FutureJournal

FUTURE STUDIES RESEARCH JOURNAL: TRENDS AND STRATEGIES

FUTURE STUDIES RESEARCH JOURNAL - FIA BUSINESS SCHOOL
Scientifc Editor: Renata Giovinazzo Spers
Evaluation: Double Blind Review, pelo SEER/OJS
Review: Preliminary
Doi: https://doi.org/10.24023/FutureJournal/2175-5825/2020.v12i3.533

Forecasting Financial Distress With Machine Learning – A Review

Recebido: 19/05/2020

Aprovado: 23/07/2020

[1]Denize Lemos Duarte
[2]Flávio Luiz de Moraes Barboza

Abstract

Purpose – Evaluate the various academic researches with multiple views on credit risk and artificial intelligence (AI) and their evolution.

Theoretical framework – The study is divided as follows: Section 1 introduces the article. Section 2 deals with credit risk and its relationship with computational models and techniques. Section 3 presents the methodology. Section 4 addresses a discussion of the results and challenges on the topic. Finally, section 5 presents the conclusions.

Design/methodology/approach – A systematic review of the literature was carried out without defining the time period and using the Web of Science and Scopus database.

Findings – The application of computational technology in the scope of credit risk analysis has drawn attention in a unique way. It was found that the demand for identification and introduction of new variables, classifiers and more assertive methods is constant. The effort to improve the interpretation of data and models is intense.

Research, Practical & Social implications – It contributes to the verification of the theory, providing information in relation to the most used methods and techniques, it brings a wide analysis to deepen the knowledge of the factors and variables on the theme. It categorizes the lines of research and provides a summary of the literature, which serves as a reference, in addition to suggesting future research.

Originality/value – Research in the area of Artificial Intelligence and Machine Learning is recent and requires attention and investigation, thus, this study contributes to the opening of new views in order to deepen the work on this topic.

Keywords: Bankruptcy. Credit Risk. Artificial Intelligence. Machine Learning.

How to cite the article:

Duarte, D., & Barboza, F. (2020). Forecasting Financial Distress With Machine Learning – A Review. *Future Studies Research Journal: Trends and Strategies, 12*(3), 528-574. doi:https://doi.org/10.24023/FutureJournal/2175-5825/2020.v12i3.533

[1] Federal University of Uberlândia – UFU, Minas Gerais, (Brasil). E-mail: denize.duarte@ufu.br Orcid id: https://orcid.org/0000-0002-5074-3385

[2] Federal University of Uberlândia – UFU, Minas Gerais, (Brasil). E-mail: flmbarboza@ufu.br Orcid id: https://orcid.org/0000-0002-3449-5297

UMA ANÁLISE DE PREVISÃO DE DIFICULDADES FINANCEIRAS COM *Machine Learning*

Resumo

Objetivo - Avaliar as diversas pesquisas acadêmicas com múltiplas visões sobre Risco de Crédito e Inteligência Artificial (IA) e sua evolução.

Quadro teórico - O estudo está dividido da seguinte forma: A seção 1 apresenta o artigo. A seção 2 trata do Risco de Crédito e sua relação com modelos e técnicas computacionais. A seção 3 apresenta a metodologia utilizada. A seção 4 aborda uma discussão dos resultados e desafios sobre a temática. Por fim, a seção 5 apresenta as conclusões.

Metodologia - Foi realizada uma revisão sistemática da literatura sem definição do período temporal e utilizando as bases de dados Web of Science e Scopus.

Resultados - A aplicação de tecnologia computacional no âmbito da análise de risco de crédito tem chamado a atenção de forma singular. Constatou-se que a demanda por identificação e introdução de novas variáveis, classificadores e métodos mais assertivos é constante. O esforço para melhorar a interpretação de dados e modelos é intenso.

Contribuições de Pesquisa, **Práticas e Sociais** - Contribui para a verificação da teoria, fornecendo informações em relação aos métodos e técnicas mais utilizadas, traz uma ampla análise para aprofundar o conhecimento dos fatores e variáveis sobre o tema. Categoriza as linhas de pesquisa e fornece um resumo da literatura, que serve de referência, além de sugerir pesquisas futuras.

Originalidade / Relevância - As pesquisas na área de Inteligência Artificial e Aprendizado de Máquina são recentes e requerem atenção e investigação, portanto, este estudo contribui para a abertura de novas perspectivas no sentido de aprofundar os estudos sobre o tema.

Palavras-chave: Falência. Risco de crédito. Inteligência artificial. Aprendizado de Máquina.

Como Citar:

Duarte, D., & Barboza, F. (2020). Uma Análise de Previsão de dificuldades financeiras com *Machine Learning*. *Future Studies Research Journal: Trends and Strategies [FSRJ], 12*(3), 528-574. doi:https://doi.org/10.24023/FutureJournal/2175-5825/2020.v12i3.533

1. Introduction

With the rapid development in information systems and the accelerated spread of economic globalization, how to conduct an effective credit risk assessment is a relevant issue in the field of finance. When the business failure rate becomes high, it can be compromising for investors, shareholders, suppliers, society and in general, for a country's economy (Geng et al., 2015; Zelenkov et al., 2017; Alaka et al., 2018).

The frequent scenario of these global financial crisis and bubbles directly impacts economic growth (Sun et al., 2017). In recent decades, there has been a significant transfer of resources from the productive sectors to the financial sector. Rochon and Rossi (2010) warn of this deregulation, with the growth of complicated financial instruments, the predatory activities of bank loans, the fraudulent practices of institutions, which brought changes to the functioning of economic systems. Thus, this process of capital reallocation in the economy is one of the elements responsible for the growing financial instability (Grilli et al., 2015; Sun et al., 2017).

In view of the weakening of economic growth, followed by the costly effect of the financial distress of organizations, markets are afraid of granting credit and the very fragility of the financial system as a whole (Antunes et al., 2017). The financial distress of the companies increase the risk, when several companies are unable to return the capital to the economy due to their financial fragility, this makes the debts become uncollectible, causing enormous problems for the entire financial system (Sun et al., 2017). This large-scale risk migration, triggered, interrupts and reduces the flow of capital, slowing the economy and causing financial crisis.

After the 2007/2008 global financial crisis, the world market sought to improve market risk and the determinations related to the pricing of financial instruments. According to the Bank for International Settlements (BIS), these guidelines aim to improve the capacity of organizations and institutions to absorb shocks and cope with possible financial crises. In this sense, Basel III emerged, covering measures that strengthen regulation, supervision and risk management.

The creation of the Basel Accord, emphasized the banking sector, but was intended to alert markets worldwide that there is concern about the future of the global financial system and its prospects in relation to its contribution to social and economic development (Barboza et al., 2017). According to the Brazilian Association of Financial and Capital Market Entities (ANBIMA), the Basel III Agreement started in Brazil in October 2013, being the beginning of the transition phase to the new prudential standards and should be completed in full in the year 2022.

In this context, presented by the vulnerability of the financial system, there is a need to quantify and interpret the variables and information related to the phenomenon of financial distress that involves credit risk and its operations. Wang et al. (2012) suggests that an improvement in the accuracy of the predictability of financial distress, even if small, can provide greater returns, minimize the negative effects on system dysfunction, on market segments and even on economic growth.

There are several studies of statistical methods, based on regressions, multivariate discriminant analyzes and based on Artificial Intelligence (AI) available, but there is no agreement on the one that brings the best strategy, greater assertiveness and precision in the face of the insolvency phenomenon (Aziz and Dar, 2006; Abdou and Pointon, 2011).

Ohlson (1980) used logistic regression to estimate the probability pattern for financial distress. Beaver (1966); Altman (1968); Taffler, R. J. (1982); Altman et al. (1994) used Multiple Discriminate Analysis (MDA) through accounting and financial indicators. Altman et al. (1979) used discriminant analysis to classify and predict financial problems in Brazilian companies. They also highlighted other studies of discriminant analysis and logistic regression such as: Martin (1977); Griffin and Lemmon (2002); Hillegeist et al. (2004); De Andrés et al. (2005). Begley et al. (1996) criticized the models presented by Altman (1968) and Ohlson (1980), with the latter showing, in this analysis, better performance.

Some studies used algorithms and artificial intelligence to classify creditworthiness and the predictability of financial distress or bankruptcy: Tam and Kiang (1992); Altman et al. (1994); Shin et al. (2005); Pan (2012); Olson et al. (2012); Sun et al. (2014); Barboza et al. (2017); Sun et al. (2018) and García, Marqués and Sánchez (2019). Addo et al. (2018) argue that algorithms based on artificial neural networks do not necessarily provide the best performance and that it is important to analyze and verify the quality of the data to avoid bias in the classification, but this demand is relevant in any model, be it based on neural networks or others based on traditional statistics. Tang et al. (n.d.) proposes a framework for forecasting financial difficulties in Chinese companies incorporating management and text factors to assess the performance of various forecasting models. The results indicated that these elements can complement the forecasting models, especially the textual ones.

Thus, this study adopted a perspective to evaluate the various academic researches with multiple views on credit risk with a focus on financial distress, credit rating, bankruptcy, machine learning and artificial intelligence (AI) and, its evolution over the years. This objective was achieved by conducting a systematic review to identify the different constructs in this field of literature, exploring and describing the use of methods and techniques of analysis involving AI, and also presenting the trends of publications, data sets, journals, authors and co-authors of this research portfolio. Bibliometric analysis contributes to the understanding of social and natural phenomena, as it is a useful and widely disseminated tool for evaluating scientific productions (Van Leeuwen et al., 2003; Liu et al., 2014).

To develop this work, the Knowledge Development Process-Constructivist methodology (ProKnow-C) was used, which resulted, as an empirical support, in a selection of 165 most relevant national and international scientific publications in relation to their citations and the databases selected for this research. In view of this investigation, it can be said that the application of computational technology in the scope of credit risk management analysis, for institutions and organizations, has been attracting attention in a unique way. It was found that the demand to identify and introduce new variables, classifiers and more assertive methods is constant. The effort to improve the interpretation of data and models is intense and seeks to address the gaps in this field of studies.

This research contributes to verifying the theory, providing information in relation to the most used methods and techniques and also performs a broad analysis, to deepen the knowledge of the factors and variables integrated to credit risk in the face of the phenomenon of financial distress and in the insolvency forecast, thus determining what are the research trends in this field of studies.

This study also contributes to the categorization of lines of research and brings a synthesis of the literature, in which the results serve as a reference, in addition to suggesting future research on the topic. In this context, the findings of this research can contribute to those interested in this topic, helping them to examine trends in this field of studies.

Considering the theoretical discussions and the historical aspect, research on credit risk has not only practical relevance, for the market and the financial system as a whole, since it is connected to operations in which it covers the entire society, but also has its importance theoretical because of its relationship, intrinsic or not, with other studies that are linked to finance theory.

We have organized this research into five sections in addition to this introduction. Section 2 deals with credit risk and its relationship with computational models and techniques. Section 3 presents the methodology used in this research. Section 4 addresses a discussion of the results and challenges on the topic. Finally, section 5 presents the conclusions.

2. Related Literature

At the beginning of capitalist activities, commercial transactions were carried out only with immediate payments, which made it a limitation on purchasing power. The granting of credit was an alternative to enable these transactions, which were previously restricted, in order to expand the economic process, in addition to making it more efficient between the parties. In this context, one of the purposes of the capital market is the transfer of funds between creditors (savers) and debtors (borrowers), which interact efficiently through financial transactions (Copleland and Weston, 1988; Grilli et al., 2015). Scherr (1989) then establishes that credit is the power or the ability to obtain goods or services in exchange for the future payment commitment.

Brealey and Myers (1996) report that, in companies, one third of current assets is composed of the customer account, which confirms the demand and relevance for constant continuous improvement in the modeling and policies of financial loss forecasts. Companies grant credit in order to: increase sales; increase the value of forward transactions; increase, retain or win customers. Thus, granting credit is a way to invest in the customer. Corroborating this statement, Copleland and Weston (1988) report that the capital providers, after exhausting their productive opportunities of greater return, are willing to give up their excess resources, since the loan rate is higher than the one that can be obtained, but the settlement of the amount assigned at the agreed time is essential to guarantee the desired profitability.

In this way, multiple variables interfere in the behavior patterns of payments, including the macroeconomic conditions in which the company is inserted, and which, in an accelerated way can impact on the forecasts of receipts, compromising the potential of the business cash, thus, the following are defined: credit losses arising from financial distress or default (Lopez and Saidenberg, 2000).

In particular, Barboza et al. (2017) report that the default estimate is a phenomenon resulting from human action, and therefore, it is something that has a certain measure of uncertainty. If it is taken into account that there is a possibility of making a mistake in this measure, so that its result leads to some damage or loss to another agent (counterparty), then there is a situation of risk migration, or better, a situation of default and, consequently, financial losses (Lopez and Saidenberg, 2000).

Faced with economic fragility, organizations are in financial distress, compromising the estimation of economic capital and increasing risk in the markets (Sun et al., 2017). Beaver (1966) defines the financial distress of firms as being the inability of a given company to pay its financial obligations to its creditors, whether they are suppliers, financial institutions and shareholders. The author also stresses that the main concern is not to directly address the problem of default, but rather the financial predictors of events that are considered for the probability of the phenomenon that involves credit risk.

Credit risk has been the focus of studies for decades and it is an important topic in the area of financial management (Alaka et al., 2018). Beaver (1966) studied financial indices as predictors of financial distress using a Bayesian perspective, which deals with the assessment of probabilistic hypotheses for reasons of likelihood, where possible events are seen to be dichotomous, in which the company suffers financial distress or not. The author concluded that the relationship between the likelihood ratio and cash flow for total debt was the best predictor. Altman (1968) used Multivariate Discriminant Analysis to classify bankrupt and non- bankrupt companies to predict the phenomenon of bankruptcy and concluded that bankruptcy can be predicted accurately up to two years before the actual failure by combining accounting and financial ratios. Ohlson (1980) applied Logistic Regression models to predict financial distress. Zmijweski (1984) used a probit model to predict bankruptcy.

There are several methodologies used to assist and improve the accuracy of credit risk forecasting. From traditional statistical methods to those models based on AI, which can be univariate or multivariate, which focus on evidence that leads to the failure of the organization and the data are extracted mainly from financial and accounting reports. The models that are theoretical focus on the qualitative causes of the failure, and extract information that can argue the failure using multivariate theory and employing statistical techniques to provide quantitative support for the theoretical argument (Aziz and Dar, 2006).

2.1. Machine Learning at Credit Risk

Machine learning involves the study of pattern recognition theory, the construction of algorithms and computational learning in AI, which can train, learn and make predictions about certain data (Zhu et al., 2017). Odom and Sharda (1990) and Tam and Kiang (1990) were one of the precursors to employing the neural network method to predict bankruptcies and address credit risk. In this article, the authors brought a neural network approach to predicting bankruptcy in financial institutions and companies and made a comparison with other more used models, such as discriminant analysis. The results presented by the authors were that the neural networks proved to be more robust and had better performance than the other models.

In this period, neural networks have already proved to be a competitive tool among traditional statistical methods to assess the financial condition of banks and companies, especially in reducing the type I classification error rate. In 1992, Tam and Kiang (1992) published another work that was even more prominent, where they also used a neural network approach comparing other models. They selected 59 banks in crisis and 59 banks in normal condition in the American market between 1985 and 1987, based on 19 financial indexes and compared four different models, such as Artificial Neural Networks (ANN), MDA, Logit and Decision Trees (DT). According to their experiments, ANN and DT had better predictive performance. The empirical results of these studies were promising for the evaluation of companies' financial conditions in terms of predictive accuracy using ANN.

Tam and Kiang (1992) argue that an ANN represents a nonlinear discriminating function as a pattern of connections between its processing units. Huang et al. (2006) comment that ANN were developed to simulate the neurophysiology of the human brain as a type of flexible, discriminant and clustering nonlinear regression model. The authors state that the ANN architecture can be represented as a three-tier system, one being input, which processes resources, the other concealment layer with appropriate weights using an activation function such as hyperbolic tangent, softmax or logistic function and the last layer pulled from the output. Brockett et al. (1994) consider that a model of neural network can adapt to data, systems or problems and will be modified, this characteristic cannot be found in other static models. Thus, many works based on ANN, Support Vector Machine (SVM) and other algorithms generally fit the data well to deal with prediction of financial distress and bankruptcy, but due to their complexity, they are considered as "black box" technologies (Olson et al., 2012).

In order to improve the accuracy in the prediction of credit risk, it is possible to find some studies that use combinations of techniques and classifiers that integrate multiple methods, including making comparisons between them. This combination of classifiers, techniques and methods brought better performance both in prediction and in reducing Type I and II errors, but in addition to these measurements, it is necessary to evaluate the performance, cost and benefit of these combinations (Wang and Ma, 2012; Marqués et al., 2012*b*; Hajek et al., 2014; Lessmann et al., 2015; Abellán and Castellano, 2017; du Jardin, 2017, 2018).

3. Sample and Methodology

In order to address the objective established for this study, which is to evaluate the various academic researches with the multiple views on credit risk with a focus on financial difficulties, credit rating, bankruptcy and artificial intelligence (AI) and, its evolution over the years . years, a systematic review was carried out to identify the different constructs of this field of literature in articles published in Brazil and worldwide.

Bibliometric revision is defined as a mathematical and statistical application of methods to deal with different topics from different media that requires the study of different perspectives (Pritchard, 1969; Bojovic´ et al., 2014). Cole and Eales (1917) argue that through statistical means, one can trace the branches of a given theme that is attracting more evidence, and that the influence is exercised by the most relevant published works. Liu et al. (2014) states that this methodology demonstrates the dynamics of publications and their evolution over time. Thus, this bibliometric review makes it possible to find gaps in the literature that can be explored in the future.

To carry out this systematic literature review, which according to Polonioli (2020), deals with the search associated with well-defined processes and standards that guarantee transparency, reproducibility and responsibility, minimizing any bias, we follow the orientation of the process for the formation of knowledge and screening of articles by

Knowledge Development Process- Constructivist method (ProKnow-C), proposed by de Azevedo et al. (2014) and Ensslin et al. (2017). This method that provides a process analysis in a structured way to build a consistent theoretical framework that is based on the delimitation of the field of knowledge, the selection of databases, keywords, time filters and criteria for inclusion and exclusion that are misaligned with regard to the central theme of this research, in addition to citations, co-citations, authorship, co-authorship and analysis of journals.

3.1. The Database

The initial phase of scientific research began with the delimitation of the research problem, which will encourage the researcher to search in detail for information and data on a given topic in bibliographic databases (Ensslin et al., 2017). To evaluate, measure and interpret the results, we performed quantitative analyzes on the Bibliometric Portfolio (BP) (Wang et al., 2017).

With the research axis defined, established criteria and filters for delimiting the data set, selecting the types of publications (articles and reviews), and we also identified the databases accessed for consultations in order to cover the largest number possible studies: Web of Science and Scopus (Elsevier). We exclude from the bibliometric portfolio: books, book covers, duplicate works and outside the subject of this research (de Azevedo et al., 2014; Ensslin et al., 2017). No time filter was defined when querying the databases.

This research adapted the methodology of Lozano et al. (2019), who applied two keywords together using network analysis with the grouping of terms. We used the grouping of keywords that led the search for the reference to form the BP: ("bankruptcy" or "business failure" or "credit risk" or "credit scoring" or "expouse at default" or "financial distress" or "loss given default" or "probability of default" or "rating" or "risk of default") and ("machine learning" or "artificial intelligence" or "support vector machine" or "genetic algorithm" or "boosting" or “Bagging” or “data mining” or “ensemble”).

After observing and reading the titles, keywords and summaries of the first articles found in the initial searches, we demonstrate the use of these keywords to define the objective of this work. But, to better explore the theme of this study and increase the data volume of the initial sample, we made the union of these two blocks of keywords that started the composition of the BP.

3.2. Method

Using the two blocks of keywords mentioned above, without using time limitation, the research initially selected 10,346 studies from the first sample, among articles and reviews, 8,717 from the Web of Science database and 1,629 from the Scopus database

(Phases 1 to 4 - Figure 1). Figure 1 presents the summary of the article selection process for the construction of the theoretical framework of this work (Bibliometric Portfolio).

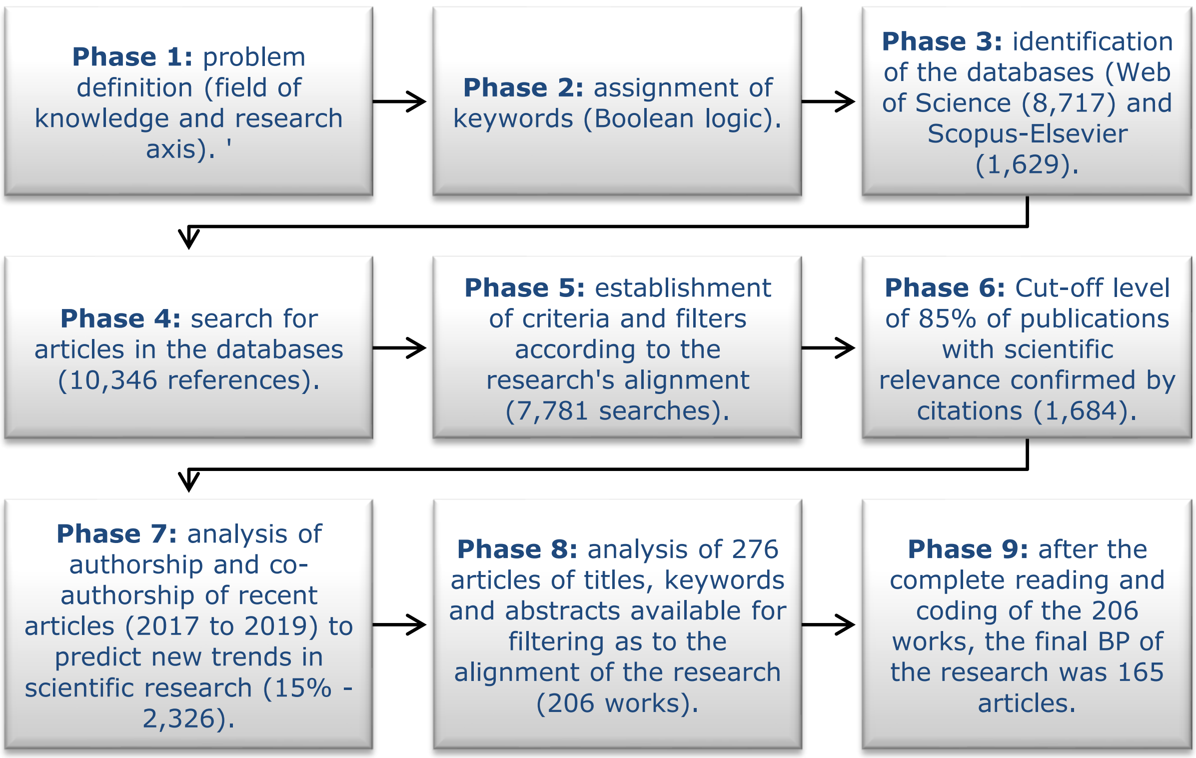

Figure 1: Summary of the ProKnow-C methodology, similar to de Azevedo et al. (2014) and Ensslin et al. (2017)

From this total base of 10,346 studies, we excluded 780 duplicate searches, which resulted in a sample of 9,566 references. After coding and ordering the titles and keywords of each article, we read the records to see how they were aligned with the theme of this study and, after this review, we excluded 1,785 papers for not adhering to the main theme of this article. research, leaving 7,781 articles for further analysis (Phase 5 - Figure 1).

Due to the large volume still existing in the database, we chose to check the scientific relevance of the works and the publication dates. We use the number of citations and co-authorships, which measure the impact of the authors' contributions and identify their scientific collaborations, whose paths can be followed through networks, and according to the evolution and improvements developed, new opportunities become available for investigation (Borgman, 1989). To identify the most relevant articles, that is, those most cited, we cut 85% of the total sum of citations of the 7,781 studies. We established this limit based on Pareto's postulate, which states that most effects originate from a small number of causes (de Azevedo et al., 2014).

Thus, we classified the papers in decreasing order of the number of individual citations, and considering the highest number of citations per document, we selected the articles until the percentage of accumulated citations was equal to the cutoff value (85%).

After this process, the bibliographic set concluded with 1,684 works (Phase 6 - Figure 1).

For the remainder of the less cited base of 15%, we use the authorships and co-authorships of the final most relevant 85% to predict new trends in scientific fields published from 2017 to 2019 (Glänzel, 2001*a*). The initial basis was 2,326 studies. We used this criterion and also the analysis of titles and keywords to align with the theme of this research, in this way, we added 30 studies resulting from this investigation and we integrated them into the BP (Phase 7 - Figure 1). In the 1684 studies, we identified by reading titles and keywords 1,438 articles that are not part of the research objective, which resulted in a base of 246 articles aligned with the theme of this research, which added to the 30 papers, totaled 276 searches (Phase 8 - Figure 1).

In the total set (276 studies), we assessed whether the summary was in line with the main objective of this research. This analysis resulted in the exclusion of 70 articles. The BP base then resulted in 206 papers aligned with the theme of this research (Phase 8 - Figure 1).

Then searched for the papers and 206 were available, we stored and updated the volume of citations per article with search on the Google Scholar website. Proceeded to complete the reading and coding of these works that resulted in a final basis of the BP 165 articles. Excluded 41 articles with an individual database, which are not linked to the central theme of this work. A table containing some information regarding the articles selected by the ProKnow-C method we present in Appendix A at the end of this research (Phase 9 - Figure 1).

4. Discussions and Results

In the survey of articles that conducted this study according to the ProKnow-C methodology, we found 165 works related to the application of AI, ML and Artificial Neural Networks (ANN) methods to address financial distress, credit risk, bankruptcy and default. This collection included a total of 302 different researchers. Among them, 55 authors published more than one work in the bibliometric portfolio related to credit risk management, bankruptcy forecast and financial distress, using machine learning models and neural networks.

Table 1 presents the 15 authors who published more than 4 articles. In this table, citations are independent of authorship or co-authorship, that is, they are additional per author and article. The author Chih-Fong Tsai is the most active, with 13 published works (citations 1,259), the most cited are those from 2008 and 2009, and his publications as the first author, were until 2014, in the most recent articles he is as co-author. Generally his works use the same data sources (Germany, Japan and Australia), and he does not inform the base time horizon, in his articles comparisons are made between individual and set techniques, hybrid or not, and most use the techniques MLP and SVM. Hui Li and Jie Sun are in second place with 10 works each (747 citations each), they publish together alternating in the works as first authors.

Table 1: Most active authors evidenced on the subject of this research by number of publications.

Author	Publications	Citations
Tsai, CF	13	1259
Li, H	10	747
Sun, J	10	747
Garcia, V	7	319
Marques, AI	7	319
Sanchez, JS	7	319
Ribeiro, B	7	229
Wang, G	6	570
Chen, N	6	215
Han, I	4	1031
Kim, MJ	4	429
Ma, J	4	319
Liang, D	4	198
du Jardin, P	4	101
Jones, S	4	101

They usually use the same Chinese data set, and the same base time horizon, try to use hybrid models and work with different techniques. Thirdly, with 7 studies published on the subject of this research, Vicente García, Ana I. Marqués, J. Salvador Sanchéz (319 citations each), usually the authors García Marqués and Vicente García alternate as first authors and Sanchéz is in the co-authorship in all works and despite being from different universities, they always publish works together. When the number of publications increases to 6, the total number of searches drops to 9 most active authors.

Figure 2 presentes the 15 most dominant authors actively and their contacts. To assess the relevance of these studies, use the number of citations per study. The number of citations is an indicator adopted as an impact factor for academic research. The highest citations suggest their level of coverage and influence in the determined field of research (Tang, 2013). Du Jardin, Jones, Liang and Bernardete, are the authors who have published the most in the last 5 years. The authors Hui Li and Jie Sun, in this BP, all their works were published together. Bernardete Ribeiro and Armando S. Vieira are always publishing as authors and co-authors.

Glänzel (2001*b*) brings a study on national characteristics in international relations of scientific co-authorship, identifying statistical evidence of symmetry and asymmetry in the links of co-publication, of the relationship between international co-authorship and the profiles of national research and the impact of the quote. The author points out in this research that co-publication maps reveal structural changes in international co-authorship links by identifying stable links and coherent clusters and new nodes and links. The author states that scientific collaboration between EU member countries promotes European integration in one of the most advanced systems in the world of science and technology and that, on the other hand, co-publications can result simply as mandatory exercises under bilateral agreements between institutions, scientific administrations or governments.

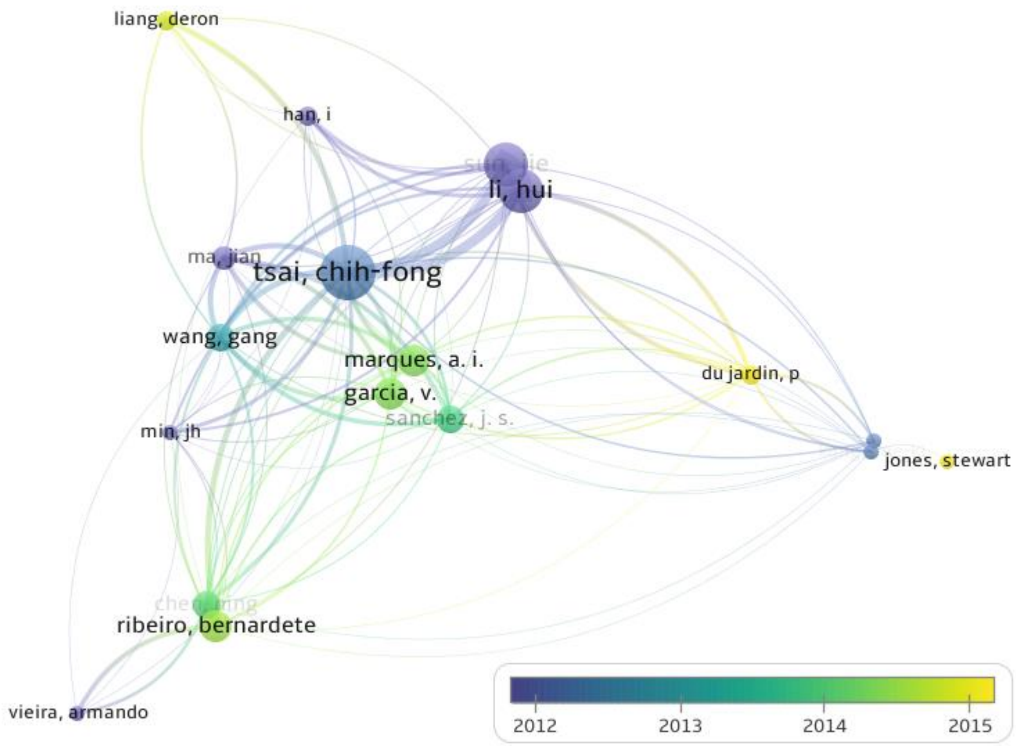

Figure 2: Map of the most active authors and their links by publication aging. The map was developed using VOSviewer software.

The author Chih Fong Tsai from Taiwan is the most cited among the most active authors, as he was a pioneer in studying the topic. His articles deal with hybrid models, which use combinations of classifying set techniques to forecast FD, bankruptcy and to predict credit risk. Hui Li, from China, specializes in ML, ANN and data mining, develops hybrid models with case based reasoning and which also use combinations of classifying set techniques to forecast financial distress and credit risk. In one of the articles selected in the BP, the author proposes an Elimination And Choice Net Translating Reality-ELECTRE approach (Li and Sun, 2009).

Table 2 presents the authors and co-authors who were most cited in relation to the theme of this research in order of citations and publications respectively, which are: Varetto, F (1,617), was the most cited, participated as a co-author in a article published in the Journal of Banking & Finance, in which Edward Altman (one of the most relevant authors in terms of credit risk) was the first author. Kiang, MY (1,444) and Tam, KY (1,444) were the precursors to using learning models to predict bankruptcy. Altman, E (1,393) is the seminal author who deals with prediction with the MDA model and reference on the topic. Marco, G (1,336) is also most quoted where he coauthored an article published in the Journal of Banking & Finance, where Altman was the first author.

Table 2: Authors most evidenced on the subject of this research by number of citations.

Author	Publications	Citations
Varetto, F	2	1.617
Kiang, MY	1	1.444
Tam, KY	1	1.444
Altman, E	2	1.393
Marco, G	1	1.336
Tsai, CF	13	1.259
Shin, KS	2	1.110
Lee, YC	3	1.052
Han, I	4	1.031
Min, JH	3	958
Chen, HC	1	949
Chen, WH	1	949
Hsu, CJ	1	949
Huang, Z	1	949
Wu, SS	1	949
Sharda, R	1	821
Wilson, RL	1	821
Pan, WT	1	815

This is a brief review that discusses the main authors who addressed the central theme of this research and we highlight China, Taiwam, USA, Spain and South Korea as the choices that lead the studies related to credit risk management, forecasting bankruptcy and FD, using machine learning models, ANN and AI. Considering the analysis so far, the authors who develop models in this area seek to improve the techniques, making crossings and clusters of different databases and models, to improve the accuracy of the models' prediction.

4.1. Articles with scientific relevance

According to the analysis of articles and reviews, we expose the fundamental characteristics of each study. These attributes include the periodicity of publications, the journals in which the works were published, the authorship and co-authorship of the documents, the characteristics of the databases used, the methods, techniques and classifiers and the countries that served as the basis for this research. The documents examined in the BP were published between 1992 and 2019.

Figure 3 presents the quantitative evolution of articles published annually. We noticed that there was a considerable increase in articles published after 2008, which can be explained by the financial crisis that occurred in mid 2007/2008, also by the improvement in infrastructure and the evolution of computational resources, which are capable of storing and transacting a vast amount of data, as well as the improvement of AI methods (Moro et al., 2015). Cole and Eales (1917) describe the question of how much research is concentrated in a specific period and in a specific region, however indifferent it may be, this is a sure indication of the fulfillment of contemporary interests and activities and that we can trace the branches of that theme that was attracting more attention. This reinforces interest in the topic, which explores the machine learning mechanisms to predict financial distress in organizations.

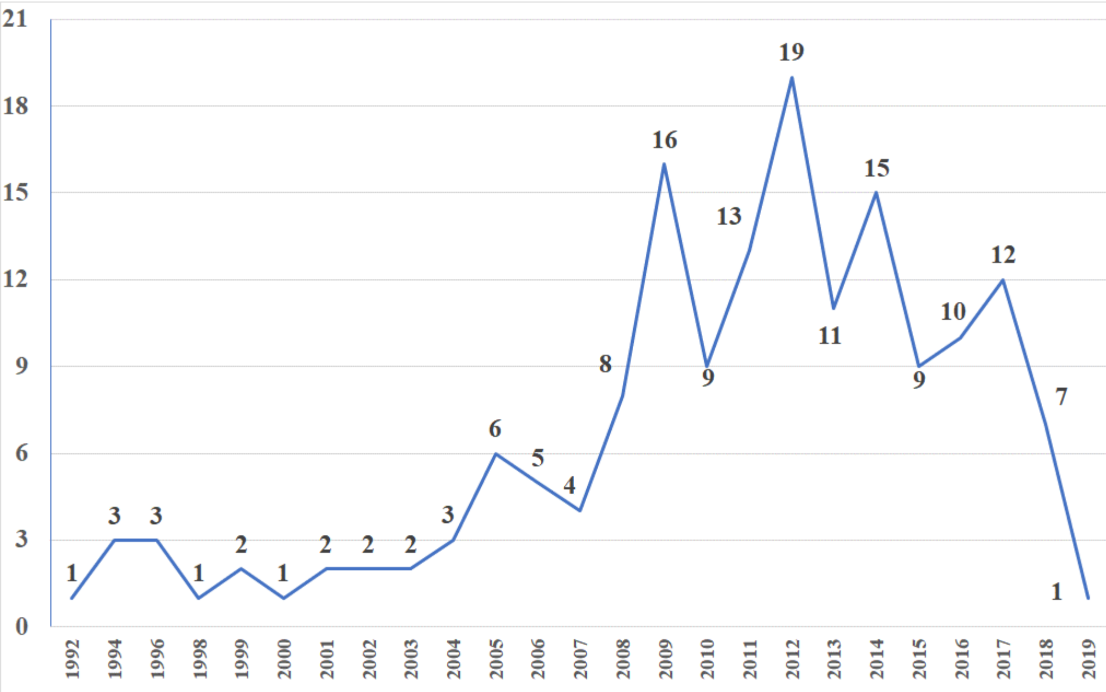

Figure 3: Annual evolution of published studies.

For us to measure the relevance of BP journals, we consider the classification indicators Journal Citation Reports (JCR), for direct comparison with journals in the area of knowledge so, we prioritize the classification by this index, and the Scientific Journal Rankings (SJR), to compare with other journals referring to other areas of knowledge.

Table 3: Featured journals on the subject of this research.

Journals	JCR	SJR
Information Fusion	10.716	2,238
Ieee Transactions on Fuzzy Systems	8.759	2,794
Ieee Transactions on Systems Man Cybernetics-Systems	7.351	2,147
Tourism Management	6.012	2,924
Neural Networks	5.785	1,970
Information Sciences	5.524	1,620
Omega-International Journal of Management Science	5.341	0,426
Knowledge-Based Systems	5.101	1,460
Artificial Intelligence Review	5.095	1,055
Applied Soft Computing	4.873	1,216
Neural Computing & Applications	4.664	0,637
Technological and Economic Development of Economy	4.344	0,774
Expert Systems With Applications	4.292	1,190
Management Science	4.219	6,080
Information & Management	4.120	0,683

Table 3 presents the position of the 15 most relevant journals according to their impact factor within the analysis of this BP. The criterion that we used to map the highlighted journals on the subject of this research was the number of citations and publications.

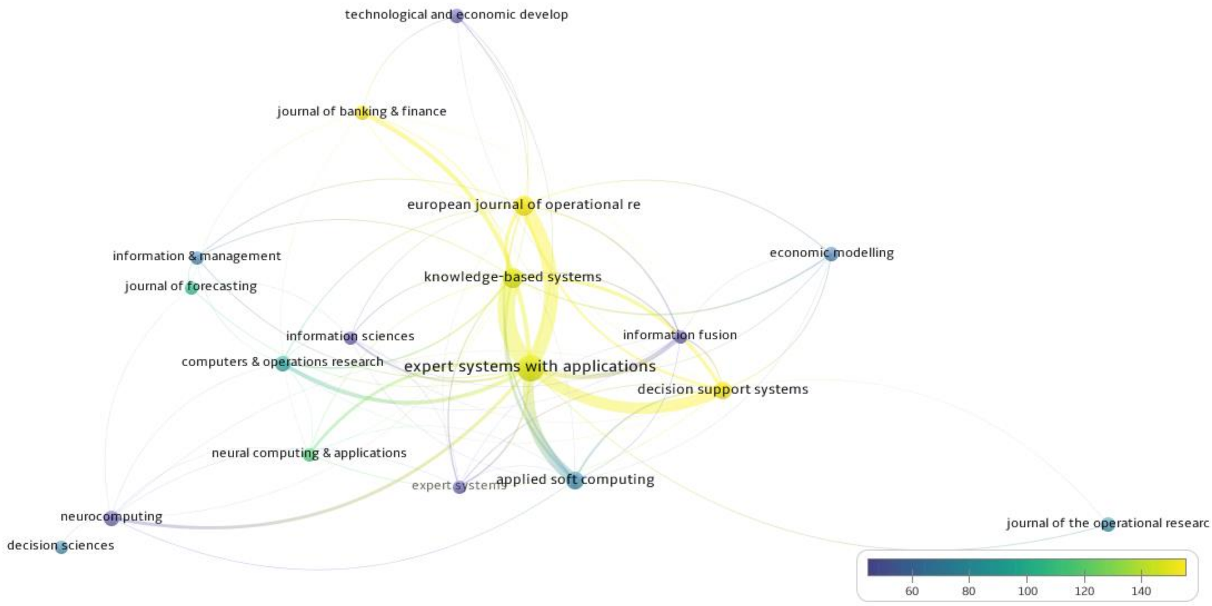

Figure 4: Map with highlighted journals on the subject of this research made in the VOSviewer software.

Furthermore, Figure 4 presents this mapping of the most cited journals and the relationship between them in the studies that make up this BP. Observed that BP's works were published in 44 different journals, with Expert Systems With Applications, Knowledge-Based Systems, European Journal of Operational Research and Decision Support Systems being the most prominent in relation to citations and also publications. Thus, we can infer that most of the publications that make up the BP of this research, are aimed at empirical studies of construction and comparability of models for forecasting and managing credit risk.

Table 4 presents numerically the distribution of BP among journals. In relation to the average number of citations per publication, Management Science has 159% more citations than the second in this position, which is the Journal of Banking & Finance, which is also 54% above the next.

Table 4: Journals with more publications and citations from BP from 1992 to 2019.

Journal	Publications	Citations	JCR	SJR
Expert Systems With Applications	57	8.193	4.292	1,190
Decision Support Systems	9	2.863	3.847	1,536
Knowledge-Based Systems	15	2.155	5.101	1,460
European Journal of Operational Research	14	2.110	3.806	2,205
Journal of Banking & Finance	3	1.671	2.205	1,599
Management Science	1	1.444	4.219	6,080
Applied Soft Computing	8	576	4.873	1,216
Ieee Transactions on Fuzzy Systems	1	361	8.759	2,794
Computers & Operations Research	4	360	3.002	1,859
Journal of Management Information Systems	1	324	3.013	2,388
Others	52	3.300		
Total	165	23.357		

Also analyzed the countries of origin of the research, which included 34 different countries, referring to the institutions in which the authors are associated, determining the origin of these works. Figure 5 presents the mapping by country of origin of the researches of this BP, highlighted by the magnitude of citations, we have the authors of the USA (5,880), Taiwan (4,421), China (2,730), South Korea (2,388) and Italy (1,801), which represents 71% of the total searches in this bibliometric portfolio.

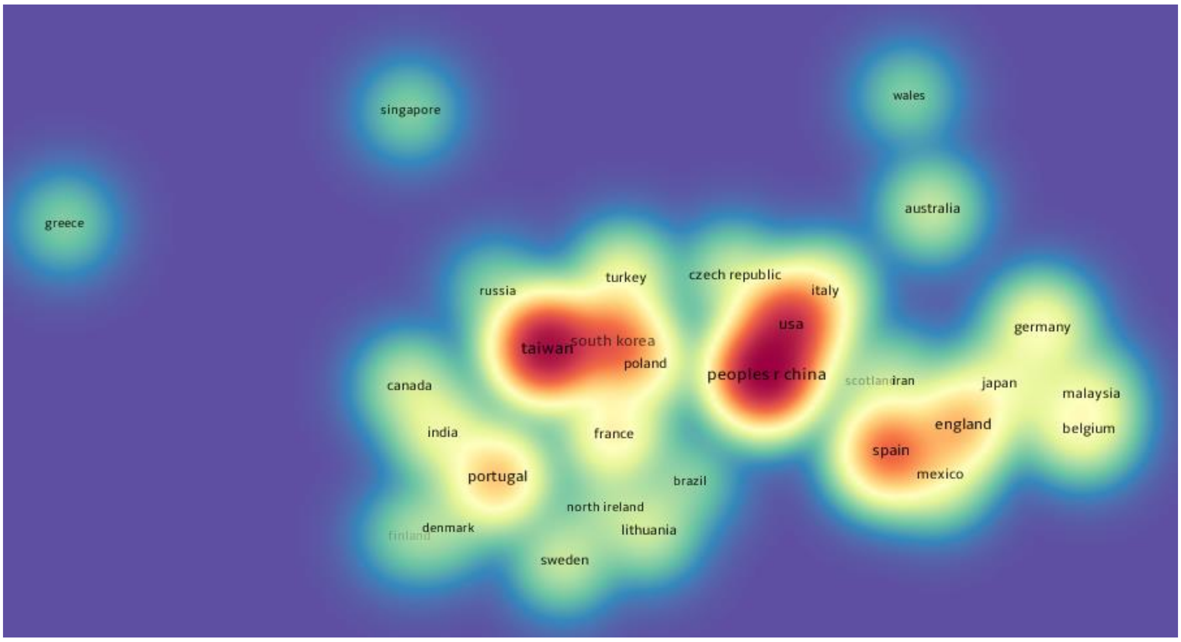

Figure 5: Map of publications by country of origin. The map was developed using VOSviewer software.

Figure 6 presents the map with the most densely populated regions, which presents where the largest research volume is concentrated, which are China (33), Taiwan (30), USA (20), Spain (14) and South Korea (13), which represents 67% of BP. According to Tollefson (2018) the Chinese government has been gradually expanding investments in science, development and technological innovation, which may have contributed to the large volume of scientific research. On the contrary, the regions with the lowest density on the map are those with the least number of searches in relation to the central theme of this BP, which are: Wales, Switzerland, South Africa, Singapore, Scotland, Russia, Northern Ireland, Iran, Greece, Finland, Denmark and Brazil, where each produced only one article with scientific relevance. Studies that include researchers from these regions and others, where there was no article related to the central theme of this work, such as countries in Latin America, Israel, Hungary, among others, that could bring enrichment to the research collection with different results and multiple views.

Figure 6 classifies the keywords of the articles that are part of this BP. It can be seen that the words that stood out the most were: neural network (97), bankruptcy prediction (91), support vector machine (70), classification (60), discriminant analysis (45), credit scoring (33), datamining (31), genetic algorithm (29), financial ratios (27), models (26), financial distress (23), feature selection (22), artificial intelligence (20), model (20), prediction (20), confirming the accuracy of the BP delimitation in relation to the central theme of this research and the adherence to the tool used (ProKnow-C). Also identified that at the extreme levels of the mapping there are the ensemble, boosting and bagging techniques that are being used more recently, because according to some studies, these methods improve performance compared to individual models (Wang et al., 2011; Liang et al., 2018).

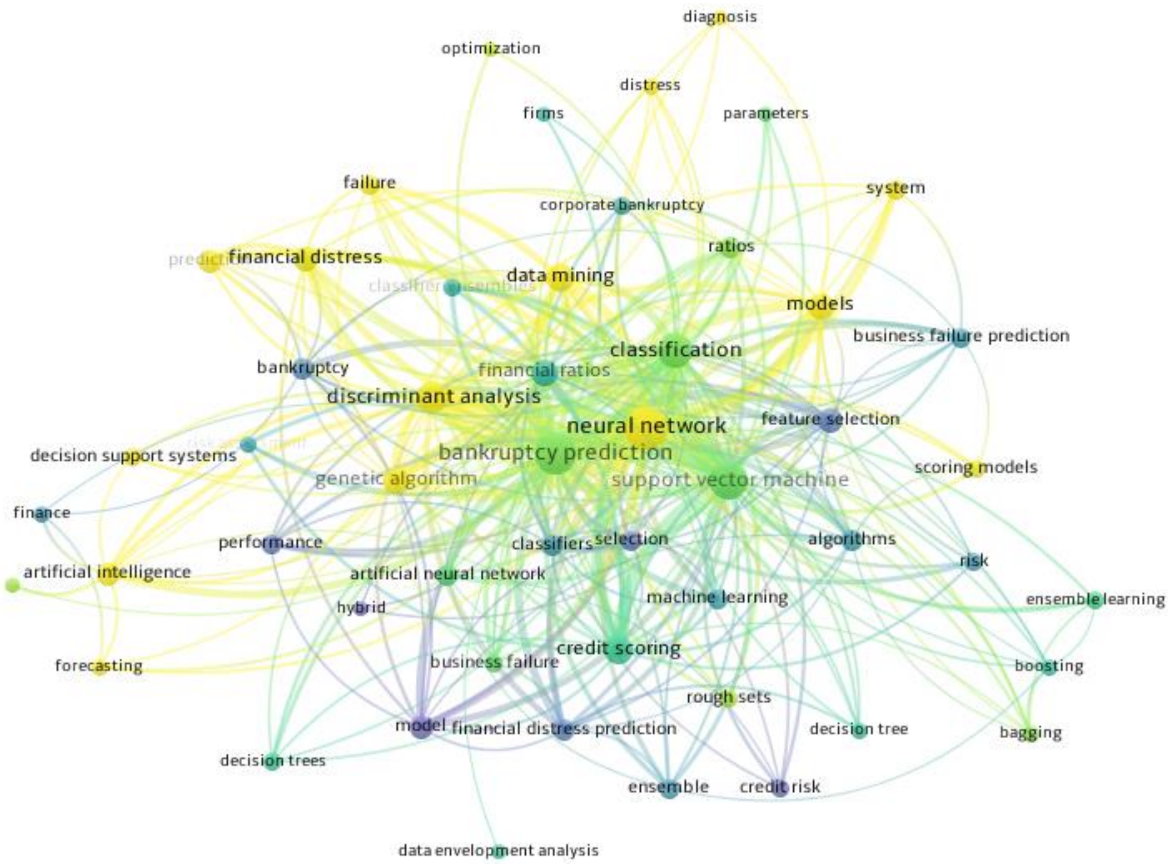

Figure 6: Map of keywords by citation. The map was developed using VOSviewer software.

4.2. Quantitative analysis of BP articles

At the end of the first stage of analysis of the main works related to the central theme of this research, we classified and coded the BP according to the categories described in 5, each of these classification categories was numbered from 1 to 8, being: 1 - Context, 2 - Type of research, 3 - Methods and techniques, 4 - Focus, 5 - Base country of origin, 6 - Dependent variable, 7 - Independent variables, 8 - Time horizon (database). Used the letters of the alphabet to codify the subcategories to further deepen the analysis. The subcategories of category 3 - Methods and techniques, where methods, techniques and classifiers are grouped, will be coded by abbreviations to facilitate data logistics.

Table 5: Categorization of the Bibliometric Portfolio for analysis.

Cat.	Meaning	Codes for alternatives	Meaning
1	Context	A	Default Corporate
		B	Financial distress
		C	Bankruptcy Forecasting
		D	Credit score
		E	Others
2	Research type	A	Empirical
		B	Review
3	Method	SVM	Support Vector Machine
		LR	Logistic Regression
		ANN	Artificial Neural Networks
		BA	Bagging
		LDA	Linear Discriminant Analysis
		MLP	MultiLayer Perceptron
		MDA	Multiple Discriminant Analysis
		DT	Decision Tree
		BOO	Boosting
		AB	AdaBoost
		NB	Naive Bayes
		CART	Classification and Regression Trees
		KNN	K-Nearest Neighbor
		RF	Random Forest
		BPNN	Back Propagation Neural Network
		GA	Genetic Algorithm
		RS	Random Subspace
		RBF	Radial Basis Function
		CBR	Case Based Reasoning
		RST	Rough Set Theory
		SOM	Self Organizing Map
		RTF	Rotation Forest
		ELM	Extreme Learning Machine
		ML	Machine Learning
		SMO	Sequential Minimal Optimization
		O	Others
4	Focus	A	Financial Institution
		B	Non-Financial Institutions
		C	Others
5	Country of origin of data	A	USA
		B	Europe
		C	Asia
		D	Latin America
		E	Australia
		F	Other / Not Mentioned
6	Dependent variables	A	Bankruptcy

		B	Financial distress
		C	Default
		D	LGD
		E	Others
7	Independent variables	A	Financial Ratios
		B	Non-Financial Indices
		C	Personal information
		D	Others
8	Time horizon (database)		Start
			End

We classify and group the variables of interest that deal with the context of the work, Figure 7, and we realize that most studies (83-45%) seek to identify the behavior of companies that went bankrupt in relation to companies that did not go bankrupt. The models that create scores and classify the risk were 55 studies (28%) and 34 studies (18%) sought to predict the financial distress of organizations.

Context	P	%	Focus	P	%	Research type	P	%
Bankruptcy Forecasting	83	45%	Financial Institution	19	13%	Empirical	154	93%
Credit score	55	30%	Non-Financial Institutions	133	81%	Review	11	7%
Financial distress	34	18%	Others	12	8%			
Default Corporate	12	6%						
Others	1	1%						
Dependent variables	**P**	**%**	**Independent variables**	**P**	**%**			
Financial Ratios	146	77%	Bankruptcy	78	47%			
Non-Financial Indices	24	13%	Financial distress	29	18%			
Personal information	5	3%	Default	47	28%			
Others	15	8%	LGD	2	1%			
			Others	9	5%			

Figure 7: Classification of the research by context, focus, type and dependent and independent variables.

Within the organization's own life cycle, it goes through phases of financial distress, when it is not fully stabilized to support such events, the outcome may result in bankruptcy or judicial recovery, therefore, the accuracy in predicting these occurrences for policyholders. decisions and other stakeholders is relevant and contributes so that the situation does not become even more serious, increasing the number of failures and becoming a contagious and devastating process for a country's economy (Altman et al., 1979; Geng et al., 2015; Zelenkov et al., 2017; Alaka et al., 2018).

Regarding the type of research as shown in Figure 7, most of the works are empirical studies (152), about 93.3% and the rest, 11 articles (6.7%), deal with literature reviews. BP's work develops models that use modern technology to simplify processes and increase efficiency in precision and prediction, seeking to proactively manage credit risk, financial distress and corporate bankruptcies (García, Marqués and Sánchez, 2019).

An important situation to stand out among the computational models would be the inter-face with the qualitative analysis or based on qualitative non-financial predictive factors of unstructured texts, such as audit reports and annual reports, to better explain and facilitate the implementation and testing of the models constructed (Hajek et al., 2014; Tang et al., n.d.). Qualitative financial analysis could meet regional perspectives and regulatory conditions, the studies by Ribeiro et al. (2019) bring an algorithm-based approach to graph construction, using binary relationships of qualitative credit risk data, and the results demonstrate that the performance of predictability accuracy needs improvement.

Figure 7 presents the articles of this BP, which focused on working, mostly (81%), using data from non-financial companies, 13% used data from financial or insurance companies and the rest, 8%, it was not possible to identify. Each segment has a specific regulatory system, they are different market segments, so specification is necessary, and to further filter this BP, we opted to exclude studies with data sets of individuals (example of a disregarded article: Guo et al. (2016)). Among the articles that did not have this information are Hu (2009) and Lahsasna et al. (2010). To evaluate the performance of the tested models, most researchers used precision (ACC), the area under the ROC curve (AUC), type I and type II errors and the Brier Score, mean square error of the forecast (Ala'raj and Abbod, 2016*a*). Karan et al. (2013) use the following performance evaluation metrics: ACC, Error Type I, Error Type II, AUC, Gini, Kolmogorov-Smirnov (KS) and Hand's H. These measures have the predictive performance of the proposed experiment and serves as a comparison between the different models (Ala'raj and Abbod, 2016*a*; Wang et al., 2018).

The studies use multiple quantitative and qualitative variables to develop the numerous models, in order to monitor the ongoing occurrences of organizations and detect sudden changes in behavior in factors that will compose the risk. Based on the BP data sets of this research, the variables of interest were classified and grouped, and we noticed that most studies (47%) seek to identify the behavior of companies that went bankrupt in relation to companies that did not go bankrupt. Models that create scores and classify credit risk were 28% and 18% sought to identify the financial distress of organizations in advance.

When classifying explanatory variables, most studies (77%) use financial and accounting indicators, due to the ease of calculating and working with the data. Qualitative variables, in addition to having a certain subjectivity, still have difficulties in collecting data, due to the intense manual work, mainly Corporate Governance data, in some countries. Thus, these variables were identified in 13% of the studies. Liang et al. (2016) and Jones (2017) used Corporate Governance data in their models such as: Board Structure, Ownership Structure, Audit Reputation, CEO Compensation, among others.

In order to accurately predict the financial distress and bankruptcy of organizations, several methods have been developed and tested. They can be divided into the following categories: traditional statistical methods and methods based on AI. Traditional statistical methods work well and are simpler and more understandable. Models based on artificial neural networks, support vector machines and other algorithms, generally present superior performance and fit satisfactorily in the database for predictability, but due to the complexity of the methods and the lack of understanding of analysts and managers, the models are considered black box technologies (Olson et al., 2012).

Table 6 presents the methods and techniques most used in BP, where Logistic Regression (10.1%) was the most used, followed by Support Vector Machine (10.0%) and Artificial Neural Network (9.7%). Although there are multiple methods and techniques, joint, individual and hybrid, to predict financial distress and bankruptcy, they behave differently in various situations, according to the database and selected variables, in this BP, 104 methods and techniques were used. Thus, it is not possible to specify a single appropriate method that can adapt and work better than any other (Wang et al., 2018).

Regarding the data set used by the studies, when it comes to forecasting financial distress and bankruptcies, it is characterized by class imbalance, where the distribution of instances of the data classes is distorted, as the number of companies in poor conditions is much smaller than companies under normal conditions (Wang et al., 2018). An algorithm used to solve this problem is the SMOTE (Synthetic Minority Over-Sampling Techinique), which helps to maintain the balance of distribution between classes, contributing to predictive results in face of unbalanced data sets (Zhou, 2013; Hajek et al., 2014; Sun et al., 2018; Li and Perez-Saiz, 2018; Wang et al., 2018).

Table 6: Methods and techniques most used in BP.

Abreviatura	Method	Frequency	% Proportion
LR	Logistic Regression	76	10,1%
SVM	Support Vector Machine	75	10,0%
ANN	Artificial Neural Networks	73	9,7%
DT	Decision Tree	50	6,7%
BA	Bagging	29	3,9%
LDA	Linear Discriminant Analysis	28	3,7%
MLP	MultiLayer Perceptron	27	3,6%
MDA	Multiple Discriminant Analysis	27	3,6%
BOO	Boosting	25	3,3%
AB	AdaBoost	20	2,7%
NB	Naive Bayes	19	2,5%
CART	Classification and Regression Trees	18	2,4%
KNN	K-Nearest Neighbor	18	2,4%
RF	Random Forest	17	2,3%
BPNN	Back Propagation Neural Network	15	2,0%
GA	Genetic Algorithm	15	2,0%
RS	Random Subspace	14	1,9%
RBF	Radial Basis Function	12	1,6%
CBR	Case Based Reasoning	10	1,3%
RST	Rough Set Theory	9	1,2%
SOM	Self Organizing Map	9	1,2%
RTF	Rotation Forest	7	0,9%
ELM	Extreme Learning Machine	6	0,8%
O	Others	150	20,1%

From this literature review, we observed that in relation to the origin of the databases, that is, the countries used for the experiments, that most studies used data sets from Germany (12.5%), Australia (11.1%), USA (10%), Japan (8.9%), Taiwan (8.1%), China (7.4%), Korea (6.6%) and France (4.8%). The use of these databases to deal with credit risk management, allows other researchers to compare new models and also their results. But when these databases are overused, they may not present the specific conditions of other countries, markets, or segments, due to regulatory issues and specific socioeconomic conditions, so the experiments can lead to distorted conclusions and may not meet the demands of several markets, because the factors that affect bankruptcy vary between different countries (Balcaen and Ooghe, 2006; García et al., 2014; Liang et al., 2018; Tian and Yu, 2017). Thus, it is necessary to add research from other countries, such as Brazil, Latin American countries, Israel, which, in this BP database, did not present any research. Figure 8 presents the countries most used as data sets for experiment and analysis.

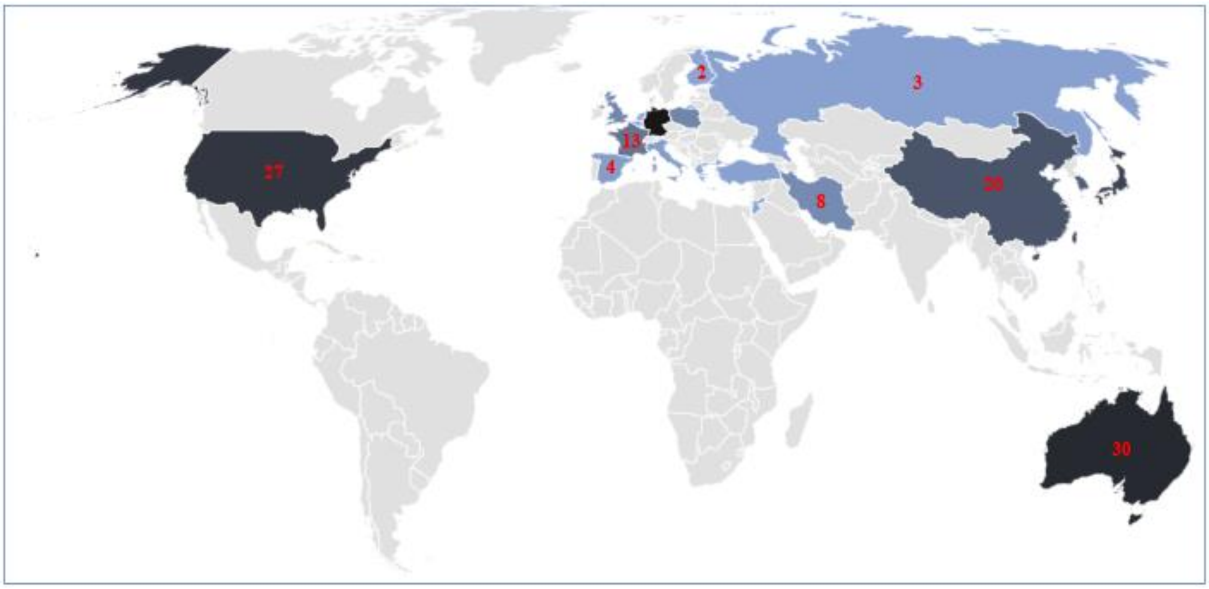

Figure 8: Mapping of countries by concentration of frequency of use of the database as a whole.

In the evaluation of the time horizon of use of the data sets that compose the BP, Figure 9, we verified that there was a concentration of data from the period 1999 to 2007, the most recent bases used were only until the year 2015 and only two searches have reached this period. The demand to update the models, with larger and more updated data sets, for a longer period of time could have broader implications in terms of results. Altman (2002) already demonstrated this concern where, for the author, the main technical motivational factors included the refinement of traditional techniques, innovations in analytical solutions, larger and improved databases, to translate credit risk management.

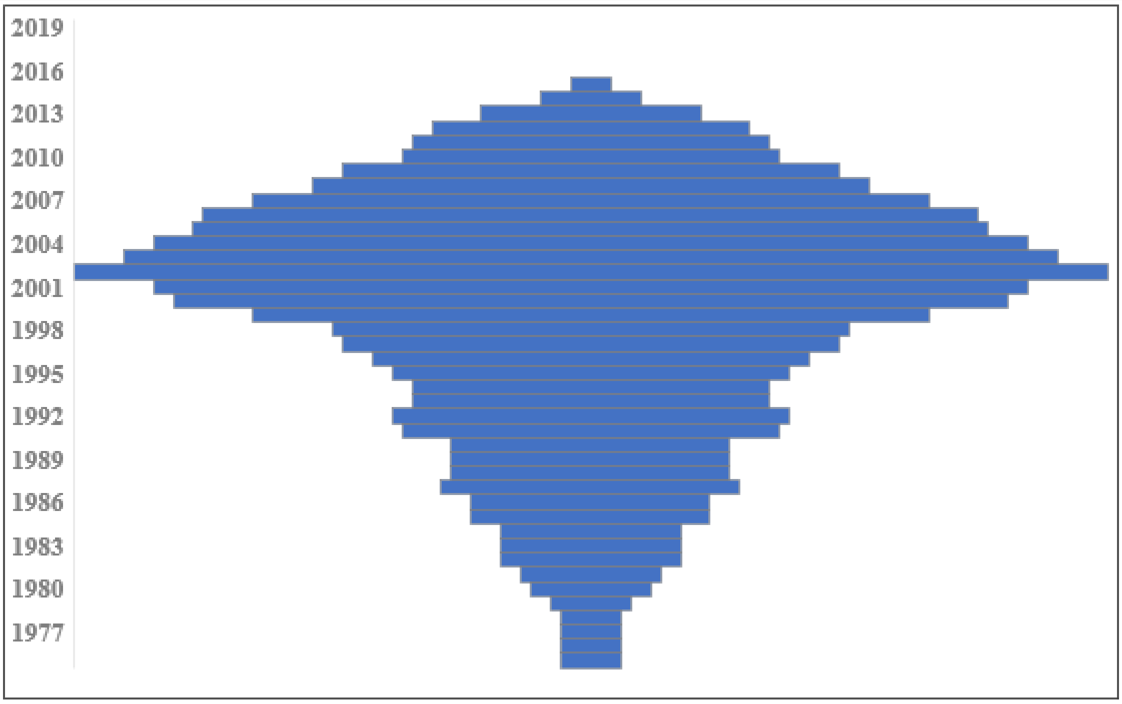

Figure 9: Time horizon (database) used in the set of databases.

A situation that may contribute to the use of the same data set is the difficulty in obtaining and updating the information of the firms for credit risk analysis, thus, several authors use the same database to analyze different methods (Li et al., 2009; Li and Sun, 2009, 2010, 2011; Ribeiro et al., 2012, 2019; Wang et al., 2011; Wang and Ma, 2011, 2012). In our portfolio 72% of the analyzed articles used the data set with an average of 7 years difference between the last year of the analyzed base and the date of publication of the article and 28%, the average of this break is 2.4 years . Thus, it can be seen that the methods are tested with older bases, it would be interesting to evaluate the behavior of the methods between the agings of the data sets, as there may be exogenous variables that influence the behavior of the models and may not have been considered.

Some researches do not detail the full description of the variables used in the experiments or did not provide the size of the database used or the period of the data set, which makes it even more difficult to understand the models and methods developed (Yu et al., 2011; Abellán and Mantas, 2014).

In this BP, methods based on machine learning and neural networks, such as SVM, DT, ANN and RF, sets, individual or hybrids, performed better in terms of accuracy than traditional statistical methods. The LR method stood out among the traditional statisticians, so in the third stage of this work, it compared the performance of the RF and LR methods, individual, in the Brazilian capital market and in companies in the region of interest.

5. Concluding Remarks

With the development of information systems and the spread of economic globalization, the way to conduct credit risk is a matter of relevance in the field of accounting and finance. The effect caused by the business failure rate can be compromising for investors, shareholders, suppliers, society and in general, for a country's economy. An improvement in the accuracy of the predictability of financial distress can provide greater returns and also minimize its negative effects for all involved.

This research was developed in order to evaluate the diverse academic researches with the multiple views on credit risk with a focus on financial distress, credit rating, bankruptcy, for which they used recent methodologies and with different perspectives from traditional statistics, such as learning machine and artificial intelligence and, its evolution over the years. This objective was achieved by conducting a systematic review to identify the different constructs in this field of literature, exploring and describing the use of methods and analysis techniques that involve AI, and also presenting the trends of publications, data sets, journals, authors and co-authors of this research portfolio. We used the ProKnow-C methodology, which resulted, as an empirical support, in the selection of 165 scientific references. In this selection there was no time limitation.

In view of the final investigation, we can observe that the application of computer technology in the scope of credit risk management, has been attracting attention in a unique way, mainly after the financial crisis that occurred in mid 2007/2008. We note that there was a considerable increase in articles published after this period, which can be explained by the crisis that occurred, another issue that can also be raised is the improvement in infrastructure and evolution of computational resources, which are capable of storing and transacting a vast amount data, as well as the improvement of AI-based methods (Moro et al., 2015).

However, despite all the benefits brought by the evolution of models and techniques based on AI and ANN, in relation to the performance in the prediction precision, some authors discuss their disadvantages and limitations. Issues to be improved are mentioned, such as the architecture used, where decisions are based on trial and error, which makes the process too slow due to the criteria and volume of data. Feedforward ANNs need to be restructured when the smallest variations in data occur, another issue to be discussed is the existence of great difficulty on the part of users in interpreting the data output in the prediction models. Another situation would be in relation to databases, where some authors claim that the models are not able to deal with the imbalance of the databases, and may even compromise their performance. These limitations make it difficult to apply these models in practice, thus making them a black box technology (Zmijweski, 1984; Wang et al., 2011; Olson et al., 2012; Zhang et al., 2013; Chuang, 2013). It would be interesting to assess the level of dependency between the variables, explaining the degree of influence of each one, within the various models and techniques, that is, to implement the perspective of fundamental analysis to computational models. The most cited articles are that of Tam and Kiang (1992), who were the precursors in the implementation of methods based on AI and ANN to predict bankruptcy and that of Altman et al. (1994), which also used methods based on ANN, but in this case, Altman was the first to apply the discriminatory analysis methodology to predict firms' financial distress. The two articles represent 11.9% of the total citations among the 165 articles in the BP.

The models presented by Altman (1968) and Ohlson (1980), LR, LDA and MDA, are still considered relevant for their predictive capacity, simplicity and consistency, which few models based on AI have managed to present (Barboza et al., 2017). Among the 104 methods and techniques presented in the analyzed BP, these three methods are still among the 8 most applied and compared in the studies, suggesting that even in the face of AI-based models, traditional statistical methods are still present due to the understanding and ease of application.

In relation to the time horizon of use of the data sets, we found that there was a concentration of use from the period 1999 to 2007, the most recent bases used were up to the year 2015 and only two surveys reached this period. The demand to update the models, with larger, more updated data sets and for a longer period of time, can have broader implications in terms of performance, this has a positive impact that can be applied in organizations and intuitions, improving operation and increasing profitability.

We observed that in relation to the origin of the data sets, that is, the countries that were used for the experiments, most of the works used data sets from Germany, Australia, US, Japan, Taiwan, China, Korea and France. The use of these bases to deal with credit risk management, allows other researchers to compare new models and their results. However, when these data are overused, they may not present the specific conditions of other regions, markets, or segments, due to regulatory issues and specific socio-economic conditions, so the experiments can lead to distorted conclusions and may not meet the demands of several markets, because the factors that affect bankruptcy vary between different countries (Balcaen and Ooghe, 2006; García et al., 2014; Liang et al., 2018; Tian and Yu, 2017). Thus, it is necessary to add research from other countries, such as Brazil, Latin American countries, Israel, which, in this BP database, had no research in this portfolio.

The present research does not intend to generalize the results in relation to the treated topic, nor to understand in detail all the production involved in the studies. Thus, this work presented, through a systematic review of the literature, several analyzes through indicators, maps and networks, demonstrating the monitoring of the evolution of a certain area of knowledge, its architecture, characteristics, distribution and relevance. Some limitations were identified, such as the restriction of databases for consultation and the formulation of other analyzes, questions and objectives, such as performance analysis by method, thus allowing the identification of new opportunities for future research.

References

Abdou, H. A. and Pointon, J. (2011), 'Credit scoring, statistical techniques and evaluation criteria: a review of the literature', *Intelligent systems in accounting, finance and management* **18**(2-3), 59–88.

Abellán, J. and Castellano, J. G. (2017), 'A comparative study on base classifiers in ensemble methods for credit scoring', *Expert Systems with Applications* **73**, 1–10.

Abellán, J. and Mantas, C. J. (2014), 'Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring', *Expert Systems with Applications* **41**(8), 3825–3830.

Addo, P. M., Guegan, D. and Hassani, B. (2018), 'Credit risk analysis using machine and deep learning models', *Risks* **6**(2), 1–20.

AghaeiRad, A., Chen, N. and Ribeiro, B. (2017), 'Improve credit scoring using transfer of learned knowledge from self-organizing map', *Neural Computing and Applications* **28**(6), 1329–1342.

Ahn, H. and jae Kim, K. (2009), 'Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach', *Applied Soft Computing Journal* **9**(2), 599–607.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O. and Bilal, M. (2018), 'Systematic review of bankruptcy prediction models: Towards a framework for tool selection', *Expert Systems with Applications* **94**, 164–184. **URL:** *https://doi.org/10.1016/j.eswa.2017.10.040*

Ala'raj, M. and Abbod, M. F. (2016*a*), 'A new hybrid ensemble credit scoring model based on classifiers consensus system approach', *Expert Systems with Applications* **64**, 36–55. **URL:** *http://dx.doi.org/10.1016/j.eswa.2016.07.017*

Ala'Raj, M. and Abbod, M. F. (2016*b*), 'Classifiers consensus system approach for credit scoring', *Knowledge-Based Systems* **104**, 89–105. **URL:** *http://dx.doi.org/10.1016/j.knosys.2016.04.013*

Alfaro, E., García, N., Gámez, M. and Elizondo, D. (2008), 'Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks', *Decision Support Systems* **45**(1), 110– 122.

Altman, E. I. (1968), 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *The journal of finance* **23**(4), 589–609.

Altman, E. I. (2002), 'Managing credit risk: A challenge for the new millennium', *Economic Notes* **31**(2), 201–214.

Altman, E. I., Baidya, T. K. N. and Dias, L. M. R. (1979), 'Assessing Potential Financial Problems for Firms in Brazil', *Journal of International Business Studies* **10**(2), 9–24.

Altman, E. I., Marco, G. and Varetto, F. (1994), 'Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)', *Journal of Banking and Finance* **18**(3), 505–529.

Antunes, F., Ribeiro, B. and Pereira, F. (2017), 'Probabilistic modeling and visualization for bankruptcy prediction', *Applied Soft Computing Journal* **60**, 831–843. **URL:** *http://dx.doi.org/10.1016/j.asoc.2017.06.043*

Aziz, M. A. and Dar, H. A. (2006), 'Predicting corporate bankruptcy: Where we stand?', *Corporate Governance* **6**(1), 18–33.

Bahrammirzaee, A. (2010), 'A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems', *Neural Computing and Applications* **19**(8), 1165–1195.

Balcaen, S. and Ooghe, H. (2006), '35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems', *British Accounting Review* **38**(1), 63–93.

Barboza, F., Kimura, H. and Altman, E. (2017), 'Machine learning models and bankruptcy prediction', *Expert Systems with Applications* **83**, 405–417. **URL:** *http://dx.doi.org/10.1016/j.eswa.2017.04.006*

Beaver, W. H. (1966), 'of Failure Financial Ratios as Predictors', *Journal of Accounting Research* **4**(1966), 71–111.

Begley, J., Ming, J. and Watts, S. (1996), 'Bankruptcy classification errors in the 1980s: an empirical analysis of altman's and ohlson's models', *Review of Accounting Studies* **1**(4), 267– 284.

Bekhet, H. A. and Eletter, S. F. K. (2014), 'Credit risk assessment model for Jordanian commer- cial banks: Neural scoring approach', *Review of Development Finance* **4**(1), 20–28. **URL:** *http://dx.doi.org/10.1016/j.rdf.2014.03.002*

Bequé, A. and Lessmann, S. (2017), 'Extreme learning machines for credit scoring: An empirical evaluation', *Expert Systems with Applications* **86**, 42–53.

Beynon, M. J. and Peel, M. J. (2001), 'Variable precision rough set theory and data discretisation: An application to corporate failure prediction', *Omega* **29**(6), 561–576.

Bojović , S., Matić , R., Popović , Z., Smiljanić , M., Stefanović , M. and Vidaković , V. (2014), 'An overview of forestry journals in the period 2006–2010 as basis for ascertaining research trends', *Scientometrics* **98**(2), 1331–1346.

Borgman, C. L. (1989), 'Bibliometrics and scholarly communication: Editor's introduction', *Communication Research* **16**(5), 583–599.

Bose, I. (2006), 'Deciding the financial health of dot-coms using rough sets', *Information and Management* **43**(7), 835–846.

Boyacioglu, M. A., Kara, Y. and Baykan, Ö. K. (2009), 'Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey', *Expert Systems with Applications* **36**(2 PART 2), 3355–3366.

Brealey, R. A. and Myers, S. C. (1996), *Principles of financial management.*, McGraw-Hill, New York.

Brockett, P. L., Cooper, W. W., Golden, L. L. and Pitaktong, U. (1994), 'A Neural Network Method for Obtaining an Early Warning of Insurer Insolvency', *The Journal of Risk and Insurance* **61**(3), 402.

Brown, I. and Mues, C. (2012), 'An experimental comparison of classification algorithms for imbalanced credit scoring data sets', *Expert Systems with Applications* **39**(3), 3446–3453. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.09.033*

Chandra, D. K., Ravi, V. and Bose, I. (2009), 'Failure prediction of dotcom companies using hybrid intelligent techniques', *Expert Systems with Applications* **36**(3 PART 1), 4830–4837. **URL:** *http://dx.doi.org/10.1016/j.eswa.2008.05.047*

Chang, T. M. and Hsu, M. F. (2018), 'Integration of incremental filter-wrapper selection strategy with artificial intelligence for enterprise risk management', *International Journal of Machine Learning and Cybernetics* **9**(3), 477–489.

Chaudhuri, A. and De, K. (2011), 'Fuzzy Support Vector Machine for bankruptcy prediction', *Applied Soft Computing Journal* **11**(2), 2472–2486. **URL:** *http://dx.doi.org/10.1016/j.asoc.2010.10.003*

Chen, F. L. and Li, F. C. (2010), 'Combination of feature selection approaches with SVM in credit scoring', *Expert Systems with Applications* **37**(7), 4902–4909. **URL:** *http://dx.doi.org/10.1016/j.eswa.2009.12.025*

Chen, M. Y. (2011), 'Predicting corporate financial distress based on integration of decision tree classification and logistic regression', *Expert Systems with Applications* **38**(9), 11261–11272. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.02.173*

Chen, N., Ribeiro, B. and Chen, A. (2016), 'Financial credit risk assessment: a recent review', *Artificial Intelligence Review* **45**(1), 1–23.

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. **URL:** *http://dx.doi.org/10.1016/j.eswa.2012.07.047*

Chen, N., Ribeiro, B., Vieira, A. S., Duarte, J. and Neves, J. C. (2011), 'A genetic algorithm- based approach to cost-sensitive bankruptcy prediction', *Expert Systems with Applications* **38**(10), 12939–12945. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.04.090*

Chen, W. H. and Shih, J. Y. (2006), 'A study of Taiwan's issuer credit rating systems using support vector machines', *Expert Systems with Applications* **30**(3), 427–435.

Chen, W. S. and Du, Y. K. (2009), 'Using neural networks and data mining techniques for the financial distress prediction model', *Expert Systems with Applications* **36**(2 PART 2), 4075– 4086. **URL:** *http://dx.doi.org/10.1016/j.eswa.2008.03.020*

Chen, Y. S. and Cheng, C. H. (2013), 'Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry', *Knowledge-Based Systems* **39**, 224– 239. **URL:** *http://dx.doi.org/10.1016/j.knosys.2012.11.004*

Cheng, C., Jones, S. and Moser, W. J. (2018), 'Abnormal trading behavior of specific types of shareholders before US firm bankruptcy and its implications for firm bankruptcy prediction', *Journal of Business Finance and Accounting* **45**(9-10), 1100–1138.

Cho, S., Kim, J. and Bae, J. K. (2009), 'An integrative model with subject weight based on neural network learning for bankruptcy prediction', *Expert Systems with Applications* **36**(1), 403– 410. **URL:** *http://dx.doi.org/10.1016/j.eswa.2007.09.060*

Chou, C. H., Hsieh, S. C. and Qiu, C. J. (2017), 'Hybrid genetic algorithm and fuzzy clustering for bankruptcy prediction', *Applied Soft Computing Journal* **56**, 298–316. **URL:** *http://dx.doi.org/10.1016/j.asoc.2017.03.014*

Chuang, C. L. (2013), 'Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction', *Information Sciences* **236**, 174–185. **URL:** *http://dx.doi.org/10.1016/j.ins.2013.02.015*

Cielen, A., Peeters, L. and Vanhoof, K. (2004), 'Bankruptcy prediction using a data envelopment analysis', *European Journal of Operational Research* **154**(2), 526–532.

Cole, F. J. and Eales, N. B. (1917), 'The history of comparative anatomy : Part i - a statistical analysis of the literature', *Science Progress* **11**(44), 578–596.

Copleland, T. E. and Weston, J. F. (1988), *Financial Theory and Corporate Policy*, Mass. Addison-Wesley Publishing Company.

Danenas, P. and Garsva, G. (2015), 'Selection of Support Vector Machines based classifiers for credit risk domain', *Expert Systems with Applications* **42**(6), 3194–3204. **URL:** *http://dx.doi.org/10.1016/j.eswa.2014.12.001*

De Andrés, J., Landajo, M. and Lorca, P. (2005), 'Forecasting business profitability by using classification techniques: A comparative analysis based on a Spanish case', *European Journal of Operational Research* **167**(2), 518–542.

De 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.

De Azevedo, R. C., Ensslin, L. and Jungles, A. E. (2014), 'A Review of Risk Management in Construction: Opportunities for Improvement', *Modern Economy* **05**(04), 367–383.

Doumpos, M. and Zopounidis, C. (2011), 'A Multicriteria Outranking Modeling Approach for Credit Rating', *Decision Sciences* **42**(3), 721–742.

Du Jardin, P. (2016), 'A two-stage classification technique for bankruptcy prediction', *European Journal of Operational Research* **254**(1), 236–252. **URL:** *http://dx.doi.org/10.1016/j.ejor.2016.03.008*

Du Jardin, P. (2017), 'Dynamics of firm financial evolution and bankruptcy prediction', *Expert Systems with Applications* **75**, 25–43. **URL:** *http://dx.doi.org/10.1016/j.eswa.2017.01.016*

Du Jardin, P. (2018), 'Failure pattern-based ensembles applied to bankruptcy forecasting', *Deci- sion Support Systems* **107**, 64–77. **URL:** *https://doi.org/10.1016/j.dss.2018.01.003*

Du Jardin, P. and Séverin, E. (2012), 'Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time', *European Journal of Operational Research* **221**(2), 378–396. **URL:** *http://dx.doi.org/10.1016/j.ejor.2012.04.006*

Ensslin, L., Ensslin, S. R., Dutra, A., Nunes, N. A. and Reis, C. (2017), 'BPM governance: a literature analysis of performance evaluation', *Business Process Management Journal* **23**(1), 71–86.

Etheridge, H. L., Sriram, R. S. and Hsu, H. Y. K. (2000), 'A Comparison of Selected Artificial Neural Networks that Help Auditors Evaluate Client Financial Viability', *Decision Sciences* **31**(2), 531–550.

Fedorova, E., Gilenko, E. and Dovzhenko, S. (2013), 'Bankruptcy prediction for Russian compa- nies: Application of combined classifiers', *Expert Systems with Applications* **40**(18), 7285– 7293. **URL:** *http://dx.doi.org/10.1016/j.eswa.2013.07.032*

Feng, X., Xiao, Z., Zhong, B., Dong, Y. and Qiu, J. (2019), 'Dynamic weighted ensemble classification for credit scoring using Markov Chain', *Applied Intelligence* **49**(2), 555–568.

Feng, X., Xiao, Z., Zhong, B., Qiu, J. and Dong, Y. (2018), 'Dynamic ensemble classification for credit scoring using soft probability', *Applied Soft Computing Journal* **65**, 139–151. **URL:** *https://doi.org/10.1016/j.asoc.2018.01.021*

Finlay, S. (2011), 'Multiple classifier architectures and their application to credit risk assessment', *European Journal of Operational Research* **210**(2), 368–378. **URL:** *http://dx.doi.org/10.1016/j.ejor.2010.09.029*

Florez-Lopez, R. (2007), 'Modelling of insurers' rating determinants. An application of ma- chine learning techniques and statistical models', *European Journal of Operational Research* **183**(3), 1488–1512.

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.

García, V., Marqués, A. I. and Sánchez, J. S. (2019), 'Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction', *Information Fusion* **47**, 88–101.

García, V., Marqués, A. I., Sánchez, J. S. and Ochoa-Domínguez, H. J. (2019), 'Dissimilarity- Based Linear Models for Corporate Bankruptcy Prediction', *Computational Economics* **53**(3), 1019–1031. **URL:** *https://doi.org/10.1007/s10614-017-9783-4*

Geng, R., Bose, I. and Chen, X. (2015), 'Prediction of financial distress: An empirical study of listed Chinese companies using data mining', *European Journal of Operational Research* **241**(1), 236–247. **URL:** *http://dx.doi.org/10.1016/j.ejor.2014.08.016*

Glänzel, W. (2001*a*), 'Coauthorship patterns and trends in the sciences (1980-1998): A biblio- metric study with implications for database indexing and search strategies', *Library Trends* **50**(3), 461–473.

Glänzel, W. (2001*b*), 'National characteristics in international scientific co-authorship relations', *Scientometrics* **51**(1), 69–115.

Gorzałczany, M. B. and Rudziń ski, F. (2016), 'A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability', *Applied Soft Computing Journal* **40**, 206–220.

Griffin, J. M. and Lemmon, M. L. (2002), '1540-6261.00497.Pdf, **LVII**(5), 2317–2336.

Grilli, R., Tedeschi, G. and Gallegati, M. (2015), 'Markets connectivity and financial contagion', *Journal of Economic Interaction and Coordination* **10**(2), 287–304.

Guo, Y., Zhou, W., Luo, C., Liu, C. and Xiong, H. (2016), 'Instance-based credit risk assess- ment for investment decisions in P2P lending', *European Journal of Operational Research* **249**(2), 417–426. **URL:** *http://dx.doi.org/10.1016/j.ejor.2015.05.050*

Hajek, P. and Michalak, K. (2013), 'Feature selection in corporate credit rating prediction', *Knowledge-Based Systems* **51**, 72–84. **URL:** *http://dx.doi.org/10.1016/j.knosys.2013.07.008*

Hajek, P., Olej, V. and Myskova, R. (2014), 'Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making', *Technological and Economic Development of Economy* **20**(4), 721–738.

Härdle, W., Lee, Y. J., Schäfer, D. and Yeh, Y. R. (2009), 'Variable selection and oversampling in the use of smooth support vector machines for predicting the default risk of companies', *Journal of Forecasting* **28**(6), 512–534.

Heo, J. and Yang, J. Y. (2014), 'AdaBoost based bankruptcy forecasting of Korean construction companies', *Applied Soft Computing Journal* **24**, 494–499. **URL:** *http://dx.doi.org/10.1016/j.asoc.2014.08.009*

Hillegeist, S. A., Keating, E. K., Cram, D. P. and Lundstedt, K. G. (2004), 'Assessing the probability of bankruptcy', *Review of Accounting Studies* **9**(1), 5–34.

Hsieh, T. J., Hsiao, H. F. and Yeh, W. C. (2012), 'Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm', *Neurocomputing* **82**, 196–206. **URL:** *http://dx.doi.org/10.1016/j.neucom.2011.11.020*

Hu, Y. C. (2009), 'Bankruptcy prediction using ELECTRE-based single-layer perceptron', *Neurocomputing* **72**(13-15), 3150–3157. **URL:** *http://dx.doi.org/10.1016/j.neucom.2009.03.002*

Hu, Y. C. and Ansell, J. (2007), 'Measuring retail company performance using credit scoring techniques', *European Journal of Operational Research* **183**(3), 1595–1606.

Huang, J. J., Tzeng, G. H. and Ong, C. S. (2006), 'Two-stage genetic programming (2SGP) for the credit scoring model', *Applied Mathematics and Computation* **174**(2), 1039–1053.

Huang, S. M., Tsai, C. F., Yen, D. C. and Cheng, Y. L. (2008), 'A hybrid financial analysis model for business failure prediction', *Expert Systems with Applications* **35**(3), 1034–1040.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H. and Wu, S. (2004), 'Credit rating analysis with support vector machines and neural networks: A market comparative study', *Decision Support Systems* **37**(4), 543–558.

Hung, C. and Chen, J. H. (2009), 'A selective ensemble based on expected probabilities for bankruptcy prediction', *Expert Systems with Applications* **36**(3 PART 1), 5297–5303. **URL:** *http://dx.doi.org/10.1016/j.eswa.2008.06.068*

Jo, H. and Han, I. (1996), 'Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction', *Expert Systems with Applications* **11**(4 SPEC. ISS.), 415– 422.

Jones, S. (2017), 'Corporate bankruptcy prediction: a high dimensional analysis', *Review of Accounting Studies* **22**(3), 1366–1422.

Jones, S., Johnstone, D. and Wilson, R. (2015), 'An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes', *Journal of Banking and Finance* **56**, 72–85. **URL:** *http://dx.doi.org/10.1016/j.jbankfin.2015.02.006*

Jones, S., Johnstone, D. and Wilson, R. (2017), 'Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks', *Journal of Business Finance and Accounting* **44**(1- 2), 3–34.

Karan, M. B., Ulucan, A. and Kaya, M. (2013), 'Credit risk estimation using payment history data: A comparative study of Turkish retail stores', *Central European Journal of Operations Research* **21**(2), 479–494.

Kim, K. J. and Ahn, H. (2012), 'A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach', *Computers and Operations Research* **39**(8), 1800–1811. **URL:** *http://dx.doi.org/10.1016/j.cor.2011.06.023*

Kim, M. J. and Han, I. (2003), 'The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms', *Expert Systems with Applications* **25**(4), 637–646.

Kim, M. J. and Kang, D. K. (2010), 'Ensemble with neural networks for bankruptcy prediction', *Expert Systems with Applications* **37**(4), 3373–3379. **URL:** *http://dx.doi.org/10.1016/j.eswa.2009.10.012*

Kim, M. J. and Kang, D. K. (2012), 'Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction', *Expert Systems with Applications* **39**(10), 9308–9314. **URL:** *http://dx.doi.org/10.1016/j.eswa.2012.02.072*

Kim, M. J., Kang, D. K. and Kim, H. B. (2015), 'Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction', *Expert Systems with Applications* **42**(3), 1074–1082. **URL:** *http://dx.doi.org/10.1016/j.eswa.2014.08.025*

Kim, S. Y. and Upneja, A. (2014), 'Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models', *Economic Modelling* **36**, 354–362. **URL:** *http://dx.doi.org/10.1016/j.econmod.2013.10.005*

Ko, P. C. and Lin, P. C. (2006), 'An evolution-based approach with modularized evaluations to forecast financial distress', *Knowledge-Based Systems* **19**(1), 84–91.

Korol, T. (2013), 'Early warning models against bankruptcy risk for Central European and Latin American enterprises', *Economic Modelling* **31**(1), 22–30. **URL:** *http://dx.doi.org/10.1016/j.econmod.2012.11.017*

Kou, G., Peng, Y. and Lu, C. (2014), 'MCDM approach to evaluating bank loan default models', *Technological and Economic Development of Economy* **20**(2), 292–311.

Koutanaei, F. N., Sajedi, H. and Khanbabaei, M. (2015), 'A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring', *Journal of Retailing and Consumer Services* **27**, 11–23. **URL:** *http://dx.doi.org/10.1016/j.jretconser.2015.07.003*

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. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.12.021*

Kwak, W., Shi, Y., Eldridge, S. W. and Kou, G. (2006), 'Bankruptcy prediction for Japanese firms: Using Multiple Criteria Linear Programming data mining approach', *International Journal of Business Intelligence and Data Mining* **1**(4), 401–416.

Lahsasna, A., Ainon, R. N. and Wah, T. Y. (2010), 'Credit scoring models using soft computing methods: A survey', *International Arab Journal of Information Technology* **7**(2), 115–123.

Lee, K. C., Han, I. and Kwon, Y. (1996), 'Hybrid neural network models for bankruptcy predictions', *Decision Support Systems* **18**(1 SPEC. ISS.), 63–72.

Lee, Y. C. (2007), 'Application of support vector machines to corporate credit rating prediction', *Expert Systems with Applications* **33**(1), 67–74.

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

Li, F. and Perez-Saiz, H. (2018), 'Measuring systemic risk across financial market infrastructures', *Journal of Financial Stability* **34**, 1–11. **URL:** *http://dx.doi.org/10.1016/j.jfs.2017.08.003*

Li, H. and Sun, J. (2009), 'Hybridizing principles of the Electre method with case-based reasoning for data mining: Electre-CBR-I and Electre-CBR-II', *European Journal of Operational Research* **197**(1), 214–224. **URL:** *http://dx.doi.org/10.1016/j.ejor.2008.05.024*

Li, H. and Sun, J. (2010), 'Business failure prediction using hybrid2 case-based reasoning (H2CBR)', *Computers and Operations Research* **37**(1), 137–151.

Li, H. and Sun, J. (2011), 'Principal component case-based reasoning ensemble for business failure prediction', *Information and Management* **48**(6), 220–227. **URL:** *http://dx.doi.org/10.1016/j.im.2011.05.001*

Li, H. and Sun, J. (2012), 'Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples - Evidence from the Chinese hotel industry', *Tourism Management* **33**(3), 622–634. **URL:** *http://dx.doi.org/10.1016/j.tourman.2011.07.004*

Li, H., Sun, J. and Sun, B. L. (2009), 'Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors', *Expert Systems with Applications* **36**(1), 643–659. **URL:** *http://dx.doi.org/10.1016/j.eswa.2007.09.038*

Liang, D., Lu, C. C., Tsai, C. F. and Shih, G. A. (2016), 'Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study', *European Journal of Operational Research* **252**(2), 561–572. **URL:** *http://dx.doi.org/10.1016/j.ejor.2016.01.012*

Liang, D., Tsai, C. F., Dai, A. J. and Eberle, W. (2018), 'A novel classifier ensemble approach for financial distress prediction', *Knowledge and Information Systems* **54**(2), 437–462.

Liang, D., Tsai, C. F. and Wu, H. T. (2015), 'The effect of feature selection on financial distress prediction', *Knowledge-Based Systems* **73**(1), 289–297. **URL:** *http://dx.doi.org/10.1016/j.knosys.2014.10.010*

Lin, F., Liang, D. and Chen, E. (2011), 'Financial ratio selection for business crisis prediction', *Expert Systems with Applications* **38**(12), 15094–15102. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.05.035*

Lin, F. Y. and McClean, S. (2001), A data mining approach to the prediction of corporate failure, *in* 'Applications and Innovations in Intelligent Systems VIII', Springer, pp. 93–106.

Lin, R. H., Wang, Y. T., Wu, C. H. and Chuang, C. L. (2009), 'Developing a business failure prediction model via RST, GRA and CBR', *Expert Systems with Applications* **36**(2 PART 1), 1593–1600. **URL:** *http://dx.doi.org/10.1016/j.eswa.2007.11.068*

Lin, S. J., Chang, C. and Hsu, M. F. (2013), 'Multiple extreme learning machines for a two-class imbalance corporate life cycle prediction', *Knowledge-Based Systems* **39**, 214–223. **URL:** *http://dx.doi.org/10.1016/j.knosys.2012.11.003*

Lin, W. C., Lu, Y. H. and Tsai, C. F. (2019), 'Feature selection in single and ensemble learning- based bankruptcy prediction models', *Expert Systems* **36**(1), 1–8.

Lin, W. Y., Hu, Y. H. and Tsai, C. F. (2012), 'Machine learning in financial crisis prediction: A survey', *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* **42**(4), 421–436.

Liou, F. M. (2008), 'Fraudulent financial reporting detection and business failure prediction models: A comparison', *Managerial Auditing Journal* **23**(7), 650–662.

Liu, W., Gu, M., Hu, G., Li, C., Liao, H., Tang, L. and Shapira, P. (2014), 'Profile of developments in biomass-based bioenergy research: a 20-year perspective', *Scientometrics* **99**(2), 507–521.

Liu, Y. and Schumann, M. (2005), 'Data mining feature selection for credit scoring models', *Journal of the Operational Research Society* **56**(9), 1099–1108.

Lopez, J. A. and Saidenberg, M. R. (2000), 'Evaluating credit risk models', *Journal of Banking & Finance* **24**(1-2), 151–165.

Lozano, S., Calzada-Infante, L., Adenso-Díaz, B. and García, S. (2019), 'Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature', *Scientometrics* **120**(2), 609–629.

Marqués, A. I., García, V. and Sánchez, J. S. (2012*a*), 'Exploring the behaviour of base classifiers in credit scoring ensembles', *Expert Systems with Applications* **39**(11), 10244–10250.

Marqués, A. I., García, V. and Sánchez, J. S. (2012*b*), 'Two-level classifier ensembles for credit risk assessment', *Expert Systems with Applications* **39**(12), 10916–10922.

Marqués, A. I., García, V. and Sánchez, J. S. (2013*a*), 'A literature review on the application of evolutionary computing to credit scoring', *Journal of the Operational Research Society* **64**(9), 1384–1399.

Marqués, A. I., García, V. and Sánchez, J. S. (2013*b*), 'On the suitability of resampling techniques for the class imbalance problem in credit scoring', *Journal of the Operational Research Society* **64**(7), 1060–1070.

Martin, D. (1977), 'Early warning of bank failure. A logit regression approach', *Journal of Banking and Finance* **1**(3), 249–276.

McKee, T. E. (2003), 'Rough sets bankruptcy prediction models versus auditor signalling rates', *Journal of Forecasting* **22**(8), 569–586.

McKee, T. E. and Lensberg, T. (2002), 'Genetic programming and rough sets: A hybrid approach to bankruptcy classification', *European Journal of Operational Research* **138**(2), 436–451.

Min, J. H. and Jeong, C. (2009), 'A binary classification method for bankruptcy prediction', *Expert Systems with Applications* **36**(3 PART 1), 5256–5263. **URL:** *http://dx.doi.org/10.1016/j.eswa.2008.06.073*

Min, J. H. and Lee, Y. C. (2005), 'Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters', *Expert Systems with Applications* **28**(4), 603– 614.

Min, J. H. and Lee, Y. C. (2008), 'A practical approach to credit scoring', *Expert Systems with Applications* **35**(4), 1762–1770.

Min, S. H., Lee, J. and Han, I. (2006), 'Hybrid genetic algorithms and support vector machines for bankruptcy prediction', *Expert Systems with Applications* **31**(3), 652–660.

Moro, S., Cortez, P. and Rita, P. (2015), 'Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation', *Expert Systems with Applications* **42**(3), 1314–1324.

Nanni, L. and Lumini, A. (2009), 'An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring', *Expert Systems with Applications* **36**(2 PART 2), 3028–3033. **URL:** *http://dx.doi.org/10.1016/j.eswa.2008.01.018*

Odom, M. D. and Sharda, R. (1990), 'A neural network model for bankruptcy prediction', *IJCNN. International Joint Conference on Neural Networks* pp. 163–168.

Ohlson, J. A. (1980), 'Financial Ratios and the Probabilistic Prediction of Bankruptcy', *Journal of Accounting Research* **18**(1), 109.

Olson, D. L., Delen, D. and Meng, Y. (2012), 'Comparative analysis of data mining methods for bankruptcy prediction', *Decision Support Systems* **52**(2), 464–473. **URL:** *http://dx.doi.org/10.1016/j.dss.2011.10.007*

Öğüt, H., Doğanay, M. M., Ceylan, N. B. and Aktaş, R. (2012), 'Prediction of bank financial strength ratings: The case of Turkey', *Economic Modelling* **29**(3), 632–640.

Paleologo, G., Elisseeff, A. and Antonini, G. (2010), 'Subagging for credit scoring models', *European Journal of Operational Research* **201**(2), 490–499. **URL:** *http://dx.doi.org/10.1016/j.ejor.2009.03.008*

Pan, W. T. (2012), 'A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example', *Knowledge-Based Systems* **26**, 69–74. **URL:** *http://dx.doi.org/10.1016/j.knosys.2011.07.001*

Pendharkar, P. C. (2005), 'A threshold-varying artificial neural network approach for classi- fication and its application to bankruptcy prediction problem', *Computers and Operations Research* **32**(10), 2561–2582.

Polonioli, A. (2020), 'In search of better science: on the epistemic costs of systematic reviews and the need for a pluralistic stance to literature search', *Scientometrics* **122**(2), 1267–1274.

Pritchard, A. (1969), 'Pritchard1969', *Journal of documentation* **25**(4), 348–349.

Ribeiro, B., Chen, N. and Kovacec, A. (2019), 'Shaping graph pattern mining for financial risk', *Neurocomputing* **326-327**, 123–131. **URL:** *http://dx.doi.org/10.1016/j.neucom.2017.01.119*

Ribeiro, B., Silva, C., Chen, N., Vieira, A. and Carvalho Das Neves, J. (2012), 'Enhanced default risk models with SVM+', *Expert Systems with Applications* **39**(11), 10140–10152. **URL:** *http://dx.doi.org/10.1016/j.eswa.2012.02.142*

Rochon, L. P. and Rossi, S. (2010), 'Has "It" Happened Again?', *International Journal of Political Economy* **39**(2), 5–9.

Salcedo-Sanz, S., Fernández-Villacañas, J. L., Segovia-Vargas, M. J. and Bousoño-Calzón,

C. (2005), 'Genetic programming for the prediction of insolvency in non-life insurance companies', *Computers and Operations Research* **32**(4), 749–765.

Scherr, F. C. (1989), *Modern Working Capital Management*, Prentice Hall.

Serrano-Cinca, C. (1996), 'Self organizing neural networks for financial diagnosis', *Decision Support Systems* **17**(3), 227–238.

Serrano-Cinca, C. and Gutiérrez-Nieto, B. (2013), 'Partial least square discriminant analysis for bankruptcy prediction', *Decision Support Systems* **54**(3), 1245–1255. **URL:** *http://dx.doi.org/10.1016/j.dss.2012.11.015*

Shin, K. S., Lee, T. S. and Kim, H. J. (2005), 'An application of support vector machines in bankruptcy prediction model', *Expert Systems with Applications* **28**(1), 127–135.

Shin, K. S. and Lee, Y. J. (2002), 'A genetic algorithm application in bankruptcy prediction modeling', *Expert Systems with Applications* **23**(3), 321–328.

Sun, J., Fujita, H., Chen, P. and Li, H. (2017), 'Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble', *Knowledge-Based Systems* **120**, 4–14.

Sun, J., Jia, M. Y. and Li, H. (2011), 'AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies', *Expert Systems with Applications* **38**(8), 9305–9312. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.01.042*

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.

Sun, J. and Li, H. (2008), 'Data mining method for listed companies' financial distress prediction', *Knowledge-Based Systems* **21**(1), 1–5.

Sun, J., Li, H., Huang, Q. H. and He, K. Y. (2014), 'Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches', *Knowledge-Based Systems* **57**, 41–56. **URL:** *http://dx.doi.org/10.1016/j.knosys.2013.12.006*

Sung, T. K., Chang, N. and Lee, G. (1999), 'Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction', *Journal of Management Information Systems* **16**(1), 63– 85.

Taffler, R. J. (1982), 'Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data', *Journal of the Royal Statistical Society* **145**(3), 342–358.

Tam, K. Y. and Kiang, M. (1990), 'Predicting bank failures: A neural network approach', *Applied Artificial Intelligence* **4**(4), 265–282.

Tam, K. Y. and Kiang, M. Y. (1992), 'Managerial applications of neural networks: The case of bank failure predictions', *Management Science* **38**(7), 926–947.

Tang, L. (2013), 'Does "birds of a feather flock together" matter—evidence from a longitudinal study on us–china scientific collaboration', *Journal of Informetrics* **7**(2), 330–344.

Tang, X., Li, S., Tan, M. and Shi, W. (n.d.), 'Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods', *Journal of Forecasting* .

Tascón Fernández, M. T. and Castaño Gutiérrez, F. J. (2012), 'Variables and models for the iden- tification and prediction of business failure: Revision of recent empirical research advances', *Revista de Contabilidad-Spanish Accounting Review* **15**(1), 7–58. **URL:** *http://dx.doi.org/10.1016/S1138-4891(12)70037-7*

Tian, S. and Yu, Y. (2017), 'Financial ratios and bankruptcy predictions: An international evidence', *International Review of Economics and Finance* **51**, 510–526. **URL:** *http://dx.doi.org/10.1016/j.iref.2017.07.025*

Tian, Y., Shi, Y. and Liu, X. (2012), 'Recent advances on support vector machines research', *Technological and Economic Development of Economy* **18**(1), 5–33.

Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M. and Martens, D. (2017), 'Bankruptcy prediction for SMEs using relational data', *Decision Support Systems* **102**, 69–81. **URL:** *http://dx.doi.org/10.1016/j.dss.2017.07.004*

Tollefson, J. (2018), 'China declared world's largest producer of scientific articles', *Nature* **553**(7689), 390–390.

Tsai, C. F. (2008), 'Financial decision support using neural networks and support vector machines', *Expert Systems* **25**(4), 380–393.

Tsai, C. F. (2009), 'Feature selection in bankruptcy prediction', *Knowledge-Based Systems,* **22**(2), 120–127. **URL:** *http://dx.doi.org/10.1016/j.knosys.2008.08.002*

Tsai, C. F. (2014), 'Combining cluster analysis with classifier ensembles to predict financial distress', *Information Fusion* **16**(1), 46–58. **URL:** *http://dx.doi.org/10.1016/j.inffus.2011.12.001*

Tsai, C. F. and Chen, M. L. (2010), 'Credit rating by hybrid machine learning techniques', *Applied Soft Computing Journal* **10**(2), 374–380.

Tsai, C. F. and Cheng, K. C. (2012), 'Simple instance selection for bankruptcy prediction', *Knowledge-Based Systems* **27**, 333–342. **URL:** *http://dx.doi.org/10.1016/j.knosys.2011.09.017*

Tsai, C. F., Hsu, Y. F. and Yen, D. C. (2014), 'A comparative study of classifier ensembles for bankruptcy prediction', *Applied Soft Computing Journal* **24**, 977–984. **URL:** *http://dx.doi.org/10.1016/j.asoc.2014.08.047*

Tsai, C. F. and Wu, J. W. (2008), 'Using neural network ensembles for bankruptcy prediction and credit scoring', *Expert Systems with Applications* **34**(4), 2639–2649.

Tung, W. L., Quek, C. and Cheng, P. (2004), 'GenSo-EWS: A novel neural-fuzzy based early warning system for predicting bank failures', *Neural Networks* **17**(4), 567–587.

Twala, B. (2010), 'Multiple classifier application to credit risk assessment', *Expert Systems with Applications* **37**(4), 3326–3336. **URL:** *http://dx.doi.org/10.1016/j.eswa.2009.10.018*

Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J. and Van Raan, A. F. (2003), 'The holy grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence', *Scientometrics* **57**(2), 257–280.

Varetto, F. (1998), 'Genetic algorithms applications in the analysis of insolvency risk', *Journal of Banking and Finance* **22**(10-11), 1421–1439.

Verikas, A., Kalsyte, Z., Bacauskiene, M. and Gelzinis, A. (2010), 'Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: A survey', *Soft Computing* **14**(9), 995– 1010.

Wang, G., Chen, G. and Chu, Y. (2018), 'A new random subspace method incorporating sentiment and textual information for financial distress prediction', *Electronic Commerce Research and Applications* **29**, 30–49. **URL:** *https://doi.org/10.1016/j.elerap.2018.03.004*

Wang, G., Hao, J., Ma, J. and Jiang, H. (2011), 'A comparative assessment of ensemble learning for credit scoring', *Expert Systems with Applications* **38**(1), 223–230. **URL:** *http://dx.doi.org/10.1016/j.eswa.2010.06.048*

Wang, G. and Ma, J. (2011), 'Study of corporate credit risk prediction based on integrating boosting and random subspace', *Expert Systems with Applications* **38**(11), 13871–13878. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.04.191*

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. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.11.003*

Wang, G., Ma, J., Huang, L. and Xu, K. (2012), 'Two credit scoring models based on dual strategy ensemble trees', *Knowledge-Based Systems* **26**, 61–68. **URL:** *http://dx.doi.org/10.1016/j.knosys.2011.06.020*

Wang, G., Ma, J. and Yang, S. (2014), 'An improved boosting based on feature selection for corporate bankruptcy prediction', *Expert Systems with Applications* **41**(5), 2353–2361. **URL:** *http://dx.doi.org/10.1016/j.eswa.2013.09.033*

Wang, J., Veugelers, R. and Stephan, P. (2017), 'Bias against novelty in science: A cautionary tale for users of bibliometric indicators', *Research Policy* **46**(8), 1416–1436.

Wang, Y., Wang, S. and Lai, K. K. (2005), 'A new fuzzy support vector machine to evaluate credit risk', *IEEE Transactions on Fuzzy Systems* **13**(6), 820–831.

Wilson, R. L. and Sharda, R. (1994), 'Bankruptcy prediction using neural networks', *Decision support systems* **11**(5), 545–557.

Wu, C. H., Tzeng, G. H., Goo, Y. J. and Fang, W. C. (2007), 'A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy', *Expert Systems with Applications* **32**(2), 397–408.

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.

Yang, Z., You, W. and Ji, G. (2011), 'Using partial least squares and support vector machines for bankruptcy prediction', *Expert Systems with Applications* **38**(7), 8336–8342. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.01.021*

Yu, L., Wang, S. and Lai, K. K. (2008), 'Credit risk assessment with a multistage neural network ensemble learning approach', *Expert Systems with Applications* **34**(2), 1434–1444.

Yu, L., Yao, X., Wang, S. and Lai, K. K. (2011), 'Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection', *Expert Systems with Applications* **38**(12), 15392–15399. **URL:** *http://dx.doi.org/10.1016/j.eswa.2011.06.023*

Yu, Q., Miche, Y., Séverin, E. and Lendasse, A. (2014), 'Bankruptcy prediction using Extreme Learning Machine and financial expertise', *Neurocomputing* **128**, 296–302.

Zelenkov, Y., Fedorova, E. and Chekrizov, D. (2017), 'Two-step classification method based on genetic algorithm for bankruptcy forecasting', *Expert Systems with Applications* **88**, 393–401. **URL:** *http://dx.doi.org/10.1016/j.eswa.2017.07.025*

Zhang, G., Hu, M. Y., Patuwo, B. E. and Indro, D. C. (1999), 'Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis', *European Journal of Operational Research* **116**(1), 16–32.

Zhang, Y., Wang, S. and Ji, G. (2013), 'A rule-based model for bankruptcy prediction based on an improved genetic ant colony algorithm', *Mathematical Problems in Engineering* **2013**.

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. **URL:** *http://dx.doi.org/10.1016/j.ejor.2014.01.044*

Zhao, H., Sinha, A. P. and Ge, W. (2009), 'Effects of feature construction on classification performance: An empirical study in bank failure prediction', *Expert Systems with Applications* **36**(2 PART 2), 2633–2644. **URL:** *http://dx.doi.org/10.1016/j.eswa.2008.01.053*

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. **URL:** *http://dx.doi.org/10.1016/j.neucom.2013.02.054*

Zhou, L. (2013), 'Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods', *Knowledge-Based Systems* **41**, 16–25. **URL:** *http://dx.doi.org/10.1016/j.knosys.2012.12.007*

Zhu, Y., Xie, C., Wang, G. J. and Yan, X. G. (2017), 'Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance', *Neural Computing and Applications* **28**(s1), 41–50.

Zieba, M., Tomczak, S. K. and Tomczak, J. M. (2016), 'Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction', *Expert Systems with Applications* **58**, 93–101. **URL:** *http://dx.doi.org/10.1016/j.eswa.2016.04.001*

Zmijweski, M. E. (1984), 'Methodological Issues Related to the Estimation of Financial Distress Prediction Models', *Journal of Accounting Research* **22**, 59–82.

Table 7: Final Bibliometric Portfolio with 165 articles selected by the ProKnow-C method.

Publ.	Author	1	2	3	4	5	6	7	8
1992	Tam and Kiang (1992)	C	A	LDA-LR-KNN-ANN-DTID3	A	A	A	A	1985-1987
1994	Altman et al. (1994)	C B	A	LDA-ANN	B	B	A	A	1982-1992
1994	Brockett et al. (1994)	B	A	LR-ANN-MDA-O	B	A	B	A B	1987-1990
1994	Wilson and Sharda (1994)	C	A	MDA-ANN	B	F	A	A	1975-1982
1996	Jo and Han (1996)	C	A	LDA-ANN-CBFS	B	C	A	A	1991-1993
1996	Lee et al. (1996)	C	A	MDA-ANN-DTID3-SOM	B	C	A	A	1979-1992
1996	Serrano-Cinca (1996)	B	A	SOM-LDA-MLP-DSS	B	B	B	A	1975-1985
1998	Varetto (1998)	B	A	LDA-GA-ANN	B	B	B	A	1982-1995
1999	Sung et al. (1999)	C	A	GA-DT-MDA-ANN-BOO	B	C	A	A	1991-1998
1999	Zhang et al. (1999)	C	A	LR-ANN-CVNN	C	F	A	D	-
2000	Etheridge et al. (2000)	B	A	BPNN-PNN-CLN	A	A	C	C	1986-1988
2001	Beynon and Peel (2001)	A	A	RST-VPRS-RPA-LR-EM-MDA	B	B	C	A	-
2001	Lin and McClean (2001)	A	A	LDA-LR-ANN-DT	B	B	C	A	1980-1999
2002	McKee and Lensberg (2002)	C	A	GA-RST	B	A	C	A	1991-1997
2002	Shin and Lee (2002)	C	A	GA	B	C	A	A	1995-1997
2003	Kim and Han (2003)	C	A	ANN-GA	B	C	A	A B	2001-2002
2003	McKee (2003)	C	A	RST	B	A	A	A	1991-1997
2004	Cielen et al. (2004)	C	A	DEA-DT-O	B	B	A	A	1994-1996
2004	Huang et al. (2004)	D	A	LR-SVM-ANN-BPNN	A	A C	C	A	1998-2002
2004	Tung et al. (2004)	C	A	GSOFNN-MCMAC-COX	A	A	A	A	1980-2000
2005	Liu and Schumann (2005)	D	A	KNN-LR-NN-MLP-DTM5	C	B	C	D	-
2005	Min and Lee (2005)	C	A	ANN-SVM-BPNN-LR-MDA	B	C	A	A	2000-2002
2005	Pendharkar (2005)	C	A	GA-ANN-DT-LDA-BPNN-TV	B	F	A	A	1987-1995
2005	Salcedo-Sanz et al. (2005)	B	A	GP-SVM-RST-GPR	A	B	B	A	1983-1993
2005	Shin et al. (2005)	C	A	SVM-BPNN	B	C	A	A	1996-1999
2005	Wang et al. (2005)	D	A	LR-ANN-FSVM-RBF	B	B C	C	A	-
2006	Bose (2006)	B	A	ANN-RST	B	F	C	A	2002

2006	Chen and Shih (2006)	A	A	SVM-BPNN	A	C	C	A B	-
2006	Ko and Lin (2006)	B	A	GA-PSO-ANN-LR-LDA	B	C	B	A	1993-2002
2006	Kwak et al. (2006)	C	A	MCLP	B	C	A	A	1989-1999
2006	Min et al. (2006)	C	A	SVM-ANN-LR-GA-CBR	B	C	A	A	1999-2000
2007	Florez-Lopez (2007)	D	A	CART-DT-MDA-LR	A	B	D	A	1999-2000
2007	Hu and Ansell (2007)	D	A	LR-ANN-SMO-NB	B	A	A	A	1994-2002
2007	Lee (2007)	D	A	BPNN-SVM-MDA-CBR	B	C	C	A	1997-2002
2007	Wu et al. (2007)	C	A	SVM-GA-MDA-LR-PM-ANN	B	C	A	A	2001-2003
2008	Alfaro et al. (2008)	C	A	AB-ANN	B	B	A	A B	-
2008	Huang et al. (2008)	C	A	LDA-DT-BPNN-HIS	B	C	A	A	2001-2004
2008	Liou (2008)	A	A	LR-ANN-DT	B	C	D	A	2003-2004
2008	Min and Lee (2008)	C	A	DEA-MDA	B	C	A	A	-
2008	Sun and Li (2008)	B	A	DM	B	C	B	A	2000-2005
2008	Tsai and Wu (2008)	C D	A	ANN-BPNN-MLP	B	B C E	A	A	-
2008	Tsai (2008)	C D	A	SVM-MLP-RBF	B	B C E F	A	A	-
2008	Yu et al. (2008)	D	A	LR-ANN-SVM-FSVM-VB-RB	B	B	C	A B	-
2009	Ahn and jae Kim (2009)	C	A	MDA-CBR-GA	A	C	A	A	1996-2000
2009	Boyacioglu et al. (2009)	C	A	MLP-CL-LVQ-SOM-ANN-SVM-MDA-KM	A	C	A	A	1997-2003
2009	Chandra et al. (2009)	A	A	MLT-CART-RF-LR-SVM-DT-BOO	B	F	C	A	-
2009	Chen and Du (2009)	C	A	ANN-DM	B	C	A	A B	1999-2006
2009	Cho et al. (2009)	C	A	MDA-LR-ML-ANN-DT-SVM-RULE-IMSW	B	C	A	A	1999-2002
2009	Härdle et al. (2009)	D	A	LR-LDA-SVM	B	B	B	A	1997-2002
2009	Hu (2009)	C	A	LDA-LR-SLP-MLP-SVM-PM-FIFLN-ELECTRE	C	F	A	A	1975-1982
2009	Hung and Chen (2009)	C	A	DT-ANN-BPNN-SVM-SMO	B	F	A	A	1997-2001
2009	Li et al. (2009)	B C	A	MDA-LR-ANN-SVM-CBR	B	C	B	A	2000-2005
2009	Li and Sun (2009)	B C	A	ELECTRE-CBR	B	C	B	A	2000-2005
2009	Lin et al. (2009)	A	A	RST-CBR-GRA-HFP	B	C	C	A B	1999-2006
2009	Lin et al. (2009)	B	A	PSO-SVM-DT	A	C	B	A B	2000-2005

2009	Min and Jeong (2009)	C	A	MDA-LR-DT-ANN-BC	B	F	A	A	2001-2004
2009	Nanni (2009)	C D	A	RS-BA-RTF-SVM-MLP-CW-RBF	B	B C E	A	A	-
2009	Tsai (2009)	C D	A	MLP	B	B C E F	A	A	-
2009	Zhao et al. (2009)	C	A	LR-DT-ANN-KNN	B	F	A	A	1991-1992
2010	Bahrammirzaee (2010)	B D	B	ANN-ES-HIS	B	F	A	C	-
2010	Chen and Li (2010)	D	A	SVM-DT-LDA-RST-FS	B	B E	C	A	-
2010	Kim and Kang (2010)	C	A	ANN-BOO-BA-DT-ML	B	C	A	A	2002-2005
2010	Lahsasna et al. (2010)	D	B	O	C	F	E	D	-
2010	Li and Sun (2010)	B C	A	CBR-LR-MDA-KNN	B	C	B	A	2000-2005
2010	Paleologo et al. (2010)	D	A	SVM-RBF-KNN-DT-AB	B	B	C	A B	-
2010	Tsai and Chen (2010)	D	A	DT-NB-LR-ANN-KM	B	C	C	A	2004-2006
2010	Twala (2010)	D	A	ANN-DT-KNN-LR-BA-BOO-NB	A	A B C E	A	A	1985-1987
2010	Verikas et al. (2010)	C	B	O	C	E	E	D	-
2011	Chaudhuri and De (2011)	C	A	LR-LDA-PNN-FSVM	B	A	A	A	2000-2001
2011	Chen (2011)	B	A	LR-DT-CART	B	C	B	A B	-
2011	Chen et al. (2011)	C	A	GA-LVQ-CSL	B	B	A	A B	2006
2011	De Andrés et al. (2011)	C	A	FCN-MARS-LDA-ANN	B	C	A	A	-
2011	Doumpos and Zopounidis (2011)	D	A	O	B	B	C	A	1998-2003
2011	Finlay (2011)	D	A	LR-LDA-CART-ANN-KNN-AB-BA-BOO	B	B	C	A	2002
2011	Li and Sun (2011)	B C	A	CBR-MDA-LR	B	C	B	A	2000-2005
2011	Lin et al. (2011)	B	A	SVM-LR-MDA-RBFN	B	C	B	A B	2000-2008
2011	Sun et al. (2011)	B	A	AB-DT-SVM-SAT	B	C	B	A	2000-2008
2011	Wang et al. (2011)	D	A	BA-BOO-LR-DT-ANN-SVM	B	C E B	C	A	2006-2007
2011	Wang and Ma (2011)	D	A	LR-DT-ANN-BA-BOO-RS	B	C	C	A	2006-2007
2011	Yang et al. (2011)	C	A	RBF-MLP-SVM-ANN-PLS	B	F	A	A	-
2011	Yu et al. (2011)	D	A	SVM	B	B E	C	D	-
2012	Brown and Mues (2012)	D	A	LR-ANN-DT-SVM-RF-LDA-GBOO	B	B	B	A	-
2012	Du Jardin and Séverin (2012)	C	A	MDA-LR-ANN-COX	B	B	A	A	1991-2009
2012	Hsieh et al. (2012)	D	A	LR-ANN-SVM-EABC-BPNN-GA	B	C	A	A B	1999-2007

2012	Kim and Ahn (2012)	D	A	SVM-MDA-LR-CBR-ANN	B	C	C	A	-
2012	Kim and Kang (2012)	C	A	BA-BOO-DT-ANN-SVM	B	C	A	A	2002-2005
2012	Koyuncugil and Ozgulbas (2012)	B	A	CHAID-EWS	B	C	B	A	2007
2012	Li and Sun (2012)	B C	A	SVM-ANN-MDA-LR-KNN	B	C	B	A	1998-2010
2012	Lin et al. (2012)	B C D	B	O	C	F	E	D	1995-2010
2012	Marqués et al. (2012*a*)	D	A	BA-RS-DECORATE-AB-RTF-NN-NB-LR-MLP-RBF-ANN-SVM-DT	B	A B C E	C	A	-
2012	Marqués et al. (2012*b*)	D	A	AB-BA-RS-RF-ANN-LR-MLP-SVM-DT	B	A B C E	C	A	-
2012	Öğüt et al. (2012)	D	A	LR-MDA-SVM-PNN	A	C	C	A	2003-2009
2012	Olson et al. (2012)	C	A	LR-RBF-DT-CART-SVM-DTJ8	B	A	A	A	2005-2009
2012	Pan (2012)	B	A	FFAO-ANN-GRNN	B	C	C	A	2003-2004
2012	Ribeiro et al. (2012)	A	A	SVM-MTL	B	B	C	A B	2002-2006
2012	Tascón Fernández and Castaño Gutierres (2012)	A	B	O	C	F	E	D	-
2012	Tian et al. (2012)	C	B	SVM	C	F	E	D	-
2012	Tsai and Cheng (2012)	C	A	MLP-DT-LR-SVM	B	B C E F	A	A	-
2012	Wang et al. (2012)	D	A	RS-BA-DT-LR-LDA-MLP-RBFN-RF-RTF	B	B E	C	A	-
2012	Wang and Ma (2012)	D	A	RSB-SVM-ANN-BA-RS-BOO-LR-DT-MLP-AB-SMO	B	C	C	A	2006-2007
2013	Chen et al. (2013)	C	A	SOM	B	B	A	A B	2003-2006
2013	Chen and Cheng (2013)	D	A	RST-ML-HIS	A	A B C F	C	A	1998-2007
2013	Chuang (2013)	A	A	CBR-RST-CART-GRA-LR	B	C	A	A	1999-2006
2013	Fedorova et al. (2013)	C	A	MDA-LR-CART-ANN-AB	B	B	A	A	2007-2011
2013	Hajek and Michalak (2013)	D	A	MLP-RBF-SVM-NB-RF-LDA-NMC-ANN	B	A B E	C	A	-
2013	Korol (2013)	B	A	LDA-CART-ANN	B	B D	B	A	1996-2009
2013	Lin et al. (2013)	B	A	RF-BPNN-LDA-ML-O	B	B C E	B	A B	2009-2011
2013	Marqués et al. (2013*b*)	D	A	SVM-LR-SMOTE	B	A B C E	C	A	-
2013	Marqués et al. (2013*a*)	D	B	O	C	F	C	D	-
2013	Serrano-Cinca and Gutierrez Nieto (2013)	B	A	LDA-LR-MLP-KNN-NB-SVM-BOO-DT-BRT-PLSDA	A	A	B	A	-
2013	Zhou (2013)	C	A	LDA-LR-DT-ANN-SVM-SMOTE-OS-US	B	A C	A	A	1981-2009

2014	Abellán and Mantas (2014)	C D	A	RS-LNMC-CDT-DT-RS-BA	B	B C E	A		-
2014	Bekhet and Eletter (2014)	D	A	LR-RBF	A	C	C	C	2006-2011
2014	García et al. (2014)	C D	B	O	C	F	E	C	2002-2013
2014	Hajek et al. (2014)	B	A	SVM-ANN-DT-NB-LR-RBF-MLP-SMO-OS-SMOTE	B	A	B	A B	2008
2014	Heo and Yang (2014)	C	A	AB-ANN-SVM-DT-ZS	B	C	A	A	2008-2012
2014	Kim and Upneja (2014)	B	A	TD-AB	B	A	B	A	1988-2010
2014	Kou et al. (2014)	A	A	NB-LR-KNN-DT-CART-MCDM	B	C	C	A	2002-2004
2014	Sun et al. (2014)	B C	B	O	E	F	E	D	-
2014	Tsai (2014)	C	A	MLP-LR-CART-SOM-KM	B	B C E F	A	A	-
2014	Tsai et al. (2014)	C	A	MLP-SVM-DT-BOO-BA	B	B C E	A	A	-
2014	Wang et al. (2014)	C	A	LR-NB-DT-ANN-SVM-BA-BOO-MLP-SMO	B	F	A	A	-
2014	Wu et al. (2014)	D	A	LR-DT-ANN	B	F	C	A	-
2014	Yu et al. (2014)	C	A	ELM-LDA-SVM	B	B	A	A	2002-2003
2014	Zhang et al. (2014)	D	A	SVM-FSVM-MCOC	C	A B E	C	D	-
2014	Zhong et al. (2014)	D	A	ELM-SVM-BPNN-ANN-EKM-SLFN	B	A	C	A	-
2015	Chen et al. (2016)	A C	B	O-CSL	C	F	E	D	-
2015	Danenas and Garsva (2015)	C	A	SVM-LR-ANN-RBF	B	A	A	A	1999-2007
2015	Geng et al. (2015)	B	A	ANN-DT-SVM-LR-LDA-MV-DM	B	C	B	A	2001-2008
2015	Jones et al. (2015)	D	A	AB-BOO-RF-ANN-SVM-LR-LDA-QDA-PROBIT	B	A E	C	A B D	1983-2013
2015	Kim et al. (2015)	C	A	AB-GMC-CB	B	C	A	A	2001-2004
2015	Koutanaei et al. (2015)	D	A	ANN-SVM-CART-NB-AB-BA-RF	A	C	C	A B	-
2015	Lessmann et al. (2015)	D	A	QDA-SVM-NB-LR-KNN-ELM-CART-ANN-RF-BOO-BA-KM-O	B	B E	C	A	-
2015	Liang et al. (2015)	B	A	KNN-CART-MLP-SVM	B	B C E	B	A	-
2015	Moro et al. (2015)	E	B	O	A	F	E	D	2002-2013
2016	AghaeiRad et al. (2017)	C D	A	SOM-FNN	B	B C E	A	A	2002-2006
2016	Ala'Raj and Abbod (2016*b*)	D	A	MCS-LR-ANN-SVM-RF-DT-NB-CA	B	B C E	A	A	-
2016	Ala'raj and Abbod (2016*a*)	D	A	GNG-MARS-CA-ANN-SVM-RF-DT-NB	B	A B C E	A	A	-
2016	Chang and Hsu (2018)	B	A	SVM-DEA-IFWTSVM-ANN-RF	B	C	B	A	2012-2014

2016	Du Jardin (2016)	C	A	MDA-LR-ANN-CART-BA-BOO-RS-PBM-SOM	B	B	A	A	2003-2012
2016	Gorzałczany and Rudzinski (2016)	D	A	FRBC	B	B C E	C	A	-
2016	Jones et al. (2017)	C	A	AB-BOO-RF-ANN-SVM-LR-LDA-QDA	B	A	A	A	-
2016	Liang et al. (2016)	C	A	SVM-KNN-NB-CART-MLP-SDA	B	C	A	A B	1999-2009
2016	Sun et al. (2017)	B	A	SVM-AB-TBOO	B	C	B	A	2000-2012
2016	Zieba et al. (2016)	C	A	AB-AC-LDA-MLP-LR-DT-RF-GB	B	B	A	A	2007-2013
2017	Abellán and Castellano (2017)	C D	A	AB-BA-RS-DECORATE-RTF-LR-DT-MLP-SVM-CDT	B	A B C E	A	A	-
2017	Antunes et al. (2017)	C	A	GP-SVM-LR	B	B	A	A	2006-2007
2017	Barboza et al. (2017)	C	A	MDA-LR-ML-SVM-BA-RF-BOO-ANN	B	A	A	A	1985-2013
2017	Bequé and Lessmann (2017)	D	A	ELM-ANN-LR	A	B E	C	A	-
2017	Chou et al. (2017)	C	A	FL-GA-BPNN	B	C	A	A	-
2017	du Jardin (2017)	C	A	MDA-LR-SVM-FNN-DT-ELM-RS-RTF-DECORATE-BA-BOO-AB-COX-SOM	B	B	A	A	1996-2001
2017	García, Marques, Sánchez and Ochoa-Domingues (2017)	C	A	FLD-LDA-SVM-LR	A	C	A	C	-
2017	Jones (2017)	C	A	GBOO	B	A	A	A B D	1987-2013
2017	Liang et al. (2018)	B	A	SVM-KNN-CART-MLP-NB-BA-BOO-MV	B	B C E	B	A	2005-2015
2017	Ribeiro et al. (2019)	A	A	MLP-SVM-RBF-CART-BOO-KNN	B	B	C	A B	2002-2006
2017	Tobback et al. (2017)	C	A	DM-WVRN	B	B	A	A B	2011-2014
2017	Zelenkov et al. (2017)	C	A	KNN-LR-NB-DT-SVM-ADB-RF-BA-QDA-ZS-MLP-GBOO-TSCM	B	B	A	A	-
2018	Cheng et al. (2018)	C	A	LR-GBOO	B	A	A	A B	1992-2012
2018	du Jardin (2018)	C	A	MDA-LR-SVM-FNN-CART-ELM-RS-RTF-BA-BOO-AB-COX-SOM	B	B	A	A	2006-2014
2018	Feng et al. (2018)	D	A	DT-ANN-SVM-NB-KNN-BA	B	B C E	C	A	-
2018	Feng et al. (2019)	D	A	DT-ANN-SVM-NB-GA-BA-BOO	B	B C E	C	A	-
2018	Lin et al. (2019)	C	A	LR-NB-BPNN-DT-SVM-KNN-BA-BOO-GA	B	B C E	A	A	-
2018	Sun et al. (2018)	D	A	DT-SMOTE-BA-OS-US	B	C	C	A	2007-2012
2018	Wang et al. (2018)	B	A	SVM-SMOTE-BA-BOO-RS-OS-US-CSL-ISTRS	B	C	B	A D	2011-2015
2019	García, Marqués and Sánchez (2019)	C D	A	RF-RTF-DECORATE-AB-BOO-BA-RS	B	B C E	A	A	-