

# **A Systematic Literature Review on Corporate Insolvency Prevention Using Artificial Intelligence Algorithms**

**Tibor Kezelj**  
**Graz University of Technology**

**Rudolf Gruenbichler**  
**Graz University of Technology**

*The research field of insolvency avoidance was particularly fuelled by the economic crisis of 2008. Corporate insolvencies cause financial as well as non-financial damage to the entrepreneurs, to creditors, but also to society through lost jobs or taxes. Neural networks, but also regressions and decision trees, represent interesting approaches to insolvency prevention due to increasing computer power and data availability. This paper presents the current state of research on these selected algorithms in insolvency avoidance. The paper therefore provides a contribution to structuring the current state of research and shows the trends of the research field. For this purpose, a systematic literature search was carried out in the scientific databases with reference to the field of business administration and management with focus on an engineering environment using defined keywords, and the results were processed and analysed. One result is that, due to the different data availability and parameters, research is being carried out into different algorithms for avoiding insolvency and the trend is towards insolvency prevention for SMEs.*

*Keywords: artificial intelligence, neural network, insolvency prevention, systematic literature review, algorithms, SME*

## **INTRODUCTION**

Over the past 20 years, the amount of data being observed is exponential rising. The reason for that is the growing need for information and optimization in every process known today. This paper will put a focus on corporate insolvency prediction and prevention using selected artificial intelligence algorithms. Insolvency is a state of not being able to pay the debts and is closely linked with bankruptcy as that is the next step after. In the research of (Jayasekera, 2018) a prediction of a company failure is being observed. But it is focused mainly on large companies that already have safety nets issued by the government as they employ many employees. This paper will look at the small and medium enterprises, mainly because the number of publications considering SMEs is still relatively new and not ultimately discovered. Small businesses have to consider many issues, from the inability to take the full effects of the economy of scale to a worse credit rating, which makes investments more difficult. One of the aims of this paper is to show the owners of SMEs that they should fight this problem together and improve and grow. As researched by (Camacho-Miñano, Segovia-Vargas and Pascual-Ezama, 2015), 94% of the business in the world are small, 5% are medium, and only 1% is large, which is an essential factor to worry about as this is also a problem

of favouring only the most prominent companies. In addition to that, the whole bankruptcy procedure is costly and has a very negative impact on the psyche of business owners.

That is why research by the likes of (Qu *et al.*, 2019) in the field of bankruptcy prediction is vital in preserving the small business by detecting troubling factors on time. By using real-time data combined with machine learning algorithms, we can prevent insolvency in its beginning and, in that way, improve the corporate ecosystem that fuels most of the population.

This paper provides a systematic literature review, which means that it is an assessment of a body of research that addresses our research question. The question is what is the current status of algorithms for predicting corporate insolvencies with artificial intelligence and what trends can be derived on the basis of the systematic literature review (SLR). SLR aims to identify what is already known about this area of study and provide a case for further research.

The following sections of this paper will discuss the following. The first Section addresses the theoretical foundation of the paper. Section 2 defines the research methodology and the article selection process. Section 3 addresses the theory behind selected machine learning algorithms. In section 4, the findings of the SLR are shown. While in section 5, a conclusion is presented.

## **THEORETICAL FOUNDATION**

Financial difficulties are something that many companies face on a daily basis. The terms that are usually used for describing those situations are “bankruptcy”, “insolvency” and “company failure”. As financial difficulties happen daily, there is a discrepancy between the data that a company receives and the actual financial state of a company. Small and medium-sized enterprises are more exposed because their insolvency prevention systems are usually not as highly monitored as their bigger equivalents.

In that regard, this paper was influenced by the works such as (Antunes, Ribeiro and Pereira, 2017) that have researched the idea of probabilistic modelling and visualization to predict bankruptcy sooner. The problem of identifying a problem early is the task of preventive diagnostics, which guides the fight against insolvency. (Camacho-Miñano, Segovia-Vargas and Pascual-Ezama, 2015) researched the principle of using artificial intelligence methods to analyse a sample of companies and determine which characteristics predict the survival of insolvent companies. The usage of machine learning techniques is essential to segregate the needed data and provide an easy gateway for the researchers such as (Qu *et al.*, 2019) to review and analyse. The sorting of information presents a significant challenge, and relation between researched firms dramatically increases the performance of research, like in the case of (Tobback *et al.*, 2017) where the relation shared are high-level managers. A multistage classification developed by (Swiderski, Kurek and Osowski, 2012) is a necessity to predict the risk of insolvency. Objective results provide the much needed accuracy of data acquisition (Obermann and Waack, 2015). But, every company has a slightly different mindset, and therefore a standardized approach to insolvency prevention is a good base but a skilled researcher needs to have the right information about the past, present and the future of a business (Jayasekera, 2018). The keyword and part of the focus in this article are SMEs and in that regard a slightly different approach to insolvency prediction is taken due to the nature of SMEs (Lee, Choi and Yoo, 2020).

## **RESEARCH METHODOLOGY**

This systematic literature review (SLR) has been carried out by applying academic databases Scopus, Science Direct and Web of Science published from January 2008 to March 2021. The initial search for identifying academic papers related to the research topic of corporate insolvency were carried out by using keyword combinations as search algorithms. The algorithms are based on previous systematic literature reviews conducted by the following authors (Shi and Li, 2019a) in their two SLR that cover a similar topic but from another angle.

The SLR methodology is an essential feature of any academic project. It is crucial to give researchers insight into the current state of academic documentation on their respected topic. The advantage is that

reliable information will be provided by critically evaluating and integrating all relevant, high-quality papers.

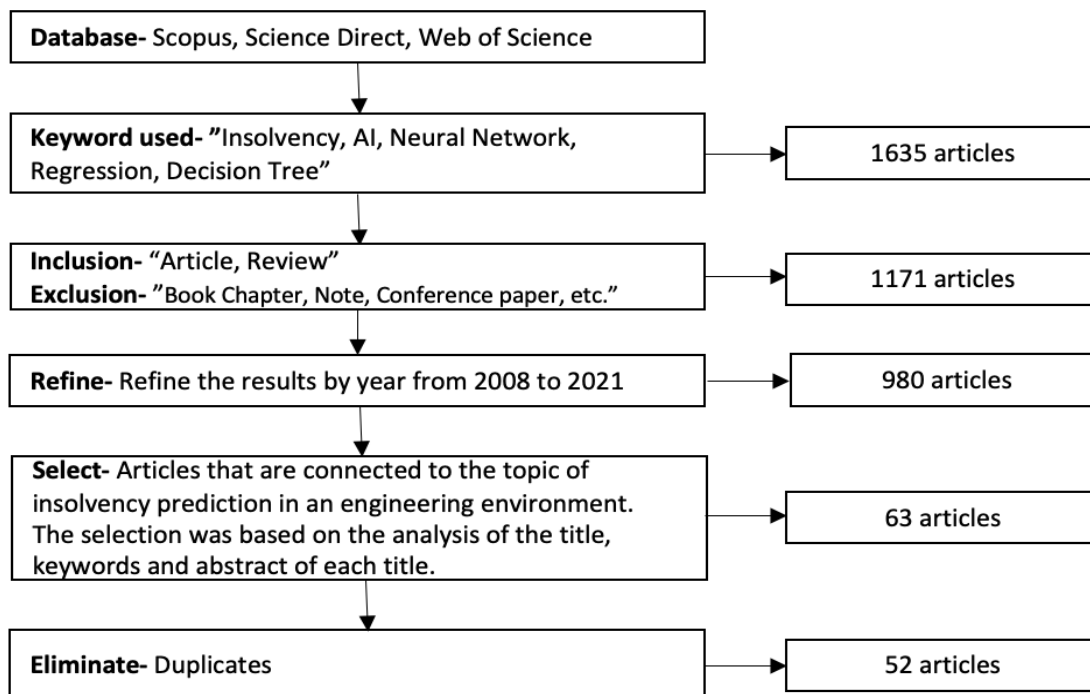
### Article Selection Process

This study will take a focused and structured approach when selecting which papers will be a part of this SLR. The studies collected were analysed concerning keywords and logics to discuss the issues and advances in insolvency prediction and prevention. All the articles were collected from the following three academic databases: Scopus, Science Direct, Web of Science.

As seen in Figure.1, the first part of the searching process was to select the keywords needed for this study and to make a search string that would incorporate the terms of Insolvency, AI, Neural Network, Regression and Decision Tree. All of the databases have a slightly different way of searching, so minimal adjustments had to be made. The first part produced 1635 articles. The second part was to determine which papers will be included and excluded from the search. Only peer-reviewed articles were used for the paper. Book chapters, notes, conference papers and such were excluded as they do not fit the parameters. The second part produced 1171 articles. The third part of the search was a refinement by the year. As this is a recent paper with AI and machine learning techniques, a decision was made to consider only newer selected approaches to insolvency prevention in particular like the neural network. A reason to take that specific year was the economic crisis that happened that year. After the refinement, the number of articles dropped to 980. The fourth part was the most interesting one. All the papers were reviewed, and only the ones connected to insolvency prevention in an engineering environment were selected. The selection was based on the analysis of the title, keywords and abstract of each title. This selection dropped the number of articles to 63. The fifth and last part was a duplicate elimination activity which narrowed the number of reports to 52.

After the search was over, each of those 52 articles was researched, and the most important topics will be reviewed in this paper.

**FIGURE 1  
PROCES FOR SELECTING RESEARCH PAPERS**



## ALGORITHMS ON CORPORATE INSOLVENCIES

There are many algorithms used for predicting and preventing corporate insolvencies. Each one has its advantages and drawbacks. In this paper the three Neural Networks, Logistic Regression and Decision tree were considered. During the research, those methods were the ones that popped out the most. As this is a study considering articles after 2008, these methods are also one of the main ones used to understand and implement the Artificial Intelligence approach.

Information and Communication Technology (ICT) tools and techniques are first applied in accounting, area that has most of matter for every company. At first, ICT applications was used in the elementary financial systems. Later, the economic modelling schemes showed that it is crucial in the analytical aspect of accounting. ICT implementation in accounting is slow and it has conservative approach. Until the end of 1990, this works was computerised to promote its efficiency, reduce expenses and to be more competitive.

Artificial intelligence was something that made an impact in accounting efficiency, so most of the big companies decided to implement it. Electronic Data Interchange, Electronic File Transfer and Image processing were some of the ICD devices that are progressively replacing traditional methods and changing the entire audit process.

There were major advances in auditing in the last hundred years. The idea of auditing is to give and insight in the true state of financial situation of a company by using the neutral approach that a third-party expert has. It is understood that the information is intensive and all the processes that implement data mining and organising and evaluating the data are presenting data in order to produce a relevant option in the process of auditing. The final audit is most commonly a mix of the grades of the auditors that have been grading the company on all aspects of financial data that was considered relevant. A relevant audit option is important in order to have everything right and organised in the right way.

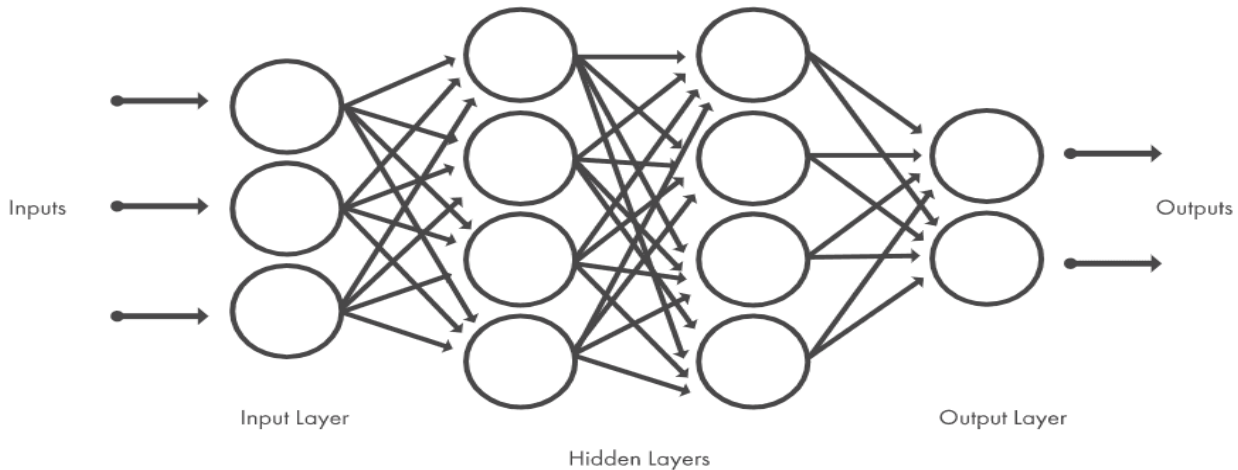
### Neural Networks

Neural Networks is arguably one of the most popular methods in machine learning and a likely role model on which other machine learning methods were developed. It is concluded that the method is similar to a human brain in the way of its processing and functioning. There are layers upon layers similar to neurons in a human brain. Those layers are influenced by input factors that guide the method (Qu et al., 2019). The philosophy behind a NN can be explained simply as a system that mimics a human brain and, in that regard, processes information. It is an inspiration for other computational models is merely its resemblance with our brain. Therefore, it is easy to adapt to other computational models according to NN.

Neural Networks use the imitation of the processing of a human brain and its neurons to function. The AI implementation of neural networks are called artificial neural networks. They work on two models which guide the reasoning that is similar to human reasoning. The first model examines the neural networks from a biological standpoint that mimics a human brain to experiment and further understand the functions of the brain. The second method sees neural networks as data handling method which has a special structure that allows a neater data processing (Alrammahi and Radif, 2019).

Learning is the basic process of every machine learning method. It symbolizes the calculations of the power of neurons in a neural network approach. The power is calculated as the value of the neuron which affects the output of the method. There are two learning varieties. Supervised and unsupervised. Supervised learning is learning where each variable is influenced by an operator and is controlled in every way. Unsupervised learning is learning where the AI takes over and the operator sees the results but not the iterations that happen inbetween (Alrammahi and Radif, 2019).

**FIGURE 2**  
**NEURAL NETWORK STRUCTURE**



### **Regression**

Logistic regression is fundamentally a linear model, which means it has an output of either a 0 or 1 (Qu et al., 2019). Logistic regression is a model that is mostly used for producing classification systems of businesses. The first usage of the method was the prediction of insolvency using data that was available in the public. Logistic Regression is a method that regulates the significance of variables of credit risk in two categories. The idea is that there are only good debtors or bad debtors (Petropoulos et al., 2020). The simplicity of the model comes from the fact that there are only two possibilities to assign data.

“In the first step of the proposed approach, we transform the multiclass task into few simple binary models of classification (similar to decision tree) and then, in the second stage, apply the additional classifier responsible for undertaking the final decision of the classification”. (Swiderski, Kurek and Osowski, 2012) The parameters are estimated based on maximum similarity. This process does not need independent variables to have multivariate normal distribution. In order to achieve a needed predictivity and reasonable results interdependent variables are used because they provide information that is needed for the results. (Nyitrai, 2019)

According to (Shi and Li, 2019b) regression is the most frequently used statistical model because it gives reliable data which can be analysed further using other machine learning methods. In other words, logistic regression has a black and white approach that researchers can easily understand.

### **Decision Tree**

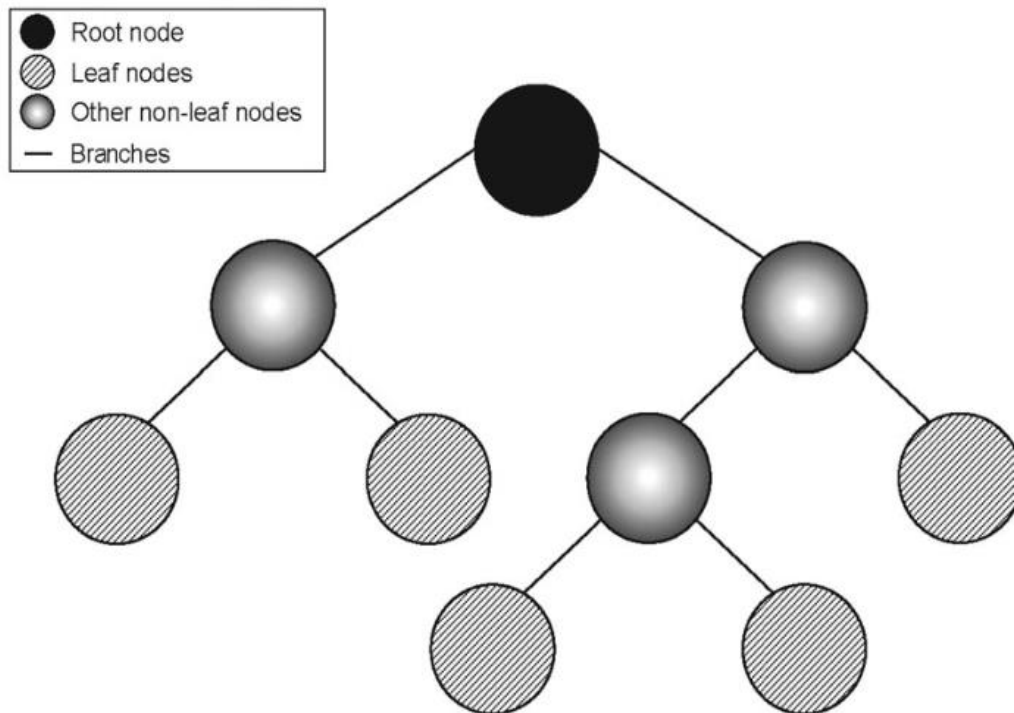
“Decision tree (DT) techniques generate tree-based classification rules to construct a DT (also known as a classification tree). DTs assign data to predefined classification groups: a DT usually gives each business to a successful or failing group. In general, DTs are binary trees, consisting of a root node, non-leaf nodes and leaf nodes connected by branches, whereby each non-leaf node has two branches leading to two distinct nodes” (Gepp, Kumar and Bhattacharya, 2010).

One of the main bonuses of decision trees is their non-parametricity, meaning that there is no need to include transforming variables. The downside is that previously researched data is needed for this method to achieve its goal. If there is no data available, the DT will not be precise, giving conflicting results.

These results are pretty easy to understand, and analysts or managers can understand how classes function (Obermann and Waack, 2015). As with every method, the goal is to gather the data that will predict

and prevent insolvency from happening in the first place. Because of that reason, great attention must be dedicated to methods that offer a structure that is understandable to anyone that observes it.

**FIGURE 3**  
**THE BASIC STRUCTURE OF A DECISION TREE**



## RESULTS AND FINDINGS

To observe and assess the trends of publications in this field of research, a systematic literature review regarding insolvency prevention and the prediction was performed. VOSviewer software was applied to analyse the papers and produce a visual representation of the co-occurrence of keywords, bibliographic coupling of countries and the overview of when specific keywords were researched.

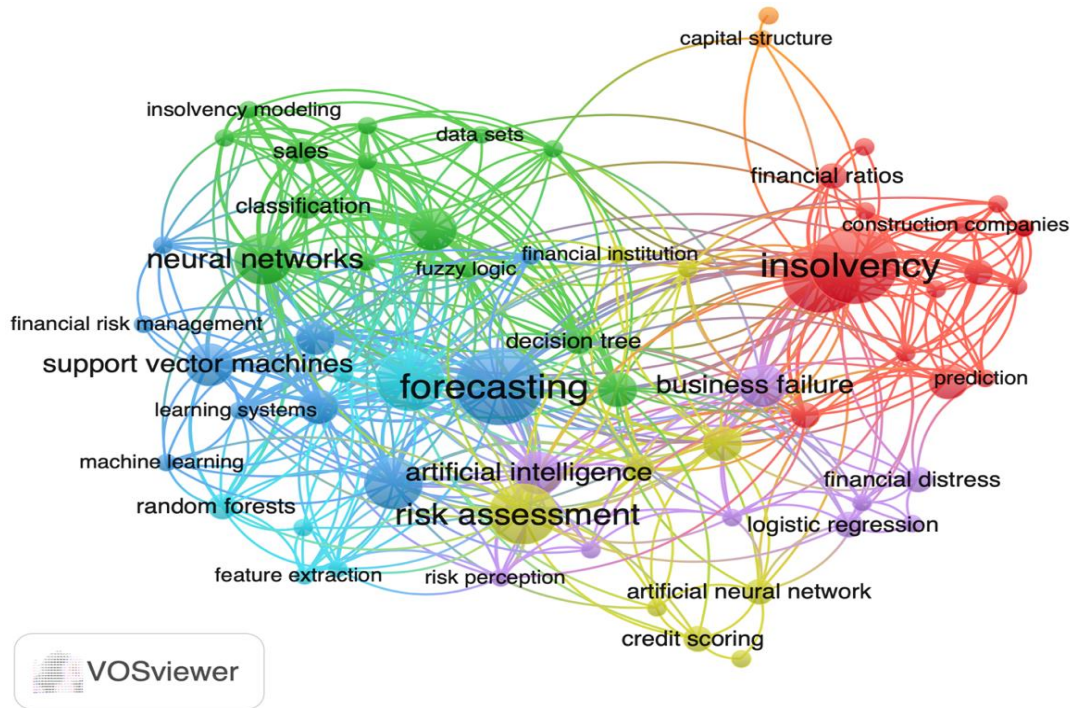
The results show a descriptive result from 52 articles obtained from Scopus, Science Direct and Web of Science. The number of publications is on the rise since the economic crisis of 2008. That shows the relevance of the topic in the eyes of researchers. During the search for the SLR articles on the points of insolvency, a large number of reports were produced (15242), but with the implementation of keywords such as neural networks, regression, decision tree, and AI drops the amount to a staggering 52. That shows that the topic could be researched even more in the grand scheme of insolvency prevention and prediction. The most relevant database was Scopus, followed by ScienceDirect and Web of Science. In the results, the complete overview of the researched data shows.

### Network Visualization Keyword Co-Occurrence Analysis

VOSviewer was used for the graphical evaluation. The logic of VOSviewer is as follows: In the network visualisation, keywords are represented by their label as well as by a circle. The higher the importance of an item, the larger the label and the circle are shown. For some items, the label is not displayed. This is done to avoid overlapping of the labels. The colour of an entry is determined by the cluster to which the entry belongs. Lines between entries represent links.

The distance between two journals in the visualization indicates the relatedness of the journals in terms of co-citation links. In general, the closer the two journals are located to each other, the stronger their relatedness. Lines also represent the strongest co-citation links between journals.

**FIGURE 4  
CO-OCCURRENCE KEYWORD NETWORK VISUALIZATION**

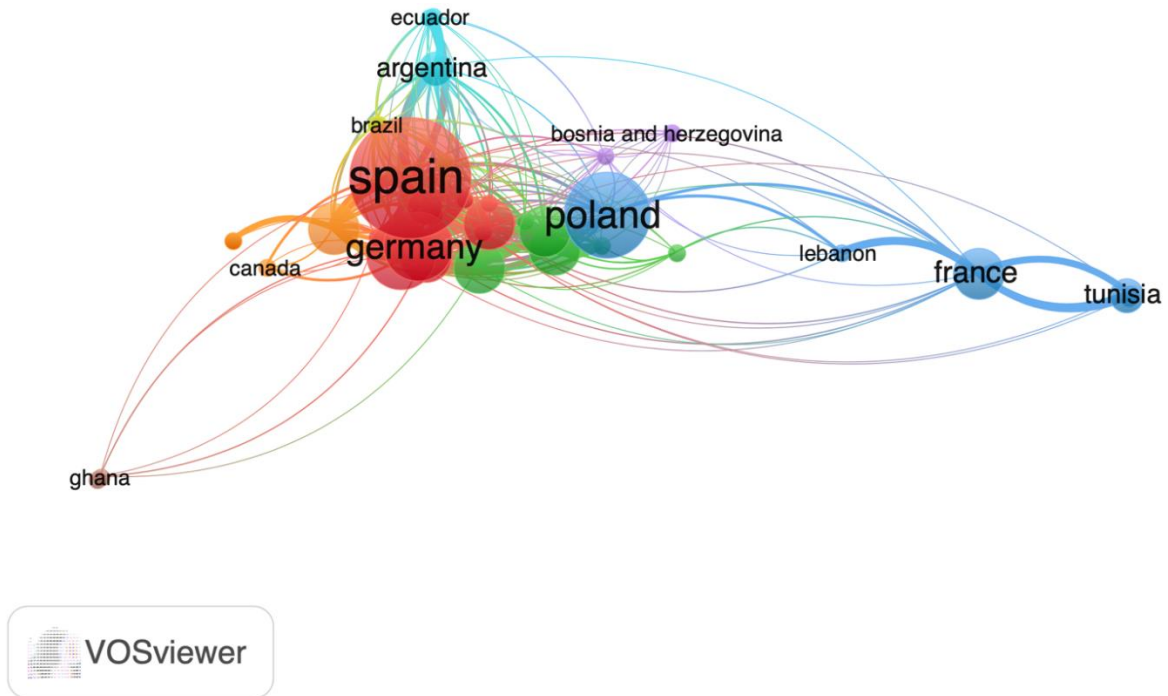


### Network Visualization Bibliographic Coupling Countries Analysis

A bibliographic coupling is a measure that uses citation analysis to determine similarities between documents. The strength is calculated regarding a reference to a shared document that they share. The link is stronger the more citations the papers share. In the case of countries bibliographic coupling analysis, we can see and understand which countries are the drivers in this research.

The data shows that the main drivers are European countries Spain, Germany and Poland, but it is easy to spot that the effort to fight insolvency is global, a concern in many different cultures. As stated by (Szetela, Mentel and Brożyna, 2016) there is a growing trend of going bankrupt because of their inability to fight insolvency. And it is a problem that even the most robust economies have to fight. In the case of (Grdić, Nižić and Mamula, 2017) where they observed a small European country Croatia, it is shown that smaller countries often lack measures that can combat the growing concern of insolvency. That is why AI and machine learning methods could significantly impact lending a helping hand to countries in that regard.

**FIGURE 5**  
**BIBLIOGRAPHIC COUPLING NETWORK VISUALIZATION**



### **Overlay Visualization of When Specific Keywords Were Researched**

The overlay visualization is identical to the network visualization, except that items are coloured differently. There are two ways in which things can be coloured in the overlay visualization. If items have scores, the colour of an object is determined by the score of the item, where by default, colours range from blue (lowest score) to green to yellow (highest score).

This overlay visualization focuses on the timeliness of specific keywords and the number of papers published in the last 13 years. The keywords included in the search algorithm have started to make a more significant impact from 2014, shown in this visualization. As stated before, decision trees are a simple method that is easy to implement, and it is no wonder that it is one of the first ones to gain the attention of researchers. In the last few years, there is a growing trend of observing and researching small and medium enterprises as they have always been hard to audit because of their dynamic structure. In the researches of (Camacho-Miñano, Segovia-Vargas and Pascual-Ezama, 2015; Tobback et al., 2017; Lee, Choi and Yoo, 2020) a lot of care is focused on the issues concerning SMEs. This is mainly because many surveyed enterprises are younger than three years, and no relevant data can be collected in that timeframe. But, that is where AI takes over a fair portion of the assignment, and new methods are being discovered that can collect the data daily and not yearly as it was done in the past.





## REFERENCES

- Alrammahi, A.H.I., & Radif, M.J. (2019). Neural networks in business applications. *Journal of Physics: Conference Series*. Institute of Physics Publishing. doi: 10.1088/1742-6596/1294/4/042007
- Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing Journal*, 60, 831–843. doi: 10.1016/j.asoc.2017.06.043
- Camacho-Miñano, M.D.M., Segovia-Vargas, M.J., & Pascual-Ezama, D. (2015). Which Characteristics Predict the Survival of Insolvent Firms? An SME Reorganization Prediction Model. *Journal of Small Business Management*, 53(2), 340–354. doi: 10.1111/jsbm.12076
- Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of Forecasting*, 29(6), 536–555. doi: 10.1002/for.1153
- Grdić, Z.Š., Nižić, M.K., & Mamula, M. (2017). Insolvency in the Republic of Croatia. *Economic Research-Ekonomska Istrazivanja*, 30(1), 1693–1704. doi: 10.1080/1331677X.2017.1383177
- Jayasekera, R. (2018). Prediction of company failure: Past, present and promising directions for the future. *International Review of Financial Analysis*, 55, 196–208. doi: 10.1016/j.irfa.2017.08.009
- Lee, S., Choi, K., & Yoo, D. (2020). Predicting the insolvency of smes using technological feasibility assessment information and data mining techniques. *Sustainability (Switzerland)*, 12(23), 1–17. doi: 10.3390/su12239790
- Nyitrai, T. (2019). Dynamization of bankruptcy models via indicator variables. *Benchmarking*, 26(1), 317–332. doi: 10.1108/BIJ-03-2017-0052
- Obermann, L., & Waack, S. (2015). Demonstrating non-inferiority of easy interpretable methods for insolvency prediction. *Expert Systems With Applications*, 42(23), 9117–9128. doi: 10.1016/j.eswa.2015.08.009
- Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N.E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3), 1092–1113. doi: 10.1016/j.ijforecast.2019.11.005
- Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science. Elsevier B.V.*, pp. 895–899. doi: 10.1016/j.procs.2019.12.065
- Shi, Y., & 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
- Shi, Y., & Li, X. (2019b). An overview of bankruptcy prediction models for corporate firms: A systematic literature review. *Intangible Capital*, 15(2), 114–127. doi: 10.3926/ic.1354
- Swiderski, B., Kurek, J., & Osowski, S. (2012). Multistage classification by using logistic regression and neural networks for assessment of financial condition of company. *Decision Support Systems*, 52(2), 539–547. doi: 10.1016/j.dss.2011.10.018
- Szetela, B., Mentel, G., & Brożyna, J. (2016). In search of insolvency among European countries. *Economic Research-Ekonomska Istrazivanja*, pp. 839–856. Taylor and Francis Ltd. doi: 10.1080/1331677X.2016.1237301
- Tobback, E., Bellotti, T., Moeyersoms, J., Stankova, M., & Martens, D. (2017). Bankruptcy prediction for SMEs using relational data. *Decision Support Systems*, 102, 69–81. doi: 10.1016/j.dss.2017.07.004