MACHINE LEARNING IN BANKRUPTCY PREDICTION – A REVIEW

Claudiu CLEMENT
University of Alexandru Ioan Cuza
Iasi, Romania
office.claudiuclement@gmail.com

Abstract: *There is an increasing interest in machine learning for bankruptcy prediction with more and more researchers contributing to the literature. Although there is a considerable amount of research, the domain does not seem to be aligned and there is still a lot of indecisiveness in terms of what is the best method to be used and on which data. Using Web of Science, Scopus and ScienceDirect databases, a systematic review of 32 texts published between 2016 and 2020 was conducted. This review shows a summary of those papers based on 9 criteria. The criteria identified include source of data, number and type of variables, models used, industry type, and timeline of dataset, sample size, aim and result as well as accuracy of the best performing model used. Overall, it has found that no model performs best on any type of data and that the domain is still away from having a conclusion about what works best and where. This paper contributes towards updating academics and practitioners with the current state of the domain, tools used for bankruptcy prediction lately and their performance.*
Keywords: *Machine learning, Bankruptcy prediction, Liquidation, Parametric modelling, Non-parametric modelling*

Acknowledgement
This work was cofinanced from the European Social Fund through Operational Programme Human Capital 2014-2020, project number POCU/380/6/13/125015 ”Development of entrepreneurial skills for doctoral students and postdoctoral researchers in the field of economic sciences”.

INTRODUCTION

The financial sector is and always was a strong pillar of social well-being and every economy is highly dependent on it. The private sector development is, as well, built on the premises of the financial sector. It can also have an important role in providing individuals and households with monetary means for access to basic needs, such as health and education, consequently impacting poverty reduction (Policy Division Working Paper, 2004). In the last more than 100 years, starting with (Bagehot, 1873) and followed by (Schumpeter, 1934) and (Hicks, 1969) literature on market development and economic growth has been getting a lot of attention making it easy for the importance of them to be understood. Considering these elements, undeniably, there has been a great amount of research from researchers in different areas to facilitate the quality of information available in the financial sector, making financial products available, helping predict financial trends, goal evaluation, asset portfolio management, pricing IPO's, finding optimal capital structure, detecting regularities in security price movements, alleviating crediting risk by predicting default and bankruptcy, etc (Bahrammirzaee, 2010). In this regard, many techniques have been developed. This paper focuses on the advancements and literature background on the methods applied in bankruptcy prediction for studies published between 2016 and 2020. In general, these techniques/methods can be classified in two main categories: parametric (multiple discriminant analysis (MDA), linear discriminant analysis (LDA), canonical discriminant analysis (CDA), logistic regression (LR) and Naïve Bayes (NB)) and non-parametric (artificial neural networks (ANN), support vector machine (SVM), decision trees (DT), k-nearest neighbour (KNN), hazard models, fuzzy models, genetic algorithms (GA) and hybrid models, where multiple models are combined).

Starting with the parametric models, logistic regression and discriminant analysis are some of the most used statistical techniques in empirical studies of economic phenomena. The difference between them comes from the fact that LR requires a logistic distribution. DA is mostly used for categorization or classification tasks where logistic regression is mostly used for obtaining the odds ratios for each categorization variable (Lo, 1986). Naïve Bayes has proved its effectiveness because of its simplicity and tractability, allowing for effective bounds (Choi *et al.*, 2019).

Secondly, non-parametric models, the ones that are recently the most used, don't make any assumption about the distribution of the underlying data as well as the fact that the number of parameters and structure of it is decided by data rather than fixed a-priori. These models are mainly multiple and depend heavily on computer technology for their implementation (Aziz and Dar, 2006). The main advantages of these models come from their ability to learn and adapt, based on the data set, capturing non-linear relationships between variables (Fejér-Király, 2015). In the same time, the weak points come from the lack of explainability, being considered black-box algorithms, they are failing to explain causal relationships between variables (i.e. financial ratios) (Lee and Choi, 2013).

Best papers in the area of bankruptcy prediction, considering number of citations, are (Balcaen and Ooghe, 2006), (Gissel, Giacomino and Akers, 2007) and (Ravi Kumar and Ravi, 2007) which are, in fact, review papers. The first two studies are centred on parametric models while the last one covers non-parametric models as well. (Balcaen and Ooghe, 2006) make a summary of the causes that led bankruptcy prediction studies to evolve. (Gissel, Giacomino and Akers, 2007) have a very important contribution to the literature by summarizing 165 papers published between 1965 and 2006. Their study includes a summarization of the papers very similar to this study, including information such as model type, number of variables used and model accuracy. In their paper, (Ravi Kumar and Ravi, 2007) treat slightly the same time frame, analysing papers published between 1968 and 2005 highlighting the following: the source of the dataset, financial ratios used, country of origin, timeline of study and the comparative performance of the techniques by presenting the accuracy.

This paper is contributing to the literature on bankruptcy prediction by summarizing the most relevant papers published in the last 5 years in the literature using a systematic review approach. The goal is providing academics and practitioners with an overview on what has been written lately by summarizing all papers intro a table including source of data/country of origin, number of variables, type of variables, models used, industry type, timeframe of the dataset used, sample size, results and accuracy of best performing model.

The remainder of this paper presents an overview on the literature review written in Chapter 2 followed by a quick theoretical presentation over the most used methods in the papers studied in Chapter 3. Chapter 4 includes the presentation of the papers studied on the premises presented in the previous paragraph. Finally, Chapter 5 concludes and provides some suggestions for future research in bankruptcy prediction.

EARLIER REVIEWS

(Balcaen and Ooghe, 2006) created a very comprehensive review paper by analysing 35 years of literature in bankruptcy prediction. The paper analyses extensively on the application of univariate analysis, risk index models, multiple discriminant analysis and conditional probability models. On the premises that, at the moment of doing the study, there were no clear and comprehensive analysis of problems related to these methods, authors treat each problem issue accordingly and discuss each of them. There are three main problems identified by the authors in their study:
The classical paradigm (i.e. the unclear definition of failure, non-stationarity and data instability, sampling bias and the choice of optimisation criteria);
The neglect of time dimension of failure (the choice of when to observe a firm may introduce a selection bias in the resulting model (Shumway *et al.*, 1999));
Problems related to the application focus (due to commercial pressure, most of the models have been developed without a holistic understanding of the reason of company failure).
(Gissel, Giacomino and Akers, 2007) have, as well, a broad study on the subject, examining 165 papers published between 1965 and 2006. This paper traces the literature on bankruptcy prediction, from the times when simple ratio analysis was used to 2006 when the usage of intelligent techniques already picked up. Authors organize the models identified in their studied papers in three categories based on the industry source of data:
General (a mix of industries); Banking; Industry-specific models.

In addition, the split between parametric and non-parametric models adopted in this paper, has inspired us to do the same in our review analysis. In their paper, it is concluded that MDA and NN are the most promising methods for bankruptcy prediction models together with the fact that in their analysis, there has not been found any correlation between the number of features and model accuracy, models with just two features being just as capable in terms of accuracy as models with 20+ features.

Another significant study in the review literature of bankruptcy prediction models is (Ravi Kumar and Ravi, 2007), their research covering papers published between 1968 and 2005. Authors categorize the papers in 8 families of techniques such as: statistical techniques, neural networks, case-based reasoning, decision trees, operational research, evolutionary approaches, rough set based techniques, other techniques including fuzzy logic, support vector machine and isotonic separation and soft computing including hybrid models based on all the previously-mentioned methods. For all papers included in the study the authors highlight the source of data sets, financial ratios used, country of origin, period of study and the prediction accuracy wherever possible.

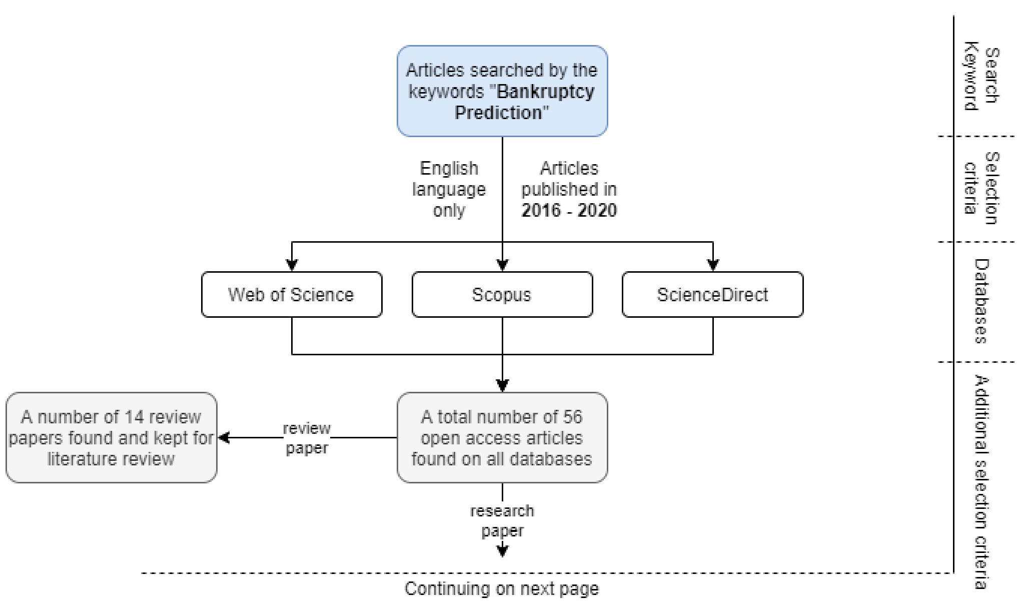
In terms of more recent review papers, worth mentioning are (Alaka *et al.*, 2018) that analysed 49 research papers published between 2010 and 2015, (Prusak, 2018) with a focus on Eastern European Bloc focused papers between Q4 2016 and Q3 2017, (Altman, 2018) making a follow-up and summarizing the 50 years history of his z-Score model. On the same note, (Qu *et al.*, 2019) with a short conference paper presenting an general overview on methods used in bankruptcy prediction, (Ptak-Chmielewska, 2019) with a focus on the addition of non-financial factors into the models, (Gruszczyński, 2019) having an overview from the unbalance sampling and sample bias perspective, (Leo, Sharma and Maddulety, 2019) focusing on banking bankruptcy risk prediction. Very important as well, (Shi and Li, 2019a) analysed papers published on bankruptcy prediction models from 1968 to 2007, same authors in (Shi and Li, 2019b) publish a bibliometric review addressing the research trends in the area of bankruptcy prediction.

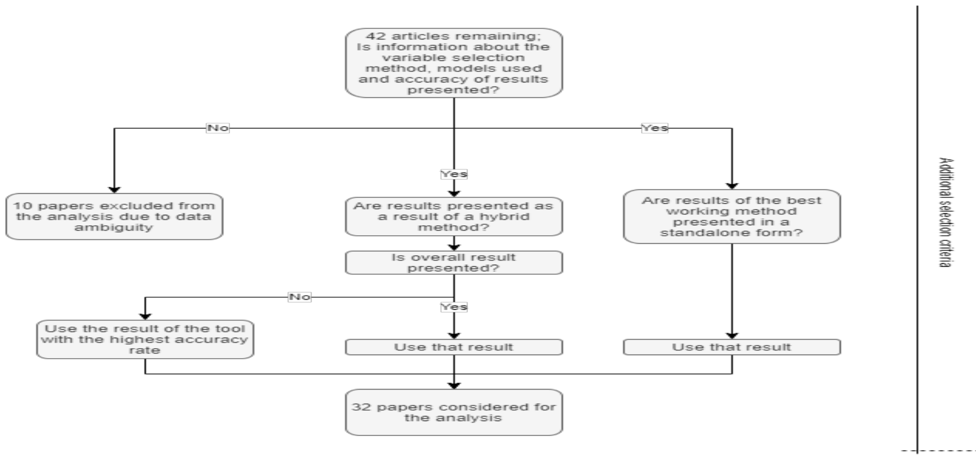
What is important to be noted, after briefing the review literature on the topic of bankruptcy prediction, is the fit of this paper in the sense of covering a period that has not been covered, at the time of writing this paper, by previous studies.

REVIEW METHODOLOGY

As mentioned earlier, this review is conducted in two broad categories: (i) parametric models and (ii) non-parametric models. Among parametric models, the methods covered are: multiple discriminant analysis (MDA), linear discriminant analysis (LDA), canonical discriminant analysis (CDA), logistic regression (LR) and naïve bayes (NB). The non-parametric, or so-called intelligent models covered in this study belong to artificial neural networks (ANN), support vector machine (SVM), decision trees (DT), k-nearest neighbour (KNN), hazard models, fuzzy models, genetic algorithms (GA) and hybrid models, where multiple models are combined. Papers are analysed chronologically. The most important dimension of the present review is the type of model applied. The review includes other dimensions also such as source of data, number of variables used in the model, type of variables (financial/relational data/textual), industry type, timeline of dataset, sample size (bankrupt vs non-bankrupt where available), accuracy of the best performing model. Further, the review focused on papers published in academic journals or conference proceedings and available in the public databases Web of Science, Scopus, ScienceDirect. More on the selection framework in fig.1:

Figure 1 Articles selection framework





Overview of Intelligent Techniques

Table 1. Parametric vs Non-parametric models

PARAMETRIC	NON-PARAMETRIC
It uses a fixed number of parameters to build the model	It uses flexible number of parameters to build the model
Considers strong assumption about the data	Considers fewer assumptions about the data
Computationally faster	Computationally slower
Require lesser data	Require more data

Source: Park, Kim and Lee, 2014

Parametric models

A wide variety of papers have studied the application of parametric models in the area of bankruptcy prediction up until the 90'ies when more complex and computationally intensive models started to be applied. The list of parametric models found in the papers studies together with a short description of the model can be found on the tab. 2 below.

Table 2 Parametric models identified in the selected papers and short description

PARAMETRIC MODELS	DESCRIPTION
Canonical discriminant analysis (CDA) Discriminant analysis (DA)	Determines how to best separate or discriminate between two or more groups of data, given their quantitative measurements of several variables of these groups (Cruz-Castillo *et al.*, 1994). Used to classify observations when the dependent variable is categorical and the independent variables is interval.
Logistic regression (LR) Cost sensitive variation of logistic regression (CLR)	LR uses the log-ratio to assign a company to either bankrupt or non-bankrupt class (Veganzones and Séverin, 2018); A variation of logistic regression which has been used for addressing class imbalance problems, mostly used in credit scoring (Zhang *et al.*, 2020).

Linear discriminant analysis (LDA) Multiple discriminant analysis (MDA)	It assumes that class-conditional densities follow Gaussian distributions and that they also have a covariance matrix (Veganzones and Séverin, 2018). It is used to determine the class membership of samples from a group of predictors by finding linear combinations of the variables that maximize the difference between classes (Brown, 1998).
Naïve Bayes (NB)	NB classification uses the probabilistic inference to assign a company to a class, given observed features, computing the probability of the decision variable (Choi *et al.*, 2019).

Source: mentioned on each method

Non-parametric models

With the growing advancements in computing power and the increasing size of samples studied, non-parametric models took-off. In the majority of previous studies, non-parametric classifiers outperform the performance measured by accuracy of their parametric counterparts, only for the case of small samples size it can be the other way around (De Andrés, Landajo and Lorca, 2005).

Table 3 Non-parametric models identified in the selected papers and short description

NON-PARAMETRIC MODELS	DESCRIPTION
AdaBoost AdaCost	Adaptive boosting (AdaBoost) is one of the machine learning algorithms designed by (Freund and Schapire, 1996). AdaBoost works as an algorithms enhancer, combined with weak classifiers to build a learning algorithm with stronger classifiers; A misclassification cost-sensitive boosting model.
Case base reasoning (CBR)	Decision tree that learns from examples using the Euclidean distance and k-nearest neighbor method (Ravi Kumar and Ravi, 2007).
Extreme learning machine (ELM)	Simple learning algorithm where the hidden layer does not need to be iteratively tuned and the training error and the norm of the weights are minimized (Yu *et al.*, 2014).
Fuzzy chance constrained least squares twin support vector machine (FCC-LSTSVM)	The chance constrained algorithm ensures the minimum misclassification for uncertain data (Song, Cao and Zhang, 2018).
Fuzzy-set qualitative comparative analysis (fsQCA)	It uses combinatorial logic, fuzzy set theory and Boolean minimization to highlight what combinations of case characteristics are sufficient to produce an outcome (Boratyńska and Grzegorzewska, 2018).
Feed-forward neural network (FNN)	Can be seen as a way to parametrize a fairly non-linear function proved to be extremely flexible in approximating smooth functions (De Andrés *et al.*, 2011);

General regression neural networks (GRNN) Multilayer neural network (MNN) Multilayer perceptron (MLP) Recurrent neural network (RNN)	A neural network with the number of neurons in the hidden layer consistent with the sample size (Song, Cao and Zhang, 2018); A neural network in which the signal flow is only in one direction (Korol, 2019); The most used class of artificial neural networks, it uses a set of input-output pairs to learn the model correlations between those groups (Tsai, Hsu and Yen, 2014). Mostly used for time series data analysis, it uses internal memory to process the incoming inputs (Ozbayoglu, Gudelek and Sezer, 2020).
Gaussian processes	Each class prediction comes in the form of a probability allowing explanatory power on how certain the model is about the state of bankruptcy (Antunes, Ribeiro and Pereira, 2017).
Classification and regression tree (CART) J48 CJ48	A decision tree algorithm developed by (Breiman L. *et al.*, 1984) that works by choosing the best separation of the population (parental node) in two sub-populations (child nodes) (Durica, Frnda and Svabova, 2019); Open-source implementation of the C4.5 algorithm; J48 optimized for cost rather than error.
Decision rule inducer (JRIP) CJRIP	It works by treating all the examples of a particular decision in the training data as a class, and finding a set of rules that cover all the members of that class. Afterwards it proceeds to the next class and does the same, repeating this until all classes have been covered (Parsania, Jani and Bhalodiya, 2014); Cost optimized JRIP.
k-Nearest neighbor (KNN)	It determines the probability of default by the proximity of cases next to each other being calculated as default cases divided by overall nearest neighbors (Kruppa *et al.*, 2013).
Principal component analysis (PCA)	It is used for dimensionality reduction while keeping much of the data set variation (Tsai, 2009);
Radial basis function network (RBFN)	Similar to MLP but in RBFN each node has its own radial basis function, such as a Gaussian function instead of the logistic function of the former (Tseng and Hu, 2010).
Random forest (RF)	A relatively new method that combines trees grown on bootstrap samples of data and a random subset of bagging of predictor variables (Yeh, Chi and Lin, 2014).
Support vector machine (SVM) CSVM Support vector regression (SVR)	Works by using statistical learning theory to perform classification and regression tasks (Ravi Kumar and Ravi, 2007); Cost sensitive support vector machine; Different than than the SVM in terms of the fact that it performs regression where SVM performs classification.

Weighted-vote relational neighbor (wvRN)	A classifier using the network structure to calculate a class probability as a weighted average of its j neighbors' probability scores (Tobback et al., 2017).
Extreme gradient boosting (XGB) XGBE EXGB	An optimized, very performant, distributed gradient boosting library; Only the last tree of XGB; Ensemble of booted trees trained with XGBE EXGB.

Source: mentioned on each method

FINDINGS

This section summarizes the reviewed articles by presenting the 9 criterias such as models used, industry type, time frame of the dataset, sample size and sample split where available, short description of the aim of the papers and results and finally the accuracy of the best performing model (tab.4).

Table 4. Summary of reviewed articles

Reference	Source of data (country of origin)	Number of variables	Type of variables	Models used	Industry type	Timeline of data set	Sample size	Aim and results	Accuracy (%)
(Liang *et al.*, 2016)	Taiwan	180	Financial ratios and corporate governance indicators	SVM, KNN, CART, MLP, NB	Mixed	1999-2009	239 bankrupt and 239 non-bankrupt	Authors used a model based on a combination of financial ratios and corporate governance indicators that proved to perform best, hence stepwise discriminant analysis (SDA) + support vector machine (SVM).	83.6
(Zięba, Tomczak and Tomczak, 2016)	Poland	64	Financial ratios	LDA, MLP, JRip, CJRip, J48, CJ48, LR, CLR, AB, AC, SVM, CSVM, RF, XGB, XGBE, EXGB	Mixed	2000-2013	700 bankrupt and 10000 non-bankrupt	Authors developed a model using Extreme Gradient Boosting and it showed results better than all methods compared. Also, they introduced a novel approach using synthetic features/variables.	95.9
(Pal *et al.*, 2016)	France	35	Financial ratios	CART, DA, LR, NN	Mixed	200	8660 bankrupt and 8660	In their paper authors proved that ensemble methods seem to capture	91

						2-2012	non-bankrupt	some variation within the decision space that individual models do not.	.2
(Sartori, Mazzucchelli and Gregorio, 2016)	Italy	6	Financial ratios	CBR, CRePERIE	Mixed	2012-2013	807 bankrupt and 11637 non-bankrupt	Authors use this new method Case Retrieval Platform Extended to RevIsE that not only proves good results in terms of accuracy but can be used because of its explainability power.	86
(Du Jardin, 2016)	USA	136	Financial ratios	SVM, CBR-SVM	Manufacturing and service	2012-2013	10 bankrupt and 188 non-bankrupt	A novel approach towards having a dynamic discriminating hyperplane matched with expert ratings (Equity Summary Score) was developed in this paper.	90
(Alaminos, Del Castillo and Fernandez, 2016)	Worldwide	12	Financial ratios	LR	Mixed (non-financial)	1990-2013	220 bankrupt and 220 non-bankrupt	Authors show that a global model proves to be more effective than a regional one.	89
(Antunes, Ribeiro and Pereira, 2017)	France	30	Financial ratios	GP, SVM, LR	Mixed	2005-2007	Multiple datasets: 1334 companies (50:50), 2000 companies (30:70) and 2000 companies (20:80)	Authors work on a visualization centric approach with three databases with different class imbalances.	92
(Barboza, Kimura and Altman, 2017)	USA	11	Financial ratios	SVM, RF, NN, LR, MDA, Bagging, Boosting	Mixed	1985-2013	612 bankrupt and 13449 non-bankrupt	Authors re-proved that the accuracy of modern machine learning methods is better than that of classical methods.	87

(du Jardin, 2017)	France	32	Financial ratios	DA, LR, DT, Cox Model, SVM, Bagging, Boosting, Random Subspace, Rotation Forest	Mixed	1997-2003	1920 bankrupt and 95910 non-bankrupt	Authors demonstrate that the accuracy of any model can be improved when the horizon of analysis exceeds two years.	82
(Wang, 2017)	N.A.	6	Financial ratios	SVM, NN, Autoencoder, LR, GA, Inductive learning	Mixed	N. A.	107 bankrupt and 143 non-bankrupt	Author shows that neural network with dropout shows best results in comparison with classical methods on the database studied.	99
(Tobback *et al.*, 2017)	Belgium/ UK	6	Relational data between companies and financial ratios	wvRN, SVM	Mixed	2011-2014	240000 bankrupt and 2200000 non-bankrupt	Authors report the potentially unused benefits of relational data in bankruptcy prediction models.	84
(Fito, Plana-Erta and Llobet, 2018)	Spain	5	Financial ratios	z-Score	Mixed	2005-2015	450 bankrupt companies	Authors analyze the difference in results between Altman z-Score and their score showing that on the dataset studied the later score is more effective.	95.8
(Song, Cao and Zhang, 2018)	China	27	Financial ratios	NN, RBF, GRNN, SVR, SVM, FCC-LSTSVm	Mixed	N. A.	398 bankrupt companies and 398 non-bankrupt companies	Authors demonstrate that effectiveness of methods depends on the type of industry.	98
(Nyitrai and Miklós, 2018)	Hungary	20	Financial ratios	DA, LR, DT, NN	Mixed	2001-2016	1468 bankrupt and 1528 non-bankrupt	Authors prove that decision trees are robust methods when faced with outliers where linear models and neural networks are sensitive.	87
(Carmona, Clim	USA	30	Financial ratios	XGB, LR, RF	Banking	2001-	78 bankrupt and 78 non-bankrupt	Authors show that XGB has a higher predictive power of bankruptcy for	98

ent and Momparler, 2018)						2015		the banking sector to the other models tested.	
(Le and Viviani, 2018)	USA	31	Financial ratios	DA, LR, ANN, SVM, KNN	Banking	2011-2016	1438 bankrupt and 1562 non-bankrupt	Authors show that KNN and ANN demonstrate their good ability of predicting bankruptcy on the dataset used while the other methods cannot.	82
(Obradović *et al.*, 2018)	Serbia	5	Financial ratios	LR	Mixed	2010-2011	43 bankrupt and 43 non-bankrupt	Authors show that the model of Logistic Regression shows promising results on predicting bankruptcy on the Serbian dataset.	88.4
(Gogas, Papadimitriou and Agrapetidou, 2018)	USA	36	Financial ratios	SVM	Banking	2007-2013	481 bankrupt and 962 non-bankrupt	The SVM model used by the authors outperforms the well-established Ohlson's score.	99.2
(du Jardin, 2018)	France	15	Financial ratios	DA, LR, DT, Cox, SVM, FNN, ELM	Mixed	2006-2014	120 bankrupt and 6000 non-bankrupt	The findings reinforce the idea that the model accuracy does not solely rely on data mining techniques but also on the way one will use some knowledge about the bankruptcy phenomenon during modeling process.	82.9
(Mai *et al.*, 2018)	USA	36	Textual disclosure and Financial Ratios	CNN, NN, LR, SVM	Mixed	1994-2014	477 bankrupt and 11350 non-bankrupt	Authors combine numerical variables with textual disclosures and show the first large-sample evidence of the predictive power of textual disclosures.	85
(Boratyńska and Grzegorze	Poland	6	Financial ratios	fsQCA, MDA, LR	Agribusiness	1996-2	14 bankrupt and 14 non-bankrupt	The study shows that fsQCA proves to be efficient in predicting bankruptcy for the agribusiness sector.	92.9

wska, 2018)						007			
(Veganzones and Séverin, 2018)	France	50	Financial ratios	LDA, LR, NN, SVM, RF	Mixed	2013-2014	2400 bankrupt and 6600 non-bankrupt	In this study authors show that prediction methods reward the classification of the majority class to the detriment of the minority class in imbalanced training datasets.	92.8
(Chang, 2019)	Poland	64	Financial ratios	SVM, RF	Mixed	2007-2013	2091 bankrupt and 45405 non-bankrupt	In a modest study authors show that the random forest method shows the highest results although the accuracy is only a bit above 70%.	70
(Lukason and Andresson, 2019)	Estonia	10 financial and 24 tax arrears	Tax arrears and financial ratios	LR, MLP	Mixed	2013-2017	512 bankrupt and 4003 non-bankrupt	The study shows that the dynamic usage of only a certain type of payment defaults (tax arrears) can substantially outrun the accuracies of financial ratio-based models.	93.4
(Affes and Hentati-Kaffel, 2019)	USA	10	Financial ratios	LR, CDA, PCA	Banking	2008-2013	410 bankrupt and 5805 non-bankrupt	The study shows that LR and CDA can predict banks failure with great accuracy.	95.6
(Charalambakis and Garrett, 2019)	Greece	7	Financial ratios	LR	Mixed	2003-2011	1770 bankrupt and 29116 non-bankrupt	The authors create 5 different logit models to test their efficiency in predicting bankruptcy on their Greek dataset one of them showing great results both over short and long run.	91.9
(Agrawal and Maheshwari, 2019)	India	1	Financial ratios	LR, MDA	Mixed	2001-2012	135 bankrupt and 135 non-bankrupt	The study used industry beta to assess its impact on default probability by regressing it with stock returns. The result shows industry beta being statistically significant in predicting default.	75.6

(Korol, 2019)	Europe	20	Financial ratios	MNN, RNN, Fuzzy Sets, DT	Mixed	2004-2017	300 bankrupt and 300 non-bankrupt	The study shows the superiority of fuzzy sets over the other developed models mostly closer the announcement of bankruptcy showing promising results on the long run as well.	96.2
(Durica, Frnda and Svabova, 2019)	Poland	37	Financial ratios	CART	Mixed	2016-2017	2698 bankrupt and 26210 non-bankrupt	Using a decision tree algorithm authors manage to create a model with great accuracy mainly useful for predicting the financial difficulties of Polish companies.	93.6
(Habachi and Benbachir, 2019)	Morocco	22	Financial ratios	LDA, Bayesian	Mixed	2017-2018	114 bankrupt and 1333 non-bankrupt	Authors proposed a quite effective method of rating model using LDA using a dataset of SMEs from a Moroccan bank.	93.7
(Hosaka, 2019)	Japan	263	Financial ratios	CNN, DT, LDA, SVM, MLP, AB, z-score	Mixed	2002-2016	102 bankrupt and 2062 non-bankrupt	The authors used a CNN based on GoogleNet that proved better results that that of comparable classical models.	88
(Muñoz-Izquierdo *et al.*, 2020)	Spain	20	Financial ratios	LR	Mixed	2004-2014	404 bankrupt and 404 non-bankrupt	Authors show that a mix of financial and auditing register a considerably higher accuracy.	87

The final goal of any bankruptcy prediction model development is to have a high accuracy of prediction. Tab.4 in its last column presents the accuracy of the best performing model out of the ones tested by the authors or the accuracy of the hybrid model implemented. Values range from 70% to 99.22% with a mean value of 87.12%, which is in line with previous findings of (Alaka *et al.*, 2018), studied papers not showing incremental increase in their results but rather on a steady trend in terms of accuracy.

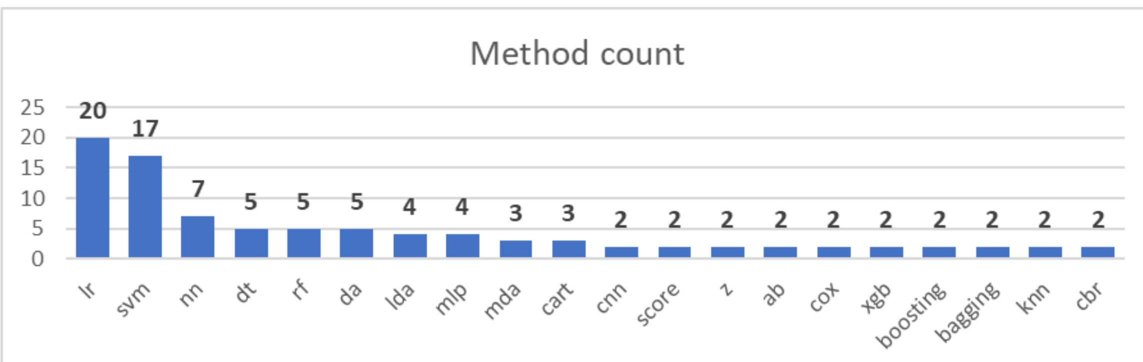
As (du Jardin, 2017) demonstrated, the longer the period of analysis is the better the accuracy of the model becomes. In the papers studied, the period of analysis ranges from as small as 1 year to 28 years, with an average of 8.7 years showing that most papers have a consistent time frame at their disposal at the time of analysis.

Looking at the size of the dataset considered by the papers studied, 19/32 (59%) of the papers did not have equal samples for bankrupt and non-bankrupt companies, this problem leading to the well-known oversampling problem where when faced with an imbalance between labels the algorithms tend to predict the oversampled label/class. As (Zhou, 2013) showed in his study, the accuracy of the models are highly dependent on the number of bankruptcies in the model thus authors should focus on managing the oversample problem before working on the algorithms themselves.

In terms of the sources of data, there seems to be no focus on a specific geographic region, which is great news for the domain as the algorithms are, tested on different databases hence no specialization or improvement of algorithms on a single area.

Starting with (Agrawal and Maheshwari, 2019) that only used 1 variable for predicting bankruptcy prediction (industry beta) and continuing with (Hosaka, 2019) that had 263 financial ratios, it is clear that there is a very wide coverage in terms of researchers preference for the numbers of variables used. There does not seem to be a clear focus on the type of industry either, 25/32 (78,12%) papers analysing companies from mixed industries. Although there seems to be a wide variety in terms of the composition elements of the papers published, not the same thing can be said about the type of variables used, 28/32 (87,5%) of papers used financial ratios as the predicting variables showing a still biased view, in a form or another, towards considering financial ratios as being the only way to go in predicting bankruptcy.

Figure 2. Count of methods present in reviewed papers



Source: Own calculations

Now, the key part of this paper and what draws the most attention is the specific models used by researchers in the literature. In the Fig.2 in previous page, it is extremely interesting the be observed that the first two methods are logistic regression and support vector machine which relates with previous literature, logistic regression being considered the key reference algorithm.

Table 5. Parametric vs Non-parametric models

METHODS	METHODS TOGETHER COUNT
da lr	5
svm rf	3
lr svm	3

lr dt	3

Source: Own calculations.

In tab.5 it can be observed that most frequently used together in the papers studied are discriminant analysis and logistic regression. This is as well in line with previous literature, discriminant analysis and logistic regression being usually tested against each other to find the best parametric model. In top 4 pairs presented above, logistic regression is present in 3 which, again, emphasises on the power of this simple algorithm, having it always in comparison with the other, later developed algorithms.

CONCLUSION

The aim of this paper was to present the summary of latest literature on bankruptcy prediction that would help practitioners and academia understand what the current trend and research focus is and what the results are. We showed that the bankruptcy prediction models continue to evolve with even broader perspectives then before and different strategies in developing the models. This study used a systematic review to highlight key elements as source of data, number and type of variables, models used, industry type, timeline of dataset, sample size, aim and result as well as accuracy. Overall, it can be concluded that there is no tool that is generally better and that the accuracy depends more on the tweaking of the algorithm based on the sample used and its properties rather than pre-defined selection based on previous studies. This idea is aligned with (Alaka *et al.*, 2018) who provided researchers with a tool selection framework but mentioned that caution should be used and the best results are achieved by trial-and-error. Future studies should consider analysing more characteristics in the published papers in the literature together with a closer look on hybrid models understanding. Although there is not currently a model that *fits all* current research shows that it might be the best way to go. Finally, as most of studies analysed in this paper considered financial ratios for their analysis, there is plenty of room for research in including more qualitative variables into the models.

References

1. Affes, Z. and Hentati-Kaffel, R. (2019) ‘Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis’, *Computational Economics*. Springer New York LLC, 54(1), pp. 199–244. doi: 10.1007/s10614-017-9698-0.
2. Agrawal, K. and Maheshwari, Y. (2019) ‘Efficacy of industry factors for corporate default prediction’, *IIMB Management Review*. Elsevier Ltd, 31(1), pp. 71–77. doi: 10.1016/j.iimb.2018.08.007.
3. Alaka, H. A. *et al.* (2018) ‘Systematic review of bankruptcy prediction models: Towards a framework for tool selection’, *Expert Systems with Applications*. Elsevier Ltd, 94, pp. 164–184. doi: 10.1016/j.eswa.2017.10.040.
4. Alaminos, D., Del Castillo, A. and Fernandez, M. A. (2016) ‘A global model for bankruptcy prediction’, *PLoS ONE*. Public Library of Science, 11(11). doi: 10.1371/journal.pone.0166693.
5. Altman, E. I. (2018) ‘Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models?’, *International Journal of Financial Studies*. MDPI AG, 6(3), p. 70. doi: 10.3390/ijfs6030070.
6. De Andrés, J. *et al.* (2011) ‘Bankruptcy forecasting: A hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS)’, *Expert Systems with Applications*, 38(3), pp. 1866–1875. doi: 10.1016/j.eswa.2010.07.117.

7. 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), pp. 518–542. doi: 10.1016/j.ejor.2004.02.018.
8. Antunes, F., Ribeiro, B. and Pereira, F. (2017) 'Probabilistic modeling and visualization for bankruptcy prediction', *Applied Soft Computing Journal*. Elsevier B.V., 60, pp. 831–843. doi: 10.1016/j.asoc.2017.06.043.
9. Aziz, M. A. and Dar, H. A. (2006) 'Predicting corporate bankruptcy: Where we stand?', *Corporate Governance*, pp. 18–33. doi: 10.1108/14720700610649436.
10. Bagehot, W. (1873) 'Lombard Street : A Description of the Money Market', pp. 1–287.
11. 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), pp. 1165–1195. doi: 10.1007/s00521-010-0362-z.
12. Balcaen, S. and Ooghe, H. (2006) '35 years of studies on business failure : an overview of the classic statistical methodologies and their related problems', 38, pp. 63–93. doi: 10.1016/j.bar.2005.09.001.
13. Barboza, F., Kimura, H. and Altman, E. (2017) 'Machine learning models and bankruptcy prediction', *Expert Systems with Applications*. Elsevier Ltd, 83, pp. 405–417. doi: 10.1016/j.eswa.2017.04.006.
14. Boratyńska, K. and Grzegorzewska, E. (2018) 'Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches', *Journal of Business Research*, 89(January), pp. 175–181. doi: 10.1016/j.jbusres.2018.01.028.
15. Breiman L. *et al.* (1984) *Classification and Regression Trees, Wadsworth and Brooks*. Monterey, CA.
16. Brown, C. E. (1998) 'Multiple Discriminant Analysis', in *Applied Multivariate Statistics in Geohydrology and Related Sciences*. Springer Berlin Heidelberg, pp. 115–128. doi: 10.1007/978-3-642-80328-4_10.
17. Affes, Z. and Hentati-Kaffel, R. (2019) 'Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis', *Computational Economics*. Springer New York LLC, 54(1), pp. 199–244. doi: 10.1007/s10614-017-9698-0.
18. Agrawal, K. and Maheshwari, Y. (2019) 'Efficacy of industry factors for corporate default prediction', *IIMB Management Review*. Elsevier Ltd, 31(1), pp. 71–77. doi: 10.1016/j.iimb.2018.08.007.
19. Alaka, H. A. *et al.* (2018) 'Systematic review of bankruptcy prediction models: Towards a framework for tool selection', *Expert Systems with Applications*. Elsevier Ltd, 94, pp. 164–184. doi: 10.1016/j.eswa.2017.10.040.
20. Alaminos, D., Del Castillo, A. and Fernandez, M. A. (2016) 'A global model for bankruptcy prediction', *PLoS ONE*. Public Library of Science, 11(11). doi: 10.1371/journal.pone.0166693.
21. Altman, E. I. (2018) 'Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models?', *International Journal of Financial Studies*. MDPI AG, 6(3), p. 70. doi: 10.3390/ijfs6030070.
22. De Andrés, J. *et al.* (2011) 'Bankruptcy forecasting: A hybrid approach using fuzzy c-means clustering and multivariate adaptive regression splines (MARS)', *Expert Systems with Applications*, 38(3), pp. 1866–1875. doi: 10.1016/j.eswa.2010.07.117.
23. 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), pp. 518–542. doi: 10.1016/j.ejor.2004.02.018.
24. Antunes, F., Ribeiro, B. and Pereira, F. (2017) 'Probabilistic modeling and visualization for bankruptcy prediction', *Applied Soft Computing Journal*. Elsevier B.V., 60, pp. 831–843. doi: 10.1016/j.asoc.2017.06.043.
25. Aziz, M. A. and Dar, H. A. (2006) 'Predicting corporate bankruptcy: Where we stand?', *Corporate Governance*, pp. 18–33. doi: 10.1108/14720700610649436.
26. Bagehot, W. (1873) 'Lombard Street : A Description of the Money Market', pp. 1–287.
27. 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), pp. 1165–1195. doi: 10.1007/s00521-010-0362-z.
28. Balcaen, S. and Ooghe, H. (2006) '35 years of studies on business failure : an overview of the classic statistical methodologies and their related problems', 38, pp. 63–93. doi: 10.1016/j.bar.2005.09.001.

29. Barboza, F., Kimura, H. and Altman, E. (2017) 'Machine learning models and bankruptcy prediction', *Expert Systems with Applications*. Elsevier Ltd, 83, pp. 405–417. doi: 10.1016/j.eswa.2017.04.006.
30. Boratyńska, K. and Grzegorzewska, E. (2018) 'Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches', *Journal of Business Research*, 89(January), pp. 175–181. doi: 10.1016/j.jbusres.2018.01.028.
31. Breiman L. *et al.* (1984) *Classification and Regression Trees*, *Wadsworth and Brooks*. Monterey, CA.
32. Brown, C. E. (1998) 'Multiple Discriminant Analysis', in *Applied Multivariate Statistics in Geohydrology and Related Sciences*. Springer Berlin Heidelberg, pp. 115–128. doi: 10.1007/978-3-642-80328-4_10.
33. Carmona, P., Climent, F. and Momparler, A. (2018) 'Predicting failure in the U.S. banking sector: An extreme gradient boosting approach', *International Review of Economics and Finance*. Elsevier Ltd, (September 2017), pp. 1–20. doi: 10.1016/j.iref.2018.03.008.
34. Chang, H. (2019) 'The application of machine learning models in company bankruptcy prediction', in *ACM International Conference Proceeding Series*. Association for Computing Machinery, pp. 199–203. doi: 10.1145/3374549.3374550.
35. Charalambakis, E. C. and Garrett, I. (2019) 'On corporate financial distress prediction: What can we learn from private firms in a developing economy? Evidence from Greece', *Review of Quantitative Finance and Accounting*. Springer New York LLC, 52(2), pp. 467–491. doi: 10.1007/s11156-018-0716-7.
36. Choi, Y. *et al.* (2019) 'Learning Fair Naive Bayes Classifiers by Discovering and Eliminating Discrimination Patterns'. Available at: http://arxiv.org/abs/1906.03843 (Accessed: 29 May 2020).
37. Cruz-Castillo, J. G. *et al.* (1994) 'Applications of canonical discriminant analysis in horticultural research', *HortScience*, 29(10), pp. 1115–1119. doi: 10.21273/hortsci.29.10.1115.
38. Durica, M., Frnda, J. and Svabova, L. (2019) 'Decision tree based model of business failure prediction for Polish companies', *Oeconomia Copernicana*. Instytut Badan Gospodarczych / Institute of Economic Research, 10(3), pp. 453–469. doi: 10.24136/oc.2019.022.
39. Fejér-Király, G. (2015) 'Bankruptcy Prediction: A Survey on Evolution, Critiques, and Solutions', 3, pp. 93–108. doi: 10.1515/auseb-2015-0006.
40. Fito, M. A., Plana-Erta, D. and Llobet, J. (2018) 'Usefulness of Z scoring models in the early detection of financial problems in bankrupt Spanish companies', *Intangible Capital*. OmniaScience, 14(1), pp. 162–170. doi: 10.3926/ic.1108.
41. Freund, Y. and Schapire, R. (1996) 'A decision-Theoretic Generalization of On-Line Learning and an Application to Boosting', *Journal of computer and System Sciences*, 55, pp. 119–139.
42. Gissel, J. L., Giacomino, D. and Akers, M. D. (2007) *A Review of Bankruptcy Prediction Studies: 1930-Present*, *Journal of Financial Education*. Publisher Link.
43. Gogas, P., Papadimitriou, T. and Agrapetidou, A. (2018) 'Forecasting bank failures and stress testing: A machine learning approach', *International Journal of Forecasting*. Elsevier B.V., 34(3), pp. 440–455. doi: 10.1016/j.ijforecast.2018.01.009.
44. Gruszczyński, M. (2019) 'On unbalanced sampling in bankruptcy prediction', *International Journal of Financial Studies*. MDPI Multidisciplinary Digital Publishing Institute, 7(2). doi: 10.3390/ijfs7020028.
45. Habachi, M. and Benbachir, S. (2019) 'Combination of linear discriminant analysis and expert opinion for the construction of credit rating models: The case of SMEs', *Cogent Business and Management*. Cogent OA, 6(1). doi: 10.1080/23311975.2019.1685926.
46. Hicks, J. (1969) *A Theory of Economic History - John Hicks, John R. Hicks*. Oxford University Press. Available at: https://books.google.ro/books/about/A_Theory_of_Economic_History.html?id=u5A6AAAAIAAJ&redir_esc=y (Accessed: 24 May 2020).
47. Hosaka, T. (2019) 'Bankruptcy prediction using imaged financial ratios and convolutional neural networks', *Expert Systems with Applications*. Elsevier Ltd, 117, pp. 287–299. doi: 10.1016/j.eswa.2018.09.039.
48. du Jardin, P. (2017) 'Dynamics of firm financial evolution and bankruptcy prediction', *Expert Systems with Applications*. Elsevier Ltd, 75, pp. 25–43. doi: 10.1016/j.eswa.2017.01.016.
49. du Jardin, P. (2018) 'Failure pattern-based ensembles applied to bankruptcy forecasting', *Decision Support Systems*. Elsevier B.V., 107, pp. 64–77. doi: 10.1016/j.dss.2018.01.003.

50. Du Jardin, P. (2016) ‘A two-stage classification technique for bankruptcy prediction’, *European Journal of Operational Research*. Elsevier B.V., 254(1), pp. 236–252. doi: 10.1016/j.ejor.2016.03.008.
51. Korol, T. (2019) ‘Dynamic Bankruptcy Prediction Models for European Enterprises’, *Journal of Risk and Financial Management*. MDPI AG, 12(4), p. 185. doi: 10.3390/jrfm12040185.
52. Kruppa, J. *et al.* (2013) ‘Consumer credit risk: Individual probability estimates using machine learning’, *Expert Systems with Applications*. Elsevier Ltd, 40(13), pp. 5125–5131. doi: 10.1016/j.eswa.2013.03.019.
53. Le, H. H. and Viviani, J. L. (2018) ‘Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios’, *Research in International Business and Finance*. Elsevier, 44(February 2017), pp. 16–25. doi: 10.1016/j.ribaf.2017.07.104.
54. Lee, S. and Choi, W. S. (2013) ‘A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis’, *Expert Systems with Applications*, 40(8), pp. 2941–2946. doi: 10.1016/j.eswa.2012.12.009.
55. Leo, M., Sharma, S. and Maddulety, K. (2019) ‘Machine learning in banking risk management: A literature review’, *Risks*. MDPI AG, 7(1). doi: 10.3390/risks7010029.
56. Liang, D. *et al.* (2016) ‘Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study’, *European Journal of Operational Research*. Elsevier B.V., 252(2), pp. 561–572. doi: 10.1016/j.ejor.2016.01.012.
57. Lo, A. W. (1986) ‘Logit versus discriminant analysis. A specification test and application to corporate bankruptcies’, *Journal of Econometrics*, 31(2), pp. 151–178. doi: 10.1016/0304-4076(86)90046-1.
58. Lukason, O. and Andresson, A. (2019) ‘Tax Arrears Versus Financial Ratios in Bankruptcy Prediction’, *Journal of Risk and Financial Management*. MDPI AG, 12(4), p. 187. doi: 10.3390/jrfm12040187.
59. Mai, F. *et al.* (2018) ‘Deep learning models for bankruptcy prediction using textual disclosures’, *European Journal of Operational Research*. Elsevier B.V. doi: 10.1016/j.ejor.2018.10.024.
60. Muñoz-Izquierdo, N. *et al.* (2020) ‘Does audit report information improve financial distress prediction over Altman’s traditional Z-Score model?’, *Journal of International Financial Management and Accounting*. Blackwell Publishing Ltd, 31(1), pp. 65–97. doi: 10.1111/jifm.12110.
61. Nyitrai, T. and Miklós, V. (2018) ‘The effects of handling outliers on the performance of bankruptcy prediction models’, *Socio-Economic Planning Sciences*. Elsevier, (August 2017), pp. 1–9. doi: 10.1016/0304-3762(82)90059-1.
62. Obradović, D. B. *et al.* (2018) ‘Insolvency prediction model of the company: The case of the republic of serbia’, *Economic Research-Ekonomska Istrazivanja* . Taylor and Francis Ltd., 31(1), pp. 138–157. doi: 10.1080/1331677X.2017.1421990.
63. Ozbayoglu, A. M., Gudelek, M. U. and Sezer, O. B. (2020) ‘Deep Learning for Financial Applications : A Survey’. Available at: http://arxiv.org/abs/2002.05786 (Accessed: 15 April 2020).
64. Pal, R. *et al.* (2016) ‘Business health characterization: A hybrid regression and support vector machine analysis’, *Expert Systems with Applications*. Elsevier Ltd, 49, pp. 48–59. doi: 10.1016/j.eswa.2015.11.027.
65. Park, H., Kim, N. and Lee, J. (2014) ‘Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over KOSPI 200 Index options’, *Expert Systems with Applications*. Elsevier Ltd, 41(11), pp. 5227–5237. doi: 10.1016/j.eswa.2014.01.032.
66. Parsania, D. V., Jani, N. and Bhalodiya, N. (2014) ‘Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis’, *IJDI-ERET*, 3.
67. Policy Division Working Paper (2004) ‘The Importance of Financial Sector Development for Growth and Poverty Reduction’, *Department for International Development*, (August).
68. Prusak, B. (2018) ‘Review of Research into Enterprise Bankruptcy Prediction in Selected Central and Eastern European Countries’, *International Journal of Financial Studies*. MDPI AG, 6(3), p. 60. doi: 10.3390/ijfs6030060.
69. Ptak-Chmielewska, A. (2019) ‘Predicting Micro-Enterprise Failures Using Data Mining Techniques’, *Journal of Risk and Financial Management*. MDPI AG, 12(1), p. 30. doi: 10.3390/jrfm12010030.
70. Qu, Y. *et al.* (2019) ‘Review of bankruptcy prediction using machine learning and deep learning techniques’, in *Procedia Computer Science*. Elsevier B.V., pp. 895–899. doi: 10.1016/j.procs.2019.12.065.

71. Ravi Kumar, P. and Ravi, V. (2007) ‘Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review’, *European Journal of Operational Research*, 180(1), pp. 1–28. doi: 10.1016/j.ejor.2006.08.043.
72. Sartori, F., Mazzucchelli, A. and Gregorio, A. Di (2016) ‘Bankruptcy forecasting using case-based reasoning: The CRePERIE approach’, *Expert Systems with Applications*. Elsevier Ltd, 64, pp. 400–411. doi: 10.1016/j.eswa.2016.07.033.
73. Schumpeter, J. A. (1934) ‘The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle’.
74. Shi, Y. and Li, X. (2019a) ‘A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms’, *Heliyon*. Elsevier Ltd. doi: 10.1016/j.heliyon.2019.e02997.
75. Shi, Y. and Li, X. (2019b) ‘An overview of bankruptcy prediction models for corporate firms: A systematic literature review’, *Intangible Capital*. OmniaScience, 15(2), pp. 114–127. doi: 10.3926/ic.1354.
76. Shumway, T. *et al.* (1999) *Forecasting Bankruptcy More Accurately: A Simple Hazard Model*.
77. Song, Y. gang, Cao, Q. lin and Zhang, C. (2018) ‘Towards a new approach to predict business performance using machine learning’, *Cognitive Systems Research*. Elsevier B.V., 52, pp. 1004–1012. doi: 10.1016/j.cogsys.2018.09.006.
78. Tobback, E. *et al.* (2017) ‘Bankruptcy prediction for SMEs using relational data’, *Decision Support Systems*. Elsevier B.V., 102, pp. 69–81. doi: 10.1016/j.dss.2017.07.004.
79. Tsai, C. F. (2009) ‘Feature selection in bankruptcy prediction’, *Knowledge-Based Systems*. Elsevier B.V., 22(2), pp. 120–127. doi: 10.1016/j.knosys.2008.08.002.
80. Tsai, C. F., Hsu, Y. F. and Yen, D. C. (2014) ‘A comparative study of classifier ensembles for bankruptcy prediction’, *Applied Soft Computing Journal*. Elsevier B.V., 24, pp. 977–984. doi: 10.1016/j.asoc.2014.08.047.
81. Tseng, F.-M. and Hu, Y.-C. (2010) ‘Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks’, *Expert Systems with Applications*. Pergamon, 37(3), pp. 1846–1853. doi: 10.1016/J.ESWA.2009.07.081.
82. Veganzones, D. and Séverin, E. (2018) ‘An investigation of bankruptcy prediction in imbalanced datasets’, *Decision Support Systems*. Elsevier, 112(May), pp. 111–124. doi: 10.1016/j.dss.2018.06.011.
83. Wang, N. (2017) ‘Bankruptcy Prediction Using Machine Learning’, *Journal of Mathematical Finance*, 07(04), pp. 908–918. doi: 10.4236/jmf.2017.74049.
84. Yeh, C. C., Chi, D. J. and Lin, Y. R. (2014) ‘Going-concern prediction using hybrid random forests and rough set approach’, *Information Sciences*. Elsevier Inc., 254, pp. 98–110. doi: 10.1016/j.ins.2013.07.011.
85. Yu, Q. *et al.* (2014) ‘Bankruptcy prediction using Extreme Learning Machine and financial expertise’, *Neurocomputing*, 128, pp. 296–302. doi: 10.1016/j.neucom.2013.01.063.
86. Zhang, L. *et al.* (2020) ‘A descriptive study of variable discretization and cost-sensitive logistic regression on imbalanced credit data’, *Journal of Applied Statistics*. Taylor and Francis Ltd., 47(3), pp. 568–581. doi: 10.1080/02664763.2019.1643829.
87. Zhou, L. (2013) ‘Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods’, *Knowledge-Based Systems*, 41, pp. 16–25. doi: 10.1016/j.knosys.2012.12.007.
88. Zięba, 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*. Elsevier Ltd, 58, pp. 93–101. doi: 10.1016/j.eswa.2016.04.001.