF**
65	Wang et al., 2012	108	LRA; LDA; MLP; RBF NN; **RS-Bagging DT**; Bagging-RS DT
66	Wang et al., 2012	48	RSFS; ANN; DT; LR
67	García *et al., 2012*	33	Data Filtering
68	Blanco et al., 2013	103	**MLP**; LDA; QDA; LR
68	Chen et al., 2013	59	SOM; FSOM; TSOM
70	Cotugno, Monferrà, and Sampagnaro, 2013	83	Descriptive Statistics; MRA
71	Harris, 2013	35	Seven Models of SVM
72	Kruppa et al., 2013	61	LR; KNN; BNN; **RF**
73	Miguéis, Benoit, and Van Den Poel, 2013	19	BBQR
74	Tinoco and Wilson, 2013	159	**LR**

Continued on next column

#	Authors	Cited By	Techniques
75	Zhu et al., 2013	26	TOPSIS; **C-TOPSIS**; LDA; QDA; DT; LR; kNN; SVM; LSSVM
76	Kou et al., 2014	389	**MCDM**
77	Bekhet and Eletter, 2014	47	**LR**; RBF NN
78	Bubb and Kaufman, 2014	71	Not Applied
79	Ferreira et al., 2014	26	MCDM; **MACBETH**
80	García, Marqués, and Sánchez, 2014	38	MLP; SVM; 1-NN; C4.5
81	Jankowitsch, Nagler, and Subrahmanyam, 2014	89	Descriptive Statistics
82	Oreski and Oreski, 2014	147	HGANN
83	Zhang, Gao, and Shi, 2014	36	MCOC; SVM; Fuzzy-SVM; **KFP-MCOC**
84	Zhong et al., 2014	76	ANN; ELM; I-ELM; **SVM**
85	Wu et al., 2014	37	**LR**; DT; ANN
86	Danenas and Garsva, 2015	46	**CSVM**; SVM; LSVM
87	Florez-Lopez and Ramon-Jeronimo, 2015	21	LDA; LR; kNN; **SVM**; ANN; CHAID; C4.5; CART; **Gradient Boosting**; RF; SVDF; WVDF; CADF
88	Ghosh, 2015	95	GMM
89	Harris, 2015	103	**CSVM**; SVM; LinR; LR; DT; ANN
90	Iturriaga and Sanz, 2015	75	DA; LR; RF; **ANN**; **SVM**; SOM
91	Laeven, Levine, and Michalopoulos, 2015	175	Proposed technique but not applied
92	Lee, Sameen, and Cowling, 2015	144	Descriptive Statistics
93	Lessmann et al., 2015	173	Not Applied
94	Zhao et al., 2015	55	MLP
95	Zhou et al., 2015	37	FVIF; WMBGA
96	Sousa et al., 2016	48	Framework
97	Guo et al., 2016	161	P2P
98	Cleofas-Sánchez et al., 2016	28	LR; ANN; SVM; **Hybrid**
99	Lahmiri, 2016	9	Feature Selection; SVM; ANN; Bayes NN
100	Beck et al., 2016	112	Not Applied
101	Abellán and Castellano, 2017	20	**LR**; **MLP**; SVM; **C4.5**; **CDT** - combined with AdaBoost, **Bagging**, Random Subspace, DECORATE, Rotation Forest

Continued on next column

Legend: 1nLSSVMRBF (1-Norm Support Vector Machines With Radial Basis Functions Kernel); 1-NN (1 Nearest Neighbor); ACO (Ant Colony Optimization); AINE (Artificial Imune Network); AIRS (Artificial Immune System); ANFIS (Adaptive Neuro Fuzzy System); ANN (Artificial Neural Networks); ATkRA (At K Ranking Applicants Algorithm); AV (Account Variables); BBQR (Bayesian Binary Quantile Regression); BC-OLS (Box–Cox Transformation Ordinary Least Squares Estimation); bNN (Bagged k-Nearest Neighbors); B-OLS (Beta Transformation Ordinary Least Squares Estimation); BPN (Back Propagation Network); BPNN (Back Propagation Neural Networks); BR (Beta Regression); C4.5 (C4.5 Decision Tree); CADF (Correlated-Adjusted Decision Forests); CART (Classification and Regression Trees); CBR (Case-Based Reasoning); CCNN (Cascade Correlations Neural Networks); CDT (Credal Decision Tree); CHAID (Chi-square automatic interaction detection); MCOC (Multi-Criteria Optimization Classifier); CPH (Cox Proportional Hazards); CRM (Cox Regression Model); CT (Classification Trees); C-TOPSIS (Classification Technique for Order Preference by Similarity to Ideal Solution); CRED (Continuous/Discrete Rule Extractor via Decision Tree Induction); DA (Discriminant Analysis); DA (Discriminant Analysis); DEA (Data Envelopment Analysis); DHNN (Double Hidden Layer Neural Networks); DLDA (Diagonal Linear Discriminant Analysis); DQDA (Diagonal Quadratic Discriminant Analysis); DT (Decision Trees); Dtab (Decision Table); ELM (Extreme Learning Machine); EM (Expectation Maximization); EmBP (Emotional Back Propagation); ESVM (Enforced Support Vector Machines Based Model); ET Boost (Error Trimmed Boosting); FSNN (Feature Selection Neural Networks); FSOM (Feature Self-Organizing Maps); FVIF (Filter Method and Variance Inflation Method); GA (Genetic Algorithm); GANN (Genetic Algorithm Neural Networks); GDM (Group Decision Making); GMDH (Group Method of Data Handling); GMM (Generalized Method of Moments); GP (Genetic Programming); HARA (Hub Authority Ranking Applicants); HBMO (Honey Bee Mating Optimization); HGADSM (Hibrid Genetic Algorithm into Dual Scoring Model); HGANN (Hybrid Genetic Algorithm Neural Network); HubAvgRA (Hub-Avb Ranking Applicants Algorithm); ID3 (Decision Trees with different Tree Construction Algorithms); I-ELM (Incremental Extreme Learning Machine); IOM (Instance-Based Model; KASNP (a kernel-based learning method called kernel affine subspace nearest point); KFP-MCOC (Kernel, Fuzzyfication and Penalty Factors Multi-Criteria Optimization Classifier); KNN (k-Nearest Neighbors); LD (Logistic Discrimination); LDA (Linear Discriminant Analysis); LinR (Linear Regression); LLR (Logarithm Logistic Regression); LR (Logistic Regression); LRA (Logistic Regression Analysis); LRF (Logistic Regression with Fixed Coefficients); LRR (Logistic Regression with Random Coefficients); LSSVMLin (Least Square Support Vector Machines with Linear Kernel); LSSVMRBF (Least Square Support Vector Machines with Radial Basis Functions Kernel); MACBETH (Measuring Attractiveness is applied by a Categorical Based Evaluation Technique); MARS (Multivariate Adaptive Regression Splines); MCDM (Multiple Criteria Decision Making); MCM (Mixture Cure Model); MLP (Multilayer Perceptron); MV (Macroeconomic Variables); MCLP (Multiple Criteria Linear Programming); MVA (Multivariate Analysis); NB (Naive Bayesian); NBC (Naive Bayes Classifier); NNGM (Generic Model for Parameters Optimization of the Artificial Neural Network); OLS (Ordinary Least Squares Estimation); P2P (Peer-to-Peer); PF (Preference Functions); PR (Probit Regression); ProbNN (Probabilistic Neural Networks); PROMETHEE (Preference Ranking Organization Method for Enrichment of Evaluations); PSO (Particle Swarm Optimization); QDA (Quadratic Discriminant Analysis); RBF NN (Radial Basis Functions Neural Networks); RBM (Basic Rating-Based Model); RBM+ (Refined Rating-Based Model); RiR (Ridge Regression); RoR (Robust Regression); RSB-SVM (Random Subspace Support Vector Machine); RSFS (Random Subset Feature Selection); RFE-SVM (recursive feature extraction with support vector machines); RSM (Random Subspace Method); RT (Regression Tree); SAIS (Simple Artificial Imune System); SHNN (Single Hidden Layer Neural Networks); SME (Small and Medium Enterprises); SOM (Self-Organizing Maps); SPINIC-CBR (Similarities to Positive and Negative Ideal Cases - Case-Based Reasoning); SVDF (Simple Majority Vote); SVM (Support Vector Machines); SVMLin (Support Vector Machines With Linear Kernel); SVMR (Support Vector Machines Regression); SVMRBF (Support Vector Machines With Radial Basis Functions Kernel); TOPSIS (Technique for Order Preference by Similarity to Ideal Solution); TS (Tabu Search); TSOM (Trajectory Self-Organizing Maps); UVA (Univariate Analysis); VIKOR (*VlseKriterijumska Optimizacija I Kompromisno Resenje* [Multi-criteria Optimization and Compromise Solution]); WMBGA (Wrapper Method Based on Genetic Algorithm); WVFD (Weighted Majority Vote).

Source: Authors

The first noticeable thing after analyzing the papers is that with all the filters applied, not many papers from 2017 until today were shown. In order to include these documents, the same research agenda was applied to the last two years

from 2017 until now. After selecting from the bases and filtering with the same 80% citation criteria and excluding theoretical and repeated papers, the remaining papers for analysis amounted to 15, as shown below in Table 5.

Table 5. Summary of analyzed papers (2017-2019)

#	Author	Cited By	Methods
1	Xia et al. 2017	149	TPE; RS; GS; MS; **XGBoost**; AdaBoost; ANN; DT; Bagging; DT; LR; RF; SVM; GBDT
2	Barboza, Kimura, and Altman, 2017	145	SVMLin; SVM RBF; **Boosting**; Bagging; **RF**; ANN; LR; MDA
3	Sun *et al*. 2018	75	**Bagging**; DTE-SBD; **DT**; SMOTE; DSR
4	Luo, Wu, and Wu D., 2017	60	SVM; **DBN**; MLR; MLP
5	Li *et al*. 2017	48	**SSVM**
6	Xia, Liu, and Liu, 2017	48	LR; RF; CSLR-T; CSRF-T; CSLR-SMOTE; CSRF-SMOTE; **CSXGBoost**
7	Bequé and Lessmann, 2017	45	ELM; ANN; KNN; **SVM-Lin**; SVM RBF; CART; J48; **LR-R**
8	Lanzarini *et al*. 2017	44	C4.5; **LVQ+PSO**
9	Xia *et al*. 2018	37	LR; GPC; SVM; DT; SVM Bagging; GPC Bagging; **RF**; **XGBoost**; **MV**
10	Kvamme *et al*. 2018	30	**CNN**
11	Tavana *et al*. 2018	30	**ANN**; BNN
12	Maldonado *et al*. 2017	25	**Logit Regression**; FS; RFE-SVM; HOSVM; SVM
13	Dirick et al., 2017	21	AFT; **CPH**
14	Moradi and Rafiei, 2019	12	ANFIS
15	Khemakhem and Boujelbene, 2018	9	SMOTE; ANN; DT

Legend: TPE (Tree-Structured Parzen Estimator); RS (Random Search); GS (Grid Search); MS (Manual Search); XGBoost (Extreme Gradient Boosting); GBDT (Gradient Boosting Decision Tree); ANN (Artificial Neural Networks); DT (Decision Trees); LR (Logistic Regression); RF (Random FoSynthetic Minority Over-Sampling Technique);; SVM (Support Vector Machines); SVMLin (Linear Support Vector Machines); SVM RBF (Radial Basis Functions Support Vector Machines); MDA (Multivariate Discriminant Analysis); SMOTE (Synthetic Minority Over-Sampling Technique); DSR (Differentiated Sampling Rates); DTE-SBD (Decision Tree Ensemble based on SMOTE, Bagging and DSR); DBN (Deep-Belief Network); MLR (Multinomial Logistic Regression); SSVM (Semi-Supervised Support Vector Machines); CSLR-T (Thresholding Logistic Regression); CSRF-T (Thresholding Random Forests); CSLR-SMOTE (Logistic Regression Balanced with Synthetic Minority Over-Sampling Technique); CSRF-SMOTE (Random Forests Balanced with Synthetic Minority Over-Sampling Technique); CSXGBoost (Cost-Sensitive Extension of Xgboost); ELM (Extreme Learning Machines); KNN (K-Nearest Neighbours); CART (Classification and Regression Trees); LR-R (Regularized Logistic Regression); C4.5 (C4.5 Decision Trees); LVQ PSO (Learning Vector Quantization Particle Swarm Optimization; GPC (Gaussian Process Classifier); MV (Majority Voting); BNN (Bayesian Neural Networks); FS (Fisher Score); RFE-SVM (Recursive Feature Elimination Support Vector Machines); HOSVM (Holdout Support Vector Machines); AFT (Accelerated Failure Time); CPH (Cox Proportional Hazards)**;** ANFIS (Adaptive Network-based Fuzzy Inference System); SMOTE (Synthetic Minority Oversampling Technique)

Source: Authors

As for the analysis of the journals from the past two years, it can be found in Table 6.

Table 6. Journals and citations amounts

Journal	H-Index	Citation
Expert Systems with Applications	162	468
Information Sciences	154	75
Engineering Applications of Artificial Intelligence	86	60
Electronic Commerce Research and Applications	62	48
Neurocomputing	110	30
Decision Support Systems	127	30
Journal of the Operational Research Society	94	25
Kybernetes	33	21

Source: Authors

Research without comparison between results

Among the analyzed papers, 30 documents did not compare the applied techniques nor the author-proposed ones, not being able to verify their performance. Thus, they will be the first papers to be assessed in this first part of the content analysis.

There were initially the papers which used only descriptive statistics as their means to evaluate credit risk, either to evaluate the broader effects of the US financial crisis on global lending to retail customers (Puri, Rocholl, and Steffen, 2011), or even to examine how the Chinese state-owned banks allocate loans to private firms (Firth, Lin, Liu, and Wong, 2009). Among the analyzed papers, authors were found whose main concern was to address the hardship that SMEs (Small and Medium Enterprises) may find in order to access financial aid or credit for investments (Lee, Sameen, and Cowling, 2015).

With more of a qualitative approach, Guo, Zhou, Luo, Liu, and Xiong (2016) used an instance-based model to assess a loan's credit risk by formulating P2P lending into portfolio optimization with boundary constraints. The authors then described the similarity between two loans by using default likelihood distance. Also, Sousa, Gama, and Brandão (2016) developed an approach to deal with changing environment in credit risk modeling by establishing a framework for this assessment. An application to a real-world financial dataset of credit cards from a financial institution in Brazil illustrates our methodology, which is able to consistently outperform the static modeling schema.

There were also authors who performed their research about the effects of organizational distance on the use of collateral for business loans by Spanish banks on the basis of the recent lender-based theory of collateral (Jiménez, Salas, and Saurina, 2009). Others considered the recovery rates of defaulted bonds in the US corporate bond market, based on a complete set of traded prices and volumes (Jankowitsch, Nagler, and Subrahmanyam, 2014), other researchers concerned with assessing how much mortgage interest rates in Italy are priced on credit risk as proxied by the probability of household mortgage delinquency, estimated by using the EU-Silc database (Magri and Pico, 2011).

There were other papers in which, due to opaque information and weak enforcement in emerging loan markets from 2012, the authors assessed the need for high collaterals, whereas borrowers lack adequate assets to pledge. For this, they found for a representative sample from Northeast Thailand where indeed most loans do not include any tangible assets as collateral (Menkhoff, Neuberger, and Rungruxsirivorn, 2012). We also found a paper that investigates the determinining factors of dispersion in interest rates on loans granted by banks to small and medium sized enterprises. The authors associated this dispersion with the loan officers' use of 'discretion' in loan rate setting process, and found that it was very important if: (i) loans were small and unsecured; (ii) firms were small and opaque; (iii) the firm operated in a large and highly concentrated banking market; and (iv) the firm was distantly located from the lender (Cerqueiro, Degryse, and Ongena, 2011). In the work developed by Cotugno, Monferrà, and Sampagnaro (2013), the authors examined the firms' credit availability during the 2007-2009 financial crisis using a dataset of 5,331 bank-firm relationships provided by borrower credit folders from three Italian banks. It aimed to test whether a strong lender-borrower relationship can produce less credit rationing for borrowing firms, even during a credit crunch period. And the final paper, which used only descriptive statistics in its analysis, provides the first systematic empirical analysis of how asymmetric information and competition in the credit market affect voluntary information sharing between lenders. Their study surrounded an experimental credit market in which information sharing can help lenders to distinguish good borrowers from bad ones (Brown and Zehnder, 2010).

There were also papers which actually developed either machine learning or statistic-based techniques but did not compare the result against what was tested. For instance, Cornett, McNutt, Strahan, and Tehranian (2011) studied how banks managed the liquidity shock that occurred during the financial crisis of 2007-2009 by adjusting their cash holdings and other liquid assets, as well as how these efforts to weather the storm affected credit availability. The authors then built a panel dataset from the quarterly Federal Financial Institutions Examination Council (FFIEC) Call Reports, which all regulated commercial bank files with their primary regulator. When the results were aggregated they found find that most of the decline in bank credit production during the height of the crisis could be explained by liquidity risk exposure.

Without comparing, but using machine learning techniques, Moradi and Rafiei (2019) used a fuzzy inference system to create a rule base using a set of uncertainty predictors. First, the authors trained an Adaptive Network-based Fuzzy Inference System (ANFIS) using monthly data from a customer profile dataset. Then, using the newly defined factors and their underlying rules, a second round of assessment began for the fuzzy inference system.

Papers were also found in which the methodology proposed by the authors themselves could not be categorized. These proposed techniques were not applied into any known database, and therefore they were not able to be compared. For example, Laeven, Levine and Michalopoulos (2015) proposed a technique through which entrepreneurs could earn profit by inventing better goods and profit-maximizing financiers arise to screen them. The model has two novel features: financiers engage in the costly but potentially profitable process of innovation (they can invent better methods for screening entrepreneurs); every screening process becomes less effective as technology advances. The model predicted that technological innovation and economic growth would eventually stop unless financiers started to innovate. Koopman, Kraussl, Lucas, and Monteiro (2009) used an intensity-based framework to study the relation between macroeconomic fundamentals and cycles in defaults and rating activity. By using Standard and Poor's U.S. corporate rating transition over the period 1980-2005, the authors estimated the default and rating cycle from micro data. They were able relate the business cycle, bank lending conditions, and financial market variables. They found that the macro variables appeared to explain part of the default cycle.

Wang et al. (2012) proposed an approach called RSFS (Random Subset Feature Selection), used for feature selection based on rough set and scatter search. In RSFS, conditional entropy is regarded as the heuristic to search for the optimal solutions. Two credit datasets in the UCI database were used to demonstrate the competitive performance of RSFS, which consisted in three credit models including Artificial Neural Networks (ANN), J48 Decision Trees (J48 DT), and Logistic Regression (LR). The experimental results showed that RSFS has a superior performance in saving the computational costs and improving classification accuracy. The last work, which had a proposed, untested technique, was a hybrid classification method based on rough sets, partial conditional probability assessments, and fuzzy sets. Their approach improved the classification capabilities of standard rough sets in credit risk (Capotorti and Barbanera, 2012).

There were papers which didn't have a technique itself or did not mention any throughout their content. Their applications varied, such as providing insight in credit risk. It might have helped practitioners to stay abreast of advancements in predictive modeling. From an academic point of view, the study provided an independent assessment of recent scoring methods and offered a new baseline to which future approaches can be compared (Lessmann, Baesens, Seow, and Thomas, 2015). Others assess the relationship between financial innovation, bank growth and fragility, and economic growth. The authors found that different measures of financial innovation are associated with faster bank growth, but also higher bank fragility and worse bank performance during the crisis (Beck, Chen, Lin, and Song, 2016). A discussion about inputs for direct marketing models was provided by describing the various types of used data, by determining the significance of the data, and by addressing the issue of selection of appropriate data (Bose and Chen, 2009). Authors also investigated the most influential evidence on the moral hazard effect of securitization, based on discontinuities in lender behavior at certain credit cores (Bubb and Kaufman, 2014).

Between the papers which did not compare results, there were the ones in which actual machine learning or statistic-based methods were applied to analyze the reasons why banks securitized on a large scale using the LR model, thus leading to indicate that liquidity and the search for improved performance are decisive factors in securitization (Cardone-Riportella, Samaniego-Medina, and Trujillo-Ponce, 2010). Some authors also examined state-level banking industry, as well as region economic determinants of non-performing loans for commercial banks and savings institutions by using both fixed effects and dynamic Generalized Method of Moments (GMM) estimations (Ghosh, 2015). Works also described an empirical study of instance sampling in predicting consumer repayment behavior, which evaluated the relative accuracies of logistic regression, discriminant analysis, DT (Decision Trees) and ANN on datasets created by gradually under- and over-sampling the good and bad, respectively (Crone and Finlay, 2012).

Another paper that applied linear programming was developed by Kwak, Shi, and Kou (2012). The authors proposed a Multiple Criteria Linear Programming (MCLP) method to predict bankruptcy, using Korean bankruptcy data after the 1997 financial crisis. The results of the MCLP approach in the Korean bankruptcy prediction study show that their method performed as well as traditional multiple discriminant analysis or logit analysis by using only financial data. In addition, this model's overall prediction accuracy is comparable to those of decision tree or support vector machine approaches.

In García, Marqués, and Sánchez, (2012) the authors did not use techniques to solve the credit risk problem. Their assessment involved dealing with the presence of noise and outliers in the training set, which may strongly affect the performance of the prediction model. Therefore, they systematically investigated whether the application of filtering algorithms leads to an increase in accuracy of instance-based classifiers in the context of credit risk assessment.

Machine Learning Applications

From the papers which used mainly machine learning techniques, Chi and Hsu (2012) selected important variables by using GA (Genetic Algorithm) to combine the bank's internal scoring model with the external credit bureau model to construct a dual scoring model for credit risk management. The results showed that the predictive ability of the dual scoring model outperforms both one-dimensional behavioral scoring and credit bureau scoring models.

Among other applications with machine learning techniques were Self-Organizing Maps (SOM), for a compact visualization of the complex behaviors in financial statements, in order to analyze the financial situation of companies over several years through a two-step clustering process (Chen, N., Ribeiro, Vieira, and Chen, A., 2013). ANN were also found among the selected papers, either to focus on enhancing credit risk models in three aspects –(i) optimizing the data distribution in datasets using a new method called Average Random Choosing; (ii) comparing effects of training-validation-test instance numbers; and (iii) finding the most suitable number of hidden units (Zhao et al., 2015)–, or combined with other techniques such as Support Vector Machines (SVM), K-Nearest Neighbours (kNN), and DT to provide some guidelines for the usage of databases, data splitting methods, performance evaluation metrics, and hypothesis testing procedures (García, Marqués, and Sánchez, 2014). And, finally, an application where a model based on binary quantile regression was proposed, using Bayesian estimation, called Bayesian Binary Quantile Regression (BBQR). The authors pointed out the distinct advantages of the latter approach: (i) the method provided accurate predictions of which customers may default in the future, (ii) the approach provided detailed insight into the effects of the explanatory variables on the probability of default, and (iii) the methodology was ideally suited to build a segmentation scheme of the customers in terms of risk of default and its corresponding uncertainty (Miguéis, Benoit, and Van Den Poel, 2013).

As for statistic-based techniques, there were probabilistic methods such as CPH (Cox Proportional Hazards) to reduce form models for credit risk in corporate lending, where the authors exploited the parallels between behavioral scores and ratings ascribed to corporate bonds (Malik and Thomas, 2010). Methods where the dependent variable was limited were also found, such as in LR for analyzing whether microfinance institutions can benefit from credit risk, been successfully adopted in retail banking (Van Gool, Verbeke, Sercu, and Baesens, 2012), or even a Probit Regression (PR) for suggesting that small firms low risk credit contracts with liquid collateral, which are their primary source of credit (Zambaldi, Aranha, Lopes, and Politi, 2011).

Considering the papers that used mainly AI (boosting techniques are not included on this section), there were works which aimed at the case of customers' default payments and compared the predictive accuracy of default probability (Yeh and Lien, 2009).

Other authors who described a credit risk evaluation system that used three supervised ANN models, each testing nine learning methods based on Back Propagation (BP) learning algorithm (Khashman, 2010), or even developed a heuristic algorithm, Hybrid Genetic Algorithm Neural Network (HGANN), which was used to identify an optimum feature subset and increase the classification accuracy in credit risk assessment (Oreski and Oreski, 2014).

Among the papers which used AI and involved these techniques as their best performance, then again, not considering the ones which applied boosting techniques, there were authors who proposed a three stage hybrid Adaptive Neuro Fuzzy System (ANFIS) credit risk model, which is based on statistical techniques and Neuro Fuzzy. Its performance was compared with conventional and commonly utilized models and showed its superiority (Akkoç 2012).

Also using AI and other techniques such as LR (Logistic Regression) and a hybrid algorithm, Cleofas-Sánchez, García, Marqués, and Sánchez (2016) explored hybrid associative memory with translation for default prediction. The performance of the hybrid associative memory with

translation is compared to four traditional neural networks, a support vector machine, and a logistic regression model in terms of their prediction capabilities.

Zhou, Lai, and Yu (2010) developed their research around testing 16 different methods and financial services datasets from companies in England. The authors found that Least Square Support Vector Machines (LSSVM) were the best performance method among other AI, statistics, and boosting techniques (combined or not). Also testing a variety of techniques, Loterman, Brown, Martens, Mues, and Baesens, (2012) showed a comparison of a total of 24 techniques using six real-life loss datasets from major international banks, where both LSSVM and ANN had the best overall performances.

Studying feature selection, Oreski, S., Oreski, D., and Oreski, G. (2012) investigated the extent to which the total data owned by a bank can be a good basis for predicting the borrower's ability to repay the loan on time, by using techniques such as Genetic Algorithm Neural Networks (GANN), Feature Selection Neural Networks (FSNN) and Generic Model for Parameters Optimization of the Artificial Neural Network (NNGM), where GANN had better accuracy than the others. Peng, Wang, Kou, and Shi, (2011) developed a two-step approach to evaluate classification algorithms for financial risk prediction. This method constructed a performance score to measure the performance of classification algorithms and introduced three Multiple Criteria Decision Making (MCDM) methods to provide a final ranking of classifiers. An empirical study was designed to assess various classification algorithms over seven real-life credit risk and fraud risk datasets from six countries where NBC (Naıve Bayes Classifiers) had better performance than the other tested methods.

Chen, Ma, and Ma (2009) proposed a hybrid support vector machine technique based on three strategies: (1) using Classification and Regression Trees (CART) to select input features, (2) using Multivariate Adaptive Regression Splines (MARS) to select input features, (3) using grid search to optimize model parameters. The authors tested their methods on a local bank and found that the hybrid of SVM + MARS was the best option to assess credit risk.

Having built several non-parametric credit risk models based on Multilayer Perceptron (MLP) and benchmarks of their performance against other models which employ the traditional Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and LR techniques, based on a sample of almost 5500 borrowers from a Peruvian microfinance institution, the results presented in Blanco et al. (2013) showed that NN (Neural Networks) models outperform the other three classic techniques both in terms of area under the receiver-operating characteristic curve (AUC) and as misclassification costs.

Harris (2015) investigated the practice of credit risk and introduced the use of the Clustered Support Vector Machine (CSVM) for credit scorecard development, comparing it with methods such as SVM, LinR (Linear Regression), LR, DT and ANN into datasets from Germany and Barbados. From among these techniques, CVSM was found to have a better performance than the rest of the techniques. There were works where four different types of hybrid models were compared by 'Classification + Classification', 'Classification + Clustering', 'Clustering + Classification', and 'Clustering + Clustering' techniques, respectively, applied on a Taiwan dataset where it was found that a Classification + Classification (LR + ANN) had a better performance than the other hybrids (Tsai and Chen, 2010).

Based on UK data from major retail credit cards, Bellotti and Crook (2012) built several models of Loss Given Default (LGD) based on account level data, including Tobit, a decision tree model, and a Beta and fractional logit transformation. The authors found that OLS (Ordinary Least Squares Estimation) models with macroeconomic variables perform best for forecasting LGD at the account and portfolio levels on independent hold-out data sets.

Lin (2009) proposed a new approach with three kinds of two-stage hybrid models of LR+ANN to explore if the two-stage hybrid model outperformed the traditional ones, and to construct a financial distress warning system for the banking industry in Taiwan. The results found factors for observable and total loans, allowance for doubtful accounts recovery rate, and interest-sensitive assets to liabilities ratio to be significantly related to the financial distress of banks in Taiwan. In the prediction of financially distressed, two-stage hybrid model (LR+ANN) giving the best performance with an 80% accuracy.

A work was also found which proposed a new type of multiple criteria CBR method for Binary Business Failure prediction (BFP) with Similarities to Positive and Negative Ideal Cases (SPNIC). The results indicate that this new CBR forecasting method can produce significantly better short-term discriminate capability than comparative methods, except for SVM, which had the best performance among the tested methods (Li, Adeli, Sun, and Han, 2011).

Wang and Ma (2012) also applied the SVM technique. Their research proposes a new hybrid ensemble approach called RSB-SVM, which is based on two popular ensemble strategies, i.e., bagging and random subspace, and uses a Support Vector Machine (SVM) as base learner. The enterprise's credit risk dataset, which included financial records from 239 companies and was collected by the Industrial and Commercial Bank of China, was selected by the authors to demonstrate the effectiveness and feasibility of the proposed method.

Other works in which the best performance involved SVM had their research either based on a comprehensive experimental comparison study over the effectiveness of learning algorithms such as ANN back propagation, Extreme Learning Machine (ELM), I-ELM, and SVM over a dataset consisting of real financial data from two corporate credit ratings not specified by the authors (Zhong, Miao, Shen, and Feng, 2014). Another one evaluated the performance of seven individual prediction techniques when used as members of five different ensemble methods, in order to suggest appropriate classifiers for each ensemble approach in the context of credit risk (Marqués, García, and Sánchez, 2012).

Some even tested only different models of SVM such as the work developed by Harris (2013), who had the research methodology based on credit-scoring models built using Broad (less than 90 days past due) and Narrow (greater than 90 days past due) default definitions.

Khashman (2009) presented a credit risk evaluation system that uses a NN model based on the back-propagation learning algorithm. Two types of ANN were tested: the first, using single hidden layers; and the second one, using two hidden layers. Analyzing the results, the author showed that the single hidden layer ANN outperformed the other method. This same author also tested six architectures of Emotional Back Propagation (EmBP) and six other ANN to investigate the efficiency of Emotional Neural Networks (EmNN) and compare their performance to conventional NNs when applied to credit risk evaluation. It was found that one of the ANN's tested architectures outperformed all the other applications (Khashman, 2011).

In Zhou, Jiang, Shi, and Tian, (2011) discussed that data mining and machine learning techniques such as SVM have been widely discussed in credit risk evaluation. The authors compared DM techniques against an optimization algorithm (kernel-based learning method called kernel affine subspace nearest point, KASNP) where they found that KASNP is an unconstrained optimal problem whose solution can be directly computed.

Iturriaga and Sanz (2015) developed a NN model to study the bankruptcy of US banks, taking into account the specific features of the recent financial crisis. The authors combined MLP and SOM to provide a tool that displays the probability of distress up to three years before bankruptcy occurs. Based on data from the Federal Deposit Insurance Corporation between 2002 and 2012, their results showed that failed banks are more concentrated in real estate loans and have more provisions. Thus, the best method to predict a non-failed bank would be ANN; to predict a failed one, SVM would be the best.

Research tried to describe what is a good or bad credit by evaluating it. The authors proposed three link analysis algorithms based on the process of SVM, to estimate an applicant's credit, so as to decide whether a bank should provide a loan. The proposed algorithms have two major phases which are called input weighted adjustor and class by SVM-based models. Among the four machine learning techniques tested, the authors found the best performance for their problem in using Hub Authority Ranking Applicants (HARA) (Xu, Zou, and Wang, 2009).

Hens and Tiwari (2012) proposed a strategy to reduce the computational time for credit risk. In this approach, the authors used SVM incorporated with the concept of reduction of features by using F score and taking a sample, instead of taking the whole dataset to create the credit risk model. The authors then compared their result with the one obtained from other methods. Their credit risk model was found to be very competitive with others due to its accuracy, as well as the fact that it takes both less computational time and that the Genetic Programing algorithm (GP) had the best performance.

Aiming to compare a new algorithm (recursive feature extraction with support vector machines, RFE-SVM) with well-known ML techniques, Derelioğlu and Gurgen (2011) proposed a knowledge discovery method that uses a MLP-based neural rule extraction (NRE) approach for credit risk analysis (CRA) of real-life small and medium enterprises (SMEs) in Turkey. In the first stage, the feature selection was achieved with the decision tree (DT), and recursive feature extraction with support vector machine (RFE-SVM) methods. The feature extraction was performed with factor analysis (FA) and principal component analysis (PCA). Then, the Continuous/Discrete Rule Extractor via Decision Tree Induction (CRED) algorithm is used to extract rules from the hidden units of a MLP for knowledge discovery.

Approaching different SVM methods, Danenas and Garsva (2015) combined CSVM, SVM and LSVM with external evaluation and sliding window testing, with focus on applications on larger datasets. The results showed that the CSVM technique had outperformed the others. In Chang and Yeh (2012), two experimental credit datasets were used to show the accuracy rate of the AINE classifier, applying a cross-validation method to evaluate its performance and compare it with other techniques. Experimental results showed that the AINE classifier is more competitive than SVM and hybrid-SVM classifiers.

In Khandani, Kim and Lo (2010), machine-learning techniques were applied to construct nonlinear, nonparametric forecasting models of consumer credit risk. By combining customer transactions and credit bureau data from January 2005 to April 2009 for a sample from a major commercial bank's customers, the authors were able to construct out-of-sample forecasts that significantly improved the classification rates of credit-card-holder delinquencies and defaults, with LR R2's of forecasted/realized delinquencies of 85%.

Hájek (2011) presented the modelling possibilities of NN on a complex real-world problem, i.e., municipal credit rating modelling. Testing ANN, Radial Basis Functions Neural Networks (RBF NN), Probabilistic Neural Networks (ProbNN), Cascade Correlations Neural Networks (CCNN), Group Method of Data Handling (GMDH), SVM, Multivariate Discriminant Analysis (MDA), LR, K-Means, and, finally, Classification Trees (CT), the results showed that the rating classes assigned to bond issuers can be classified with a high accuracy rate using a limited subset of input variable. Furthermore, the best technique for the proposed application would be ProbNN.

Tserng, Lin, Tsai, and Chen (2012) proposed an Enforced SVM-based model (ESVM model) for the default prediction in the construction industry using all available firm-years data in our ten-year sample period to solve the between-class imbalance. The empirical results of this paper show that the ESVM model always outperforms the logistic regression model and is more convenient to use because it is relatively independent of the selection of variables.

In Bijak and Thomas (2012), two-step approaches were applied, as well as a new, simultaneous method, in which both segmentation and scorecards were optimized at the same time: Logistic Trees with Unbiased Selection (LOTUS). For reference purposes, a single-scorecard model was used. The model performance measures were then compared to examine whether there was any improvement due to the used segmentation methods. Both CART and Chi-square automatic interaction detection (CHAID) had the best overall performance among the four tested models.

Koyuncugil and Ozgulbas (2012) also used the CHAID technique, while eveloping a financial early warning system through data mining, and SMEs were classified in 31 risk profiles. They also determined 2 financial early warning signs: profit before tax to owned funds and return on equity.

Also using the ML technique, Khemakem and Boujelbene (2018) used the Synthetic Minority Oversampling Technique (SMOTE). It was used to solve the problem of class imbalance and improve the performance of the classifier. The ANN and DT were designed to predict default risk. Results showed that profitability ratios, repayment capacity, solvency, duration of a credit report, guarantees, size of the company, loan number, ownership structure, and corporate banking relationship duration turned out to be the key factors in predicting default. Also, both algorithms were found to be highly sensitive to class imbalance. However, with balanced data, the decision trees displayed higher predictive accuracy for the assessment of credit risk than artificial neural networks.

As for mainly AI techniques tested in research from the last two years, the work developed by Li, Tian, Li, Zhou, and Yang (2017) was found. This paper extended studies in two main ways: firstly, it proposed a method involving machine learning to solve the reject inference problem; secondly, the Semi-Supervised Support Vector Machines (SSVM) model was found to improve the performance of scoring models compared to the industrial benchmark of LR.

In Bequé and Lessmann (2017), the authors explored the potential of ELM for consumer credit risk management. They found that ELM possesses some interesting properties, which might enable them to improve the quality of model-based decision support. To test this, they empirically compared ELM to established scoring techniques according to three performance criteria: ease of use, resource consumption, and predictive accuracy. The mathematical roots of ELM suggest that they are especially suitable as a base model within ensemble classifiers.

Kvamme, Sellereite, Aas, and Sjursen (2018) investigated, by using ANN, how transaction data can be used to assess credit risk. In a joint research with Norway's largest financial service group, DNB, they used transaction data to predict mortgage defaults. In 2012, the average Norwegian made 323 card transactions, where 71% of the value transferred was through debit payments. Hence, transactional data provided a useful description of user behavior, and subsequently consumer credit risk. Therefore, they predicted mortgage default by applying Convolutional Neural Networks (CNN) to consumer transaction data.

The main goal of Tavana, Abtahi, Caprio, and Poortarigh (2018) was the design of a system capable of warning about probable liquidity risk based only on raw data available in the bank's book or balance sheet without any predefined function. The implementation of two intelligent systems (ANN and Bayesian Neural Networks, BNN) comprised several algorithms and tests for validating the proposed model. A real-world case study was presented to demonstrate applicability and exhibit the efficiency, accuracy, and flexibility of data mining methods when modeling ambiguous occurrences related to bank liquidity risk measurement.

Another paper dealt with feature selection for credit risk assessment. Lahmiri (2016) aimed to compare several predictive models that combined feature selection techniques with data mining classifiers in the context of credit risk assessment, namely in terms of accuracy, sensitivity, and specificity statistics. The selected features werw used to train the SVM classifier, backpropagation neural network, radial basis function neural network, linear discriminant analysis and naive Bayes classifier.

Finally, the last paper that applied and had an AI method involved in its best performance was developed by Antonakis and Sfakianakis (2009). The authors examined the effectiveness of NBR as a method for constructing classification rules (credit scorecards) in the context of screening credit applicants (credit risk). For this purpose, the study used two real-world credit risk datasets to benchmark NBR against LDA, logistic regression analysis, k-nearest neighbours, classification trees, and neural networks. The results showed that, although NBR is definitely a competitive method, it was outperformed by CT and ANN applications.

Ensemble Techniques

Among the papers which used machine learning techniques, there were also the ones which showed that ensemble techniques were differential in order to make one method better than the other. For instance, Wang, Hao, Ma, and Jiang (2011) conducted a comparative assessment of the performance of three popular ensemble methods, i.e., Bagging, Boosting, and Stacking, based on four base learners, namely LR, DT, ANN, and SVM. Their experimental results revealed that the three ensemble methods can substantially improve individual base learners. Regarding the Australian database, the best performance was obtained by LR ,combined with Bagging. On the Chinese one, it was DT and Bagging, and only for the German database, the best performance method was SVM without ensemble techniques.

Twala (2010) explored the predicted behavior of five classifiers for different types of noise in terms of credit risk prediction accuracy, and how such accuracy could be improved by using classifier ensembles. Benchmarking results on four credit datasets and a comparison with the performance of each individual classifier on predictive accuracy at various attribute noise levels were presented. The experimental evaluation showed that the best overall performance was attributed to DT combined with feature selection algorithms and boosting techniques. As in Wang, G., Ma, Huang, and

Xu, (2012), two dual strategy ensemble trees were proposed: RS-Bagging DT and Bagging-RS DT, which were based on two ensemble strategies (Bagging and random subspace) in order to reduce the influence of noise data and redundant data attributes, as well as to get a relatively higher classification accuracy. Two real world credit datasets were selected to demonstrate the effectiveness and feasibility of proposed methods. Experimental results revealed that single DT gets the lowest average accuracy among five single classifiers, but, when combined with Bagging, things would go differently.

In Finlay (2011), the performance of several multiple classifier systems was evaluated in terms of their ability to correctly classify consumers as good or bad credit risks. Empirical results suggest that some multiple classifier systems deliver significantly better performance than the single best classifier, where ET Boost had better performance than others. Also assessing machine learning techniques for credit risk analysis, a research went one step beyond by introducing composite ensembles that jointly use different strategies for diversity induction. Accordingly, the combination of data resampling algorithms (Bagging and AdaBoost) and attribute subset selection methods (random subspace and rotation forest) for the construction of composite ensembles was explored, with the aim of improving prediction performance, where Bagging combined with RF had the best tested performance (Marqués et al., 2012).

The research developed by Florez-Lopez and Ramon-Jeronimo (2015) introduced an ensemble approach based on merged decision trees, the Correlated-Adjusted Decision Forest (CADF), to produce both accurate and comprehensible models. As its main innovation, this proposal explored the combination of complementary sources of diversity as mechanisms to optimize model structure, which led to a manageable number of comprehensive decision rules without sacrificing performance. The approach was evaluated in comparison to individual classifiers and alternative ensemble strategies (gradient boosting and random forests), and the best performance was developed by SVM and Gradient Boosting. However, empirical results suggested CADF might be an encouraging solution for credit risk problems, being able to compete in accuracy with more complex proposals while producing a rule-based structure directly useful for managerial decisions.

And finally, the last research which happened to involve ensemble techniques was developed by Abellán and Castellano (2017). The authors showed that a very simple base classifier attained a better trade-off in some aspects of interest for this type of studies, such as accuracy and area under the ROC curve (AUC). The AUC measure could be considered more appropriate in this ground, where different type of errors have different costs or consequences. The results presented this simple classifier as an interesting choice to be used as a base classifier in ensembles for credit risk and bankruptcy prediction, proving that individual performance of a classifier is not the only key point to be selected for an ensemble scheme. In six different datasets, a diversity of results were obtained. For instance, the best performance ensemble for the Australian database was MLP combined with Random Subspace; for the German one, LR with DECORATE; as for the Japanese, LR combined with Bagging; for the the Iranian, C4.5 (C4.5 Decision Tree) with Rotation Forest; for the Polish dataset, MLP with Bagging; and finally, for UCSD, the CDT method combined with Rotation Forest.

Now, as for the papers collected after the first selection, Xia Y., Liu C., Li, and Liu N. (2017) proposed a sequential ensemble credit risk model based on a Variant of Gradient Boosting Machine (i.e., Extreme Gradient Boosting, XGBoost). The tested methods were Tree-Structured Parzen Estimator (TPE), Random Search (RS), Grid Search (GS), Manual Search (MS), XGBoost, Gradient Boosting Decision Tree (GBDT), ANN, DT, LR, RF, and SVM.

Barboza, Kimura and Altman (2017) tested models to predict bankruptcy one year in advance, and compare their performance with results from SVMLin, SVM RBF, MDA, LR, ANN, Boosting, Bagging, and RF by using data from 1985 to 2013 on North American firms. Comparing the best models, with all predictive variables, the ensemble with RF led to an 87% accuracy, whereas logistic regression and linear discriminant analysis led to 69% and 50%, respectively, in the testing sample.

Another case where an ensemble technique combined with rule-based machine learning happened to have the best results is shown by Sun, Lang, Fujita, and Li (2018). In that paper, different times of iteration for base DT classifier training, new positive (high-risk) samples were produced to different degrees by SMOTE with Differentiated Sampling Rates (DSR), and different numbers of negative (low-risk) samples are drawn with replacement by Bagging with DSR. The experimental results indicate that DTE-SBD (Decision Tree Ensemble based on SMOTE, Bagging and DSR) significantly outperforms the other five models and is effective for imbalanced enterprise credit evaluation.

Also among the papers was the introduction of Deep-Belief Network (DBN) as a credit rating algorithm to generate fast and accurate individual classification results, compared with more traditional methods such as SVM, MLP and Multinomial Logistic Regression (MLR) (Luo, Wu, and Wu, 2017). The goal of the paper was to provide a set of descriptive results and tests that lay a foundation for future theoretical and empirical work on DBN in credit risk in Credit Default Swap (CDS) markets. The authors investigated the performances of different credit risk models by conducting experiments on a collection of CDS data.

Another research about XGBoost was also found, this time as CSXGBoost (Cost-Sensitive Extension of XGboost). In the work, developed by Xia, Liu C., and Liu N. (2017). The authors proposed a cost-sensitive boosted tree loan evaluation model by incorporating cost-sensitive learning and XGBoost to enhance the capability of discriminating potential default borrowers. Therefore, a portfolio allocation model that converts the portfolio optimization problem into an integer linear programming was proposed as a decision support system for unprofessional lenders.

Xia, Liu, Da, and Xie (2018) propose a novel heterogeneous

ensemble credit model (RF, XGBoost, and MV, Majority Voting) that integrated the Bagging algorithm with the stacking method. The proposed model differs from the extant ensemble credit models in three aspects: pool generation, selection of base learners, and trainable fuser. To confirm the efficiency of this proposed approach, a wide range of models, including individual classifiers and homogeneous and heterogeneous ensemble models, were introduced as benchmarks.

Rule-Based Machine Learning

Besides ensemble and AI techniques, there were papers which had a rule-based machine learning algorithm applied in its research, such as the one by Huysmans, Dejaeger, Mues, Vanthienen, and Baesens (2011), who, based on a number of observations, constructed a decision table model that allowed the analysts to provide classifications or predictions for new observations. The Decision Table (Dtab) algorithm was compared with the DT technique. The first one had a superior performance. As for DT as the best performance technique, we found the research of Paleologo, Elisseeff and Antonini (2010), where several classification techniques were shown to perform well on credit risk – e.g. support vector machines. While the investigation of better classifiers is an important research topic, the specific methodology chosen in real-world applications has to deal with the challenges arising from the data collected within the industry.

Also, algorithms based on swarm optimization, such as Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO), were also found among the selected papers. A study used two nature-inspired methods (ACO and PSO) for this credit risk assessment. The modelling context was developed, and its performance of the methods tested in two financial classification tasks involving credit risk assessment and audit qualifications. ACO was proposed in this study for solving this feature subset selection problem. These two nature-inspired techniques had the best performance among the others (Tabu Search, TS, and GA).

Nature-inspired methods are approaches used in various fields for the solution for a number of problems. Marinaki, Marinakis and Zopounidis (2010) used a nature-inspired method, namely Honey Bee Mating Optimization (HBMO), that was based on the mating behavior of honey bees for a financial classification problem. Being compared with PSO, ACO, GA, and TS, the HBMO method had the best performance for the analyzed problem.

Vukovic, Delibasic, Uzelac, and Suknovic (2012) proposed a Case-Based Reasoning (CBR) model that used preference theory functions for similarity measurements between cases. As it is hard to select the right preference function for every feature and set the appropriate parameters, a genetic algorithm was used to choose the right preference functions, or more precisely, to set the parameters of each preference function, such as setting attribute weights. The proposed model was compared to the well-known k-NN model, based on the Euclidean distance measure. It was evaluated on three different benchmark datasets, while its accuracy was measured with 10-fold cross-validation tests. The experimental results show that the proposed approach can, in some cases, outperform the traditional k-NN classifier.

In Kruppa, Schwarz, Arminger, and Ziegler (2013) a general framework was presented to estimate individual consumer credit risks by means of machine learning methods. Since a probability is an expected value, all nonparametric regression approaches which are consistent for the mean are consistent for the probability estimation problem. Among others, random forests RF, KNN, and Bagged k-Nearest Neighbors (bagged bNN) belong to this class of consistent nonparametric regression approaches. From the tested algorithms, RF had a better development and performance than the rest of the methods.

Zhou, Lu, and Fujita (2015) investigated the performance of different financial distress prediction models with feature selection approaches based on domain knowledge or data mining techniques. The empirical results showed that there is no significant difference between the best classification performance of models with feature selection guided by data mining techniques and the ones guided by domain knowledge.

Sánchez-Lasheras, de Andrés, Lorca, and de Cos Juez (2012) proposed a new approach to firm bankruptcy forecasting. Their proposal was a hybrid method in which sound companies were divided in clusters using SOM. Each cluster was then replaced by a director vector which summarized all of them. Once the companies in clusters had been replaced by director vectors, the authors estimated a classification model through MARS.

Considering now the second batch of papers from the past two years, Lanzarini, Villa Monte, Bariviera, and Jimbo Santana (2017) presented an alternative method that could generate rules that work not only on numerical attributes but also on nominal ones. The key feature of this method, called Learning Vector Quantization and Particle Swarm Optimization (LVQ + PSO), was their finding of a reduced set of classifying rules. Their findings indicate that the reduced quantity of rules made this method useful for credit officers aiming to make decisions about granting a credit.

Statistical Methods Applications

As for the last portion of the analyzed papers, there were the ones in which statistical methods were involved in achieving the best performance. Initially, there were papers which did not compare methods, such as Louzis, Vouldis, and Metaxas (2011); Tinoco and Wilson (2013); and Ferreira, Santos, Marques, and Ferreira J. (2014). The first work was motivated by the hypothesis that both macroeconomic and bank-specific variables have an effect on loan quality and that these effects vary between different loan categories. By applying GMM, the results showed that, for all loan categories, NPLs in the Greek banking system can be explained mainly by macroeconomic variables (GDP, unemployment, interest rates, and public debt) and management quality. In Tinoco and Wilson (2013), using a sample of 23,218 company-year

observations of listed companies during the period 1980-2011, the paper investigated empirically, using LR, the utility of combining accounting, market-based and macro-economic data to explain corporate credit risk. The paper developed risk models for listed companies that predict financial distress and bankruptcy. In Ferreira et al. (2014), the authors proposed a methodological framework allowing for the readjustment of trade-offs within risk evaluation criteria, considered of extreme importance in the lending decision process of mortgage loans. Measuring attractiveness is performed with a categorical based evaluation technique (MACBETH) to a pre-established structure of credit-scoring criteria for mortgage lending risk evaluation. This pre-established structure was used by one of the largest banks in Portugal and the framework allowed the authors to provide credit experts who participated in the study with a more informed, transparent and accurate mortgage lending risk evaluation system.

Following the papers which had statistical methods as best performance algorithms, there were the ones which actually compared different techniques. In Yu, Wang, and Lai (2009) a novel intelligent-agent-based fuzzy Group Decision Making (GDM) model was proposed as an effective Multicriteria Decision Analysis (MCDA) tool for credit risk evaluation. For comparison, the authors also tested the original GDM, SVMR (Support Vector Machines Regression), RBF NN, Back Propagation Neural Networks (BPNN), LR, and LinR. Finally, the authors found that the novel method had the best performance among the tested algorithms. Andrés, Lorca, de Cos Juez, and Sánchez-Lasheras (2011) proposed a hybrid system which combines fuzzy clustering and MARS. Both models were especially suitable for the bankruptcy prediction problem, due to their theoretical advantages when the information used for the forecasting is drawn from company financial statements. The authors tested the accuracy of their approach in a real setting consisting of a database made up of 59,336 non-bankrupt Spanish companies and 138 distressed firms which went bankrupt during 2007, and found that the hybrid Fuzzy C-Means, combined with MARS, had the best performance.

Six papers were found in which LR was the best technique. One of them assessed LR and compared it with SVM, LDA and kNN on a large credit database (Bellotti and Crook, 2009). Another one investigated whether productive inefficiency measured as the distance from the industry's 'best practice' frontier is an important *ex-ante* predictor of business failure; there was research that tested DEA (Data Envelopment Analysis) and LR as its methodology (Psillaki, Tsolas, and Margaritis, 2010). Using data mining to improve the assessment of credit worthiness using credit risk models, Yap, Ong and Husain (2011) compared the classification performance of the credit scorecard model, the LR model, and the DT model. The classification error rates for credit scorecard model, logistic regression and decision tree were 27.9%, 28.8% and 28.1%, respectively.

Kou, Peng, and Wang (2014) presented an MCDM-based (Multiple Criteria Decision Making) approach to rank a selection of popular clustering algorithms in the domain of financial risk analysis. An experimental study is designed to validate the proposed approach using three MCDM methods, six clustering algorithms, and eleven cluster validity indices from three real-life credit risk and bankruptcy risk datasets. The results demonstrate the effectiveness of MCDM methods in evaluating clustering algorithms and indicate that the repeated bisection method leads to good 2-way clustering solutions on the selected financial risk datasets.

Tong, Mues and Thomas (2012), estimated a mixture cure model predicting time to default on a UK personal loan portfolio, and compare its performance against the Cox Proportional Hazards (CPH) method and standard logistic regression. Following their experimental results, the authors found that standard LR performed better than CPH. Lessmann and Vob (2009) proposed a hierarchical reference model for SVM-based classification in this field. The approach balances the conflicting goals of transparent, yet accurate models, and compares favorably to alternative classifiers in a large-scale empirical evaluation in real-world customer relationship management applications. Among all tested models (RBF SVM, SVM, LR and CART), the LR algorithm had the better performance.

The last paper which had LR as its best performing algorithm was developed by Bekhet and Eletter (2014), where two credit risk models using data mining techniques to support loan decisions for the Jordanian commercial banks were proposed. For this research, algorithms such as LR and RBF NN were tested; the first one had better performance than the other.

Tsai, Lin, Cheng, and Lin P.(2009), constructed the consumer loan default predicting model by conducting an empirical analysis on the customers of unsecured consumer loans from a certain financial institution in Taiwan, and adopted the borrower's demographic variables and money attitude as real-time discriminant information. Furthermore, the authors used four predicting methods, such as Discriminant Analysis (DA), LR, ANN and DEA–DA, to compare their suitability. The results showed that DEA–DA and NN possessed better predicting capability, with DEA-DA being better than the second one. Thus, they proved to be the optimal predicting models that this study was longing for.

In Wu, Olson, and Luo (2014), three different approaches were used: artificial intelligence (ANN); rule-based machine learning (DT) and statistical models (LR). The paper described and demonstrated a model to support risk management of accounts receivable. Accuracy results of this model were presented, enabling accounts receivable managers to confidently apply statistical analysis through data mining to manage the risk.

Zhang, Gao, and Shi (2014) proposed a novel Multi-Criteria Optimization Classifier based on Kernel, Fuzzification, and Penalty factors (KFP-MCOC). Firstly, a kernel function was used to map input points into a high-dimensional feature space. Then an appropriate fuzzy membership function was introduced to MCOC and associated with each data point in the feature space, and the unequal penalty factors were added to the input points of imbalanced classes. The experimental results of credit risk evaluation and their comparison with MCOC, SVM and fuzzy SVM showed that KFP-MCOC could

enhance the separation of different credit applicants, the efficiency of credit risk scoring, and the generalization of predicting the credit rank of a new applicant.

Dong, Lai and Yen (2010), tried to improve the prediction accuracy of logistic regression by combining it with random coefficients. The LRR model showed to improve LR prediction accuracy without sacrificing desirable features. Finally, the last research to be analyzed in this paper was developed by Zhu et al. (2013), where the objective was to put forward a classification approach named Classification Technique for Order Preference by Similarity to Ideal Solution (C-TOPSIS). It is based on the rationale of Technique for Order Preference by Similarity to Ideal Solution TOPSIS), which is famous for reliable evaluation results and quick computing processes, and it is easy to understand and use. In comparison with 7 popular approaches on 2 widely used UCI credit datasets, C-TOPSIS ranked 2nd in accuracy, 1st in complexity, and 1st rank in interpretability. Only C-TOPSIS ranked among the top 3 in all the three aspects, which verified that C-TOPSIS could balance them well.

Considering now the second search of papers (from the past 2 years) where Statistical methods had better performance than the others compared within the research, Maldonado, Bravo, López, and Pérez (2017) proposed a profit-driven approach for classifier construction and simultaneous variable selection based on SVMLin. Their proposal incorporates a group penalty function in the SVM formulation in order to simultaneously penalize the variables that belong to the same group. The framework used algorithms such as Recursive Feature Elimination Support Vector Machines (RFE-SVM), Holdout Support Vector Machines (HOSVM), SVM, Logit Regression, and Fisher Score (FS). It was then studied in a credit risk problem for a Chilean bank, and it led to superior performance with respect to business-related goals.

Finally, Dirick, Claeskens and Baesens (2017) contributed to the existing literature by analyzing ten different data sets from five banks, using both statistical (CPH) and economic evaluation measures (Accelerated Failure Time, AFT), applicable to all considered model types: the "plain" survival models, as well as the mixture cure models.

With that last paper, we are able to bring the content analysis from all the collected research to a close. In the next section, the research questions will be answered, based on the findings of this analysis.

Answering the research questions

As shown at the beginning of this paper, two main questions were asked in order to direct this research. They are discussed below.

Are machine learning techniques being effectively applied in research about credit risk evaluation?

At the start of the analysis, a total of 102 different techniques were found, among them, statistical techniques, boosting methods, MCD makers, multivariate analysis, but mostly machine learning techniques. Those techniques were classified in two main groups: Statistic-Based and Machine Learning, as shown below in Tables 7 and 8, respectively.

Table 7. Summary of Statistic-Based Methods used by the authors, where the acronyms are in Table 4 (legend)

Multivariant Analysisr	**LDA, DA, DLDA, QDA, DQDA, MVA, MARS, MDA**
Dependent Variable Limited	LR, LRA, PR, LD, RiR, RoR, LLR, LRF, LRR, Logit Regression, FS
Probabilistic Methods	MCM, CPH
Non-Linear Regression	GMM
Linear Regression	LinR, BR
Non-Parametric Statistics	DEA
Univariate Analysis	Univariate Analysis
Discriminant Analysis	Discriminant Analysis
Sampling Techniques	SMOTE, DSR
Multiple Criteria Decision Making	MCDM, TOPSIS, PROMETHEE, VIKOR, MCOC, KFP-MCOC, C-TOPSIS

Source: Authors

Table 8. Summary of Machine Learning Techniques Applied to the selected papers where the acronyms are in Table 4 (legend)

Rule-Based Machine Learning	
	DT, DTab, CT, GDM, CART, ID3, EM, RSFS, RT, HARA, HubAvgRA, ATkRA, ACO, PSO, GA, TS, HBMO, HGADSM, bNN, SOM, FSOM, TSOM, RSM, C4.5, GP, 1-NN, CADF, CDT, kNN, Chi-Square Automatic Interaction Detection, RS, GS, MS, GBDT, RF, DTE-SBD, DBN, CSLR-T, CSRF-T, CSLR-SMOTE, CSRF-SMOTE, J48, LVQ+PSO, GPC
Boosting Techniques	ET Boost, Bagging, AdaBoost, Gradient Boosting, Random Subspace, DEC-ORATE, Rotation Forrest, XGBoost, CSXGBoost
Artificial Intelligence	SVM, SVMLin, SVM RBF, ANN, NB, SVMR, RBF NN, BPNN, HGANN, ANFIS, ProbNN, LSSVMRBF, LSSVM-Lin, 1nLSSVMRBF, SVMLin, SVMRBF, NBC, GANN, FSNN, NNGM, MLP, OLS, B-OLS, BC-OLS, CCNN, GMDH, ELM, 1-ELM, EmBP, SHNN, DHNN, AINE, AIRS, SAIS, ESVM, BBQR, Fuzzy C-Means, RFE-SVM, HOSVM, CNN

Source: Authors

From those methods, we were able to identify around 93 machine learning-based techniques, which outnumbered the 36 different statistic-based techniques. Those allowed us to answer this first question, concluding that it is agreeable to assume to the premise which surrounds the high usage of machine learning techniques durin the past ten years of research in Credit Risk Analysis.

Regarding effectiveness, from the 102 papers analyzed, 72 of them used machine learning techniques at some point. From

those 72 papers, 57 involved machine learning or a machine learning hybrid as their best performance technique,. That shows us the effectuality of these techniques and answers our first research question.

Which of these quantitative techniques have been mostly applied over the last ten years of research?

As for the most applied types of quantitative methods, it was found that the use of AI techniques prevailed. Considering Machine Learning techniques, 72 papers used these types of algorithms, and the papers that applied one AI method amounted to 57.

Regarding which AI techniques were used, it was found that ANN was the most applied. This technique appeared 47 times in the papers, either comparing different architectures or different types of ANN. Following ANN, there were the SVM techniques, which appeared 33 times along the review.

Considering Machine Learning methods apart from AI-related ones, e.g. rule-based algorithms, the most used was DT with 16 applications, and kNN followed, appearing in 11 documents. And, finally, concerning boosting techniques, both Bagging and AdaBoost were the most common among the studied papers.

All things considered, the most common technique was ANN, being extensively applied among the found papers, either in machine learning applications or overall techniques.

Conclusions

At the beginning of this research, two questions were presented surrounding credit risk research and the applied methods in order to successfully assess the problem. The first question aimed to determine whether machine learning techniques were being effectively applied in research about credit risk evaluation, and the second one, which of these quantitative techniques have been mostly applied over the last ten years of research.

As expected by the authors, the number of research papers using AI overcame other types of techniques, but more recent papers used less of these methods, suggesting that other approaches are being more accurate than what AI can provide.

Another possible reason why this expectation was not fulfilled happens to concern the filters and techniques; only papers with a higher volume of citations were selected, which could lead to older research. Moreover, this work avoided papers that were used more as a concept review than actually being innovative or showing what actually happens regarding machine learning in credit risk assessment.

An extensive literature review was presented with a protocol including different selection criteria for analyzing papers from three different databases. After the sample was collected, the content analysis was preceeded by a bibliometric review, presenting the journals and keywords. Following this step, the true content of each selected paper was reviewed both in the form of Tables 6 and 7 and the description of the main points of every research.

During the discussion presented above, every amount of different techniques was assessed, and, through that, we were able to find that, not only AI techniques were more applied than the others found, but also ANN is the most common type of AI method found among the papers.

Within the discussion, statistic-based techniques were also assessed, showing that LR is the most common between them. This is reasonable, since the nature of the problem demands for algorithms that are able to classify different client profiles for the decision-maker be able to best select the suiters for the bank credit.

As for future work, other systematic reviews may be developed focusing on AI methods for credit risk assessment, questioning differences between it, and other types of problems involving bank issues. Another option would be to use the reviewed datasets and test different hybrids in order to extend the knowledge barrier of this problem, thus stepping forward in the development of solutions for this type of problem.

Acknowledgments

This study was partially funded by PUCPR and by the Coordination for the Improvement of Education Personnel – Brazil (CAPES, represented by thefirst author) and by the National Council for Scientific and Technological Development – Brazil (CNPq, represented by second).

References

Abellán J., Castellano J.G. (2017). A comparative study on base classifiers in ensemble methods for credit risk. *Expert Systems with Applications, 73*, 1-10. 10.1016/j.eswa.2016.12.020

Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three-stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit risk analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222*(1), 168-178. 10.1016/j.ejor.2012.04.009

Andrés, J., Lorca, P., de Cos Juez, F. J., and Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications, 38*(3), 1866-1875. 10.1016/j.eswa.2010.07.117

Antonakis A.C., and Sfakianakis M. E. (2009). Assessing naıve Bayes as a method for screening credit applicants. *Journal of applied Statistics, 36*(5), 537-545. 10.1080/02664760802554263

Barboza F., Kimura H., and Altman E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications. 83*, 405-417. 10.1016/j.eswa.2017.04.006

Beck, T., Chen, T., Lin, C., and Song, F.M. (2016). Financial innova-tion: The bright and the dark sides. *Journal of Banking & Finance, 72*, 28-51. 10.1016/j.jbankfin.2016.06.012

Bekhet, H. A. and Eletter, S. F. K. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Review of Development Finance, 4*(1), 20-28. 10.1016/j.rdf.2014.03.002

Bellotti, T. and Crook, J. (2009). Support vector machines for credit risk and discovery of significant features. *Expert Systems with Applications, 36*(2), 3302-3308. 10.1016/j.eswa.2008.01.005

Bellotti, T. and Crook, J. (2012). Loss given default models incorpo-rating macroeconomic variables for credit cards. *International Journal of Forecasting, 28*(1), 171-182. 10.1016/j.ijforecast.2010.08.005

Bequé A. and Lessmann S. (2017). Extreme learning machines for credit risk: An empirical evaluation. *Expert Systems with Applications, 86*, 42-53. 10.1016/j.eswa.2017.05.050

Bijak K. and Thomas L.C. (2012). Does segmentation always im-prove model performance in credit risk? *Expert Systems with Applications, 39*(3), 2433-2442. 10.1016/j.eswa.2011.08.093

Blanco, A., Pino-Mejías, R., Lara, J., and Rayo, S. (2013). Credit risk models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications, 40*(1), 356-364. 10.1016/j.eswa.2012.07.051

Bose, I. and Chen, X. (2009). Quantitative models for direct market-ing: A review from systems perspective. *European Journal of Operational Research*. *195*(1), 1-16. 10.1016/j.ejor.2008.04.006

Brown, M. and Zehnder, C. (2010). The emergence of information sharing in credit markets. *Journal of Financial Intermediation, 19*(2), 255-278. 10.1016/j.jfi.2009.03.001

Bubb, R. and Kaufman, A. (2014). Securitization and moral hazard: Evidence from credit score cutoff rules. *Journal of Monetary Economics, 63*, 1-18. 10.1016/j.jmoneco.2014.01.005

Capotorti A. and Barbanera E. (2012). Credit risk analysis using a fuzzy probabilistic rough set model. *Computational Statistics & Data Analysis*. *56*(4), 981-994. 10.1016/j.csda.2011.06.036

Cardone-Riportella, C., Samaniego-Medina, R., and Trujillo-Ponce, A. (2010). What drives bank securitisation? The Spanish experience. *Journal of Banking & Finance, 34*(11), 2639-2651. 10.1016/j.jbankfin.2010.05.003

Cerqueiro, G., Degryse, H., and Ongena, S. (2011). Rules versus dis-cretion in loan rate setting. *Journal of Financial Intermediation*. *20*(4), 503-529. 10.1016/j.jfi.2010.12.002

Chang, S.-Y. and Yeh, T.-Y. (2012). An artificial immune classifier for credit risk analysis. *Applied Soft Computing, 12*(2), 611-618. 10.1016/j.asoc.2011.11.002

Chen W., Ma C., and Ma L. (2009). Mining the customer credit using hybrid support vector machine technique. *Expert systems with applications, 36*(4), 7611-7616. 10.1016/j.eswa.2008.09.054

Chen, N., Ribeiro, B., Vieira, A., and Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert systems with applications, 40*(*1*), 385-393. 10.1016/j.eswa.2012.07.047

Chi B.-W. and Hsu C.-C. (2012). A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit risk model. *Expert systems with applications, 39* (3), 2650-2661. 10.1016/j.eswa.2011.08.120

Cleofas-Sánchez L., García V., Marqués A.I., and Sánchez J.S. (2016). Financial distress prediction using the hybrid associa-tive memory with translation. *Applied Soft Computing 44*, 144-152 10.1016/j.asoc.2016.04.005

Cornett, M. M., McNutt, J. J., Strahan, P. E., and Tehranian, H. (2011). Liquidity risk management and credit supply in the financial crisis. *Journal of Financial Economics, 101*(2), 297-312. 10.1016/j.jfineco.2011.03.001

Cotugno, M., Monferrá, S., and Sampagnaro, G. (2013). Relation-ship lending, hierarchical distance and credit tightening: Ev-idence from the financial crisis. *Journal of Banking & Finance, 37* (5), 1372-1385. 10.1016/j.jbankfin.2012.07.026

Crone, S. F. and Finlay, S. (2012). Instance sampling in credit risk: An empirical study of sample size and balancing. *International Journal of Forecasting, 28*(1), 224-238. 10.1016/j.ijforecast.2011.07.006

Danenas, P., Garsva, G. (2015). Selection of Support Vector Machines based classifiers for credit risk domain. *Expert systems with applications, 42*(6), 3194-3204. 10.1016/j.eswa.2014.12.001

Derelioğlu G., Gurgen F. (2011). Knowledge discovery using neural approach for SME's credit risk analysis problem in Turkey. *Expert Systems with Applications, 38*(8) 9313-9318. 10.1016/j.eswa.2011.01.012

Dirick L., Claeskens G., and Baesens B. (2017). Time to default in credit risk using survival analysis: A benchmark study. *Journal of the Operational Research Society, 68*(6),652-665. 10.1057/s41274-016-0128-9

Dong G., Lai K.K., Yen J. (2010). Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science, 1*(1), 2463-2468. 10.1016/j.procs.2010.04.278

Ferreira F. A. F., Santos S. P., Marques C. S. E., and Ferreira J. (2014). Assessing credit risk of mortgage lending using MACBETH: A methodological framework. *Management Decision, 52*(2),182-206. 10.1108/MD-01-2013-0021

Finlay, S. (2011). Multiple classifier architectures and their appli-cation to credit risk assessment. *European Journal of Operational Research, 210*(2), 368-378. 10.1016/j.ejor.2010.09.029

Firth, M., Lin, C., Liu, P., and Wong, S. M. L. (2009.) Inside the black box: Bank credit allocation in China's private sector. *Journal of Banking & Finance, 33*(6), 1144-1155. 10.1016/j.jbankfin.2008.12.008

Florez-Lopez R. and Ramon-Jeronimo J.M. (2015). Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal. *Expert Systems with Applications, 42*(13), 5737-5753. 10.1016/j.eswa.2015.02.042

G., Peng Y., and Wang G. (2014). Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Information Sciences, 275*(10) 1-12. 10.1016/j.ins.2014.02.137

García V., Marqués A.I., and Sánchez J.S. (2014). An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. *Journal of Intelligent Information Systems, 44*(1), 159-189. 10.1007/s10844-014-0333-4

García, V., Marqués, A. I., and Sánchez, J. S (2012). On the use of data filtering techniques for credit risk prediction with in-stance-based models. *Expert Systems with Applications, 39*(18), 13267-13276. 10.1016/j.eswa.2012.05.075

Ghosh, A. (2015). Banking-industry specific and regional economic determinants of non-performing loans: Evidence from US states. *Journal of Financial Stability, 20*, 93-104. 10.1016/j.jfs.2015.08.004

Guo Y., Zhou W., Luo C., Liu C., and Xiong H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research, 249*(2) 417-426. 10.1016/j.ejor.2015.05.050

Hájek, P. (2011). Municipal credit rating modelling by neural networks. *Decision Support Systems, 51*(1), 108-118. 10.1016/j.dss.2010.11.033

Harris T. (2013). Quantitative credit risk assessment using support vector machines: Broad versus Narrow default definitions. *Expert Systems with Applications, 40*(11), 4404-4413. 10.1016/j.eswa.2013.01.044

Harris T. (2015). Credit risk using the clustered support vector machine. *Expert Systems with Applications, 42*(2), 741-750. 10.1016/j.eswa.2014.08.029

Hens, A. B., and Tiwari, M. K. (2012). Computational time reduction for credit risk: An integrated approach based on support vector machine and stratified sampling method. *Expert Systems with Applications, 39*(8), 6774-6781. 10.1016/j.eswa.2011.12.057

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., and Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems, 51*(1), 141-154. 10.1016/j.dss.2010.12.003

Iturriaga, F. J. L., and Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications, 42*(6), 2857-2869. 10.1016/j.eswa.2014.11.025

Jankowitsch, R., Nagler, F., and Subrahmanyam, M. G. (2014). The determinants of recovery rates in the US corporate bond market. *Journal of Financial Economics, 114*(1),155-177. 10.1016/j.jfineco.2014.06.001

Jiménez, G., Salas, V., and Saurina, J. (2009). Organizational dis-tance and use of collateral for business loans. *Journal of Banking & Finance, 33*(2), 234-243. 10.1016/j.jbankfin.2008.07.015

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance, 34*(11), 2767-2787. 10.1016/j.jbankfin.2010.06.001

Khashman A. (2009). A neural network model for credit risk evaluation. *International Journal of Neural Systems, 19*(4), 285-294. 10.1142/S0129065709002014

Khashman A. (2011). Credit risk evaluation using neural net-works: Emotional versus conventional models. *Applied Soft Computing, 11*(8), 5477-5484. 10.1016/j.asoc.2011.05.011

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications, 37*(9), 6233-6239. 10.1016/j.eswa.2010.02.101

Koopman, S. J., Kraussl, R., Lucas, A., and Monteiro, A. B. (2009). Credit cycles and macro fundamentals. *Journal of Empirical Finance, 16*(1), 42-54. 10.1016/j.jempfin.2008.07.002

Koyuncugil A.S. and Ozgulbas N. (2012). Financial early warning system model and data mining application for risk detection. *Expert systems with Applications, 39*(6), 6238-6253. 10.1016/j.eswa.2011.12.021

Kruppa J., Schwarz A., Arminger G. and Ziegler A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications, 40*(13), 5125-5131. 10.1016/j.eswa.2013.03.019

Kvamme H., Sellereite N., Aas K., and Sjursen S. (2018). Predicting mortgage default using convolutional neural networks. *Expert Systems with Applications, 102*, 207-217. 10.1016/j.eswa.2018.02.029

Kwak, W., Shi, Y., and Kou, G. (2012). Bankruptcy prediction for Korean firms after the 1997 financial crisis: using a multiple criteria linear programming data mining approach. *Review of Quantitative Finance and Accounting, 38,* 441-453. 10.1007/s11156-011-0238-z

Laeven, L., Levine, R., and Michalopoulos, S. (2015). Financial inno-vation and endogenous growth. *Journal of Financial Intermediation, 24*(1), 1-24. 10.1016/j.jfi.2014.04.001

Lahmiri S. (2016). Features selection, data mining and finacial risk classification: a comparative study. *Intelligent Systems in Accounting, Finance and Management, 23*(4) 265-275. 10.1002/isaf.1395

Lanzarini L.C., Villa Monte A., Bariviera A.F., and Jimbo Santana P. (2017). Simplifying credit risk rules using LVQ + PSO. *Kybernetes, 46* (1), 8-16. 10.1108/K-06-2016-0158

Lee, N., Sameen, H., and Cowling, M. (2015). Access to finance for innovative SMEs since the financial crisis. *Research Policy, 44*(2) 370-380. 10.1016/j.respol.2014.09.008

Lessmann, S. and Vob, S. (2009). A reference model for customer-centric data mining with support vector machines. *European Journal of Operational Research, 199*(2), 520-530. 10.1016/j.ejor.2008.12.017

Lessmann, S., Baesens, B., Seow, H.-V., and Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit risk: An update of research. *European Journal of Operational Research, 247*(1), 124-136. 10.1016/j.ejor.2015.05.030

Li Z., Tian Y., Li K., Zhou F., and Yang W (2017). Reject inference in credit risk using Semi-supervised Support Vector Machines. *Expert Systems with Applications, 74*, 105-114. 10.1016/j.eswa.2017.01.011

Li, H., Adeli, H., Sun, J., and Han, J.-G. (2011) Hybridizing principles of TOPSIS with case-based reasoning for business failure prediction. *Computers & Operations Research, 38*(2), 409-419. 10.1016/j.cor.2010.06.008

Lin, S. L. (2009) A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications, 36*(4) 8333-8341. 10.1016/j.eswa.2008.10.015

Loterman, G., Brown, I., Martens, D., Mues, C., and Baesens, B. (2012) Benchmarking regression algorithms for loss given de-fault modeling . *International Journal of Forecasting, 28*(1), 161-170. 10.1016/j.ijforecast.2011.01.006

Louzis, D. P., Vouldis, A. T., and Metaxas, V. L. (2011) Macroeco-nomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking & Finance, 36*(4), 1012-1027. 10.1016/j.jbankfin.2011.10.012

Luo C., Wu D., and Wu D. (2017) A deep learning approach for credit risk using credit default swaps. *Engineering Applications of Artificial Intelligence, 65*, 465-470. 10.1016/j.engappai.2016.12.002

Luo, S., Kong, X., and Nie T. (2016) Spline Based Survival Model for Credit Risk Modelling. *European Journal of Operational Research, 253*(*3*), 869-879. 10.1016/j.ejor.2016.02.050

Magri, S. and Pico, R. (2011) The rise of risk-based pricing of mort-gage interest rates in Italy. *Journal of Banking & Finance, 35*(5), 1277-1290. 10.1016/j.jbankfin.2010.10.008

Maldonado S., Bravo C., López J., and Pérez J. (2017) Integrated framework for profit-based feature selection and SVM classification in credit risk. *Decision Support Systems, 104*, 113-121. 10.1016/j.dss.2017.10.007

Malik M. and Thomas L. C. (2010) Modelling credit risk of portfolio of consumer loans. *Journal of the Operational Research Society. 61*(3), 411-420. 10.1057/jors.2009.123

Marinaki, M., Marinakis, Y., and Zopounidis, C. (2010) Honey Bees Mating Optimization algorithm for financial classification problems. *Applied Soft Computing, 10*(3), 806-812. 10.1016/j.asoc.2009.09.010

Marinakis, Y., Marinaki, M., Doumpos, M., and Zopounidis, C. (2009). Ant colony and particle swarm optimization for financial classification problems. *Expert Systems with Applications, 36*(7), 10604-10611. 10.1016/j.eswa.2009.02.055

Marqués A. I., García V., and Sánchez J. S. (2012). Exploring the behaviour of base classifiers in credit risk ensembles. *Expert Systems with Applications, 39*(11), 10244-10250. 10.1016/j.eswa.2012.02.092

Marqués A. I., García V., and Sánchez J. S. (2012b). Two-level classi-fier ensembles for credit risk assessment. *Expert Systems with Applications, 39*(12), 10916-10922. 10.1016/j.eswa.2012.03.033

Menkhoff, L., Neuberger, D., and Rungruxsirivorn, O. (2012). Collat-eral and its substitutes in emerging markets' lending. *Journal of Banking & Finance, 36*(3), 817-834. 10.1016/j.jbankfin.2011.09.010

Miguéis V. L., Benoit D. F., and Van Den Poel D. (2013). Enhanced decision support in credit risk using Bayesian binary quantile regression. *Journal of the Operational Research Society, 64*(9), 1374-1383. 10.1057/jors.2012.116

Moradi S., and Rafiei F.M. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation, 5*(15). 10.1186/s40854-019-0121-9

Oreski, S. and Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Systems with Applications, 41*(4), 2052-2064. 10.1016/j.eswa.2013.09.004

Oreski, S., Oreski, D., and Oreski, G. (2012). Hybrid system with ge-netic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications 39*(16),12605-12617. 10.1016/j.eswa.2012.05.023

Paleologo, G., Elisseeff, A., and Antonini, G. (2010). Subagging for credit risk models. *European Journal of Operational Research, 201*(2), 490-499. 10.1016/j.ejor.2009.03.008

Peng, Y., Wang, G., Kou, G., and Shi, Y. (2011). An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing, 11*(2), 2906-2915. 10.1016/j.asoc.2010.11.028

Psillaki, M., Tsolas, I. E., and Margaritis, D. (2010). Evaluation of credit risk based on firm performance. *European Journal of Operational Research, 201*(3), 873-881. 10.1016/j.ejor.2009.03.032

Puri, M., Rocholl, J., and Steffen, S. (2011). Global retail lending in the aftermath of the US financial crisis: Distinguishing between supply and demand effects. *Journal of Financial Economics, 100*(3), 556-578. 10.1016/j.jfineco.2010.12.001

Sánchez-Lasheras, F., de Andrés, J., Lorca, P., and de Cos Juez, F. J. (2012). A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications, 39*(8), 7512-7523. 10.1016/j.eswa.2012.01.135

Sousa M.R., Gama J., and Brandão E. (2016). A new dynamic modeling framework for credit risk assessment. *Expert Systems with Applications, 45,* 341-351. 10.1016/j.eswa.2015.09.055

Steiner, M. T. A., Nievola, J. C., Soma, N. Y., Shimizu, T., and Steiner Neto, P. J. (2007). Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. *Pesquisa Operacional, 27*(3), 407-426. 10.1590/S0101-74382007000300002

Sun J., Lang J., Fujita H., and Li H. (2018). Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. *Information Sciences, 425*, 76-91. 10.1016/j.ins.2017.10.017

Tavana M., Abtahi A. R., Caprio D., and Poortarigh M. (2018). An Artificial Neural Network and Bayesian Network model for liquidity risk assessment in banking. *Neurocomputing, 275*, 2525-2554. 10.1016/j.neucom.2017.11.034

Tinoco, M. H. and Wilson, N. (2013). Financial distress and bank-ruptcy prediction among listed companies using account-ing, market and macroeconomic variables. *International Review of Financial Analysis, 30*, 394-419. 10.1016/j.irfa.2013.02.013

Tong, E. N. C., Mues, C., and Thomas, L. C. (2012). Mixture cure models in credit risk: If and when borrowers default. *European Journal of Operational Research*, *218*(1), 132-139. 10.1016/j.ejor.2011.10.007

Tsai, C.-F., Chen, M.-L. (2010). Credit rating by hybrid machine learning techniques. *Applied Soft Computing, 10* (2), 374-380. 10.1016/j.asoc.2009.08.003

Tsai, M.-C., Lin, S.-P., Cheng, C.-C., and Lin, Y.-P. (2009). The consumer loan default predicting model-An application of DE-ADA and neural network. *Expert Systems with Applications, 36*(9), 11682-11690. 10.1016/j.eswa.2009.03.009

Tserng, H. P., Lin, G.-F., Tsai, L. K., and Chen, P.-C. (2011). An en-forced support vector machine model for construction con-tractor default prediction. *Automation in Construction, 20*(8), 1242-1249. 10.1016/j.autcon.2011.05.007

Twala, B. (2010). Multiple classifier application to credit risk assessment. *Expert Systems with Applications, 37*(4), 3326-3336. 10.1016/j.eswa.2009.10.018

Van Gool J., Verbeke W., Sercu P., and Baesens B. (2012). Credit risk for microfinance: Is it worth it? *International Journal of Finance & Economics, 17*(2), 103-123. 10.1002/ijfe.444

Vukovic, S., Delibasic, B., Uzelac, A., and Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit risk. *Expert Systems with Applications, 39*(9), 8389-8395. 10.1016/j.eswa.2012.01.181

Wang G. and Ma J. (2012). A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine. *Expert Systems with Applications, 39*(5) 5325-5331. 10.1016/j.eswa.2011.11.003

Wang, G., Hao, J., Ma, J., and Jiang, H. (2011). A comparative assessment of ensemble learning for credit risk. *Expert Systems with Applications, 38*(1), 223-230. 10.1016/j.eswa.2010.06.048

Wang, G., Ma, J., Huang, L. and Xu, K. (2012). Two credit risk mod-els based on dual strategy ensemble trees. *Knowledge-Based Systems, 26,* 61-68. 10.1016/j.knosys.2011.06.020

Wang, J., Hedar, A.-R., Wang, S., and Ma, J. (2012). Rough set and scatter search metaheuristic based feature selection for credit risk. *Expert Systems with Applications, 39*(6), 6123-6128. 10.1016/j.eswa.2011.11.011

Wu D.D., Olson D.L., and Luo C. (2014). A Decision Support Approach for Accounts Receivable Risk Management. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(12), 1624-1632. 0.1109/TSMC.2014.2318020

Xia Y., Liu C., and Liu N. (2017). Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications, 24*, 30-49. 10.1016/j.elerap.2017.06.004

Xia Y., Liu C., Da B., and Xie F. (2018). A novel heterogeneous en-semble credit risk model based on b-stacking approach. *Expert Systems with Applications, 93*, 182-199. 10.1016/j.eswa.2017.10.022

Xia Y., Liu C., Li Y., and Liu N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit risk. *Expert Systems with Applications, 78*, 225-241. 10.1016/j.eswa.2017.02.017

Xu, X., Zhou, C., and Wang, Z. (2009). Credit risk algorithm based on link analysis ranking with support vector machine. *Expert Systems with Applications, 36*(2), 2625-2632. 10.1016/j.eswa.2008.01.024

Yap, B. W., Ong, S. H., and Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit risk models. *Expert Systems with Applications, 38*(10), 13274-13283. 10.1016/j.eswa.2011.04.147

Yeh, I.-C. and Lien, C.-H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications, 36*(2), 2473-2480. 10.1016/j.eswa.2007.12.020

Yu, L., Wang, S., and Lai, K. K. (2009). An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: The case of credit risk. *European Journal of Operational Research, 195*(3), 942-959. 10.1016/j.ejor.2007.11.025

Zambaldi, F., Aranha, F., Lopes, H., and Politi, R. (2011). Credit granting to small firms: A Brazilian case. *Journal of Business Research, 64*(3), 309-315. 10.1016/j.jbusres.2009.11.018

Zhang Z., Gao G., and Shi Y. (2014). Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. *European Journal of Operational Research, 237*(1), 335-348. 10.1016/j.ejor.2014.01.044

Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y., and Wasinger, R. (2015). Investigation and improvement of multilayer perceptron neural networks for credit risk. *Expert Systems with Applications, 42*(7), 3508-3516. 10.1016/j.eswa.2014.12.006

Zhong, H., Miao, C., Shen, Z., and Feng, Y. (2014). Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. *Neurocomputing, 128*, 285-295. 10.1016/j.neucom.2013.02.054

Zhou L., Lu D., and Fujita H. (2015). The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches. *Knowledge-Based Systems, 85*, 52-61. 10.1016/j.knosys.2015.04.017

Zhou X., Jiang W., Shi Y., Tian Y. (2011). Credit risk evaluation with kernel-based affine subspace nearest points learning method. *Expert Systems with Applications, 38*(4), 4272-4279. 10.1016/j.eswa.2010.09.095

Zhou, L., Lai, K. K., and Yu, L. (2010). Least squares support vector machines ensemble models for credit risk. *Expert Systems with Applications, 37*(1), 127-133. 10.1016/j.eswa.2009.05.024

Zhu X., Li J., Wu D., Wang H., and Liang C. (2013). Balancing accu-racy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach. *Knowledge-Based Systems, 52,* 258-267. 10.1016/j.knosys.2013.08.004

Fernanda Medeiros Assef, Maria Teresinha Arns Steiner

Ten-year evolution on credit risk research: a Systematic Literature Review approach and discussion
Diez años de evolución en la investigación de riesgo de crédito: un enfoque y discusión de revisión sistemática de literatura

Ingeniería e Investigación
vol. 40, no. 2, p. 50 - 71, 2020
Facultad de Ingeniería, Universidad Nacional de Colombia.,
ISSN: 0120-5609
ISSN-E: 2248-8723

DOI: https://doi.org/10.15446/ing.investig.v40n2.78649