

Big data and credit risk assessment: a bibliometric review, current streams, and directions for future research

Haitham Nobanee, Hiba Shanti, Hind Aldhanhani, Abdulrahman Alblooshi & Essa Alali

To cite this article: Haitham Nobanee, Hiba Shanti, Hind Aldhanhani, Abdulrahman Alblooshi & Essa Alali (2022) Big data and credit risk assessment: a bibliometric review, current streams, and directions for future research, Cogent Economics & Finance, 10:1, 2132638, DOI: 10.1080/23322039.2022.2132638

To link to this article: <https://doi.org/10.1080/23322039.2022.2132638>



© 2022 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.



Published online: 12 Oct 2022.



Submit your article to this journal [↗](#)



Article views: 1186



View related articles [↗](#)



View Crossmark data [↗](#)



Received: 04 April 2022
Accepted: 02 October 2022

*Corresponding author: Haitham Nobanee, College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates
E-mail: haitham.nobanee@liverpool.ac.uk; haitham.nobanee@adu.ac.ae; haitham.nobanee@oxcis.ac.uk; nobanee@gmail.com

Reviewing editor:
David McMillan, University of Stirling, Stirling, United Kingdom

Additional information is available at the end of the article

FINANCIAL ECONOMICS | REVIEW ARTICLE

Big data and credit risk assessment: a bibliometric review, current streams, and directions for future research

Haitham Nobanee^{1,2,3*}, Hiba Shanti¹, Hind Aldhanhani¹, Abdulrahman Alblooshi¹ and Essa Alali¹

Abstract: This study aims to track the structural development of academic research on credit risk assessment and big data using bibliometric analysis. The bibliography is obtained from the Scopus database and contains all studies with citations published between 2012 and 2021. The study's findings suggest that credit risk assessment and big data are vast fields that have increased significantly in the last nine years. Chinese researchers and organizations contributed the most to the documents. The current study concludes that several possibilities exist to improve the knowledge of credit risk assessment and big data.

Subjects: Corporate Finance; Banking; Credit & Credit Institutions

Keywords: credit risk; network mapping; risk assessment; bibliometric; big data

Subject classification codes: G32; C89

1. Introduction

Credit risk develops when the debtor fails or delays debt repayment, whether in entirety or part, in a debt contract. Credit risk is defined by (Anderson, 2013) as the “probability that a legally enforceable contract would become worthless (or at least significantly decreased in value) due to the counterparty defaulting and going out of business.” As stated by Saunders & Cornett, (2017), it is “the risk of not being able to fully pay out the promised cash flow from financial institutions’ loans and securities.” Hence, credit risk develops because of default in derivative agreements between debt issuers and counterparties.

The relevance of credit risk has been widely acknowledged among academics across disciplines. Consequently, credit risk measures have received considerable attention, particularly in the corporate finance sector. Big data can be utilized to obtain a deeper understanding of the processes and credit status of various companies. (Wang et al., 2020) This will aid in better risk assessment thus decreasing the level of default.

Hence, several studies have attempted to use big data to develop credit risk assessment models that will aid lenders in decreasing the level of credit default. These models were designed to enhance the entire process and improve the quality of assessment.

Therefore, this work provides a novel approach to further explore credit risk assessment and big data by conducting a bibliometric assessment to uncover distinct notions within this large body of literature, as a comprehensive bibliometric analysis on this topic is missing. This is necessary to identify what has been accomplished and what is required in future research.

The current study listed and analyzed all the articles and studies related to credit risk assessment and big data in journals published between 2012 and 2021 (July) using the Scopus database. This study explains the recent research progress in credit risk assessment and big data, which articles were more cited, and which authors and journals contributed the most to the field.

This study contributes significantly to credit risk assessment and big-data research. First, the work describes, organizes, and identifies key institutions, journals, publications, and authors for future studies. Additionally, we identify significant keywords and cite studies that could prepare future researchers to conduct new research on this topic. The study also offers an overview of the history of the research and summarizes and identifies present and developing research streams in credit risk assessment and big data.

This study combines the big data and credit risk assessment aspects. It performs a bibliometric analysis, different from other bibliometric studies published in this regard, as existing studies only focused on one of the two dimensions.

The results of this study show that the National Natural Science Foundation of China is the top sponsoring entity, and the Journal of Physics: Conference Series has the most published studies associated with credit risk assessment and big data research. The Electronic Commerce Research and Applications Journal has the most cited articles.

Author Zhang Y. published the most articles on the topic, whereas Yu Y had the most citations. Finally, China is the top country in credit risk assessment and big data research. Hence, it can be concluded that Chinese authors, institutes, and organizations have dominated credit risk assessment, big data, and related research in the last few years.

2. Literature review

2.1. History of credit risk assessments approaches

Credit risk is an essential form of financial risk and is frequently seen as the earliest type of financial market risk from the 1800 BCE period in the ancient Egyptian era (Caouette et al., 2011). Altman's Z-score (Benzschawel, 2012), established on the multivariate discriminant analysis of five accounting measures, was the first contemporary quantitative credit risk assessment model. Even though it is 50 years old, the Z-score remains an important instrument for many market participants (Benzschawel, 2012). However, because accounting ratios depend on past data, the approach has been criticized for being backward-looking and sporadic. Other credit risk models, such as structural and reduced-form models, have been developed owing to this.

(Black & Scholes, 1972) and (Merton, 1973) are credited with developing structural credit risk-assessment models. According to capital structure theory (Modigliani & Miller, 1958), structural models imply that a default occurs when a company's assets are worth less than its debt. (Black & Scholes, 1972) employ an options pricing model to price debt and equity, proving that call options on equity may affect the debt value. The issue with the Black-Scholes model is that a company's asset values cannot be directly monitored. (Merton, 1974) continues his work by showing that the asset value may, on some assumptions, be calculated and then used to calculate the probability of a default, which he calls the "distance to the default." However, the Merton model is considered the most important for credit risk modeling in terms of structural models.

In contrast to structural models, reduced-form models can identify the risk of default without assuming the cause of the credit risk premium (Benzschawel, 2012). The reduced-form models are based on risk-neutral pricing theory, which asserts that a risky investment's market value equals the current value of future risk-free cash fluxes (e.g., the US Treasury rate). (Jarrow & Turnbull, 1995), (Duffie & Singleton, 1999) and many others use risk-neutral pricing theory to estimate credit

risk. The reduced-form method has also been a prominent paradigm (Diaz Weigel & Gemmill, 2006).

2.2. Components of credit risk

Default, spread, and downgrade risks are the three main components of credit risk (Anson et al., 2004). Default risk is the risk that the issuer or counterparty will not fulfill the terms of a financial contract's obligation. Loss or failure to perform a problem caused by an increased lending spread is known as the risk of credit spread. The credit spread illustrates how financial markets react to an issue's projected credit quality deterioration. The danger of credit rating degradation is known as the downgrade risk. When a rating agency assigns a lower grade than the prior grade, the issuer is at risk of being downgraded. These three forms of credit risk are inextricably linked.

2.3. Types of credit risk assessment models

Credit risk assessment models are often divided into two different groups: qualitative and quantitative (Saunders & Cornett, 2017). In analyzing credit risk, the value of the variables includes the characteristics of the borrower (e.g., reputation, levers of funds, volatility in income, and collateral.) and those of the market (such as business cycles, interest rate level.). A subjective assessment is conducted on these variables to determine whether a candidate may be granted credit. Quantitative models seek to obtain a credit score to establish the chance of default or categorize borrowers into distinct default risk categories (Saunders & Cornett, 2017).

Quantitative credit risk modeling is not a simple process because default risk, a component of credit risk, seldom occurs (Anson et al., 2004). However, the concept of credit risk has evolved over the years in several quantitative credit risk models. They are frequently based on other models on a certain theoretical basis, such as ratios, theory of options, econometrics, or expert systems (Caouette et al., 2011). Econometrics, simulation, optimization, or a combination of these are frequently used to build financial models. Therefore, credit risk models may be classified into three components: methodologies utilized, application fields, and products engaged (Caouette et al., 2011).

(Caouette et al., 2011) defined econometric methods as statistical models in which the likelihood of default is the dependent variable. The available models include linear probability, logit, probit, linear discrimination analysis, and other regression methods (Altman & Saunders, 1997; Saunders & Cornett, 2017). Neural networks are computer-based systems (that function similarly to the human brain) that employ the same data in econometric models to make choices, typically through trial and error. They frequently look for connections between discrete choice model variables (e.g., the logit model). Hybrid models integrate structural models with additional financial variables to approximate a company's probability of default (e.g., the book value of assets and liabilities, net income, and return on equity; Benzschawel, 2012). The KMV and HPD models (Sobehart & Keenan, 2001) are two examples.

The multiple classifier systems (MCS) have evolved in the last few years to enhance the accomplishment of a single prediction model in financial credit risk assessment. Many researchers have demonstrated that the MCS technique may generate superior results than individual credit risk evaluation models (Verikas et al., 2010; Zhou, 2012). In four different classifiers on three financial data sets, (Nanni & Lumini, 2009) explores four ensemble methods using four different classifiers on three financial datasets. They identify a random subspace that generates the largest area under the ROC curve (AUC).

Credit risk assessment models are further classified into two types: i) consumer and ii) corporate. Although both models have identical underlying assumptions, much of the existing literature focuses on the corporate perspective. Credit risk analysis was first used to assess consumer credit risk in the period after 1950 to estimate the creditworthiness of retail customers (Thomas et al., 2005). Consumer credit risk models include neural networks and expert system/decision tree

models. Altman's (Altman & Saunders, 1997) is the 1st quantitative model of business credit (Benzschawel, 2012). Then came the publications of (Black & Scholes, 1972), (Merton, 1974), and (Jarrow & Turnbull, 1995), which provided the groundwork for corporate credit risk modeling research (Kealhofer, 2003).

2.4. Big data and credit risk assessment

A study by (Wagdi & Tarek, 2022) examines the efficiency of technology models in credit risk-scoring modeling in developing markets. It suggests evaluation approaches for credit risk-scoring modeling for present and possible debtors through an examination of the Egyptian banking field by proposing and investigating a framework for the integration of big data and artificial neural networks based on systematic and unsystematic risk for the macroeconomic environment and features of present and possible debtors.

Moreover, a study by (Pérez-Martín et al., 2018a) highlights the massive number of databases financial corporations manage. It has become essential to resolve this issue by deploying big data methods to enormous financial datasets for segmenting risk groups. Several Monte Carlo experiments are applied to massive datasets using known techniques and algorithms. Additionally, a linear mixed model (LMM) is employed as the latest incremental contribution to assess the credit risk of financial firms. The results show that big data can assist in extracting the value of data; therefore, superior choices can be made without the runtime element.

Additionally, a study by (Md et al., 2020) examines the present trend of how financial sectors deal with big data and demonstrates how big data affects diverse financial areas. More precisely, it demonstrates its influence on financial markets, financial institutions, and the relationship with Internet finance, financial management, Internet credit service companies, fraud detection, risk analysis, and financial application management.

Likewise, (Addo et al., 2018a) state in their study that as advanced technology related to big data, data availability and computing power is rising, and several banks and lending organizations are updating their business models accordingly. Credit risk forecasts, supervision, model consistency, and efficient loan management are crucial to decision-making and transparency. Addo et al., (2018a) focused on building a binary classifier model based on machine and deep learning real data simulations to forecast the loan default probability.

However, (Mhlanga, 2021) reveals that artificial intelligence and machine learning strongly influence credit risk assessments by employing alternative data sources, such as public data, to address the issues of information asymmetry, adverse selection, and moral hazard. Therefore, lenders perform a serious credit risk analysis, evaluate the consumer's behavior, and validate customer's competence to reimburse loans, enabling less fortunate individuals to access credit.

Big data approaches have recently been added to enhance credit risk assessment procedures. Big data is fundamentally a collection of data that can be obtained, saved, controlled, and analyzed by a computer in a short period. Big data has enhanced the effectiveness of data transmission, storage, management, and sharing. Big data analysis technology can recognize these massive amounts of data in credit risk, enhancing the exactness and scientificity of risk forecasting and early warning (Du et al., 2021). Additionally, (Cui, 2015) states that the idea, technique, and means of big data are inserted in recent evaluation systems. This is to increase the capturing width, depth, and real-time information and utilize the scientific method to mine the key value in huge data, which is effective for enhancing risk evaluation and prediction.

Hence, this study uses a longitudinal approach because articles are linked through time (Small, 1999), and credit risk models incorporate various ideas and notions from other fields. Moreover, because big data is also evolving in the credit risk assessment process, it is essential to study the link and how the topic is evolving. The study addresses three research questions, adapted from

(Fetscherin & Heinrich, 2015) but adjusted for the study, using a multidisciplinary assessment of the literature:

- (1) What is the evolution of the idea of credit risk assessment and big data, what are the main research streams, and which require more attention?
- (2) Which studies, articles, and authors are the most referenced and worth reading in this subject for future studies?
- (3) What are the most significant institutions, organizations, and countries contributing to credit risk assessment and big data?

3. Methodology

The quantitative study of science aims to improve understanding, and bibliometric analysis plays an essential part in this field (Van Raan, 2004). The quantitative examination of technical terminology, production, development, cooperation, and usage of scientific publications is known as a bibliometric analysis. In recent years, bibliometric analysis has become popular for measuring and analyzing scientists' output, collaboration among authors and institutions, output comparison, highly cited outputs, and co-citation analysis. According to Ellegaard & Wallin, (2015), bibliometric analysis is an important aspect of research assessment techniques, especially in the scientific and practical domains. The use of bibliometric analysis for collaboration between industry and institutes was discussed by (Skute et al., 2019). Hence, bibliometrics analyzes the quantitative characteristics of research using statistical and mathematical approaches (Broadus, 1987). According to Cobo et al., (2011), Bibliometric mapping is a 3D depiction of the relationships between disciplines, fields, areas, and specific publications or authors.

Moreover, the bibliometric analysis identifies focus studies and objectively depicts the links among publications concerning a particular research subject by assessing how many occasions they have been co-cited by other published works (Apriliyanti & Alon, 2017; Fetscherin & Heinrich, 2015). The findings may be used to determine the popularity of important authors, their articles, and their effects. Thus, the bibliometric analysis makes it easier to evaluate meta-analytics, create and discover key research streams, and develop basic theoretical agendas (Apriliyanti & Alon, 2017; Fetscherin & Heinrich, 2015; Nobanee, 2020, 2021; Nobanee et al., 2021)

Academics publish their crucial discoveries and findings in research journals and typically base their study on studies/documents formerly published in similar publications according to bibliometric analysis (Van Raan, 2003). The analysis unit in any citation analysis is considered a citation (Kim & McMillan, 2008) that goes beyond the basic listing of research articles to incorporate centers of excellence (Fetscherin & Usunier, 2012) and map connectivity among research field publications.

Thus, it was found that the bibliometric analysis approach is the most suitable methodology for our research, as it will help identify critical studies on this topic. Moreover, it will help illustrate the connections among publications regarding this topic by evaluating how many instances have been co-cited by other published works. Thus, it will facilitate the evaluation of key research streams and develop basic theoretical agendas, unlike other methods, which are limited to delivering only a literature review on this topic.

Hence, the current study uses information from the Scopus database of Elsevier abstracts and citations. The data from Scopus are obtained using keywords, which can be summed up in Table 1, as a subject for 2012–2021. The most frequently used keywords and their occurrences are summarized in Table 2. Table 2 shows that the most frequent word in the search was “Big Data,” with a total occurrence of 145. The keyword “Risk Assessment” was used 80 times. The details of other major keywords are shown in the table.

Table 1. Search queries using keywords

Description	Conditions	Number of documents
The search query of TITLE - ABS—KEY	TITLE-ABS-KEY (("big data" AND "credit risk") OR ("big data" AND "credit expos*") OR ("big data" AND " loan risk") OR ("big data" AND "default risk") OR ("big data" AND "credit default") OR ("big data" AND "risk of default") OR ("big data" AND "insolvency risk") OR ("big data" AND "exposure at default") OR ("big data" AND "downgrade risk") OR ("big data" AND "credit rating") OR ("big data" AND "credit score") OR ("big data" AND "credit rate") OR ("big data" AND "credit assessment") OR ("big data" AND "credit evaluation") OR ("big data" AND "credit analysis") OR ("big data" AND "credit valuation") OR ("big data" AND "credit appraisal") OR ("big data" AND "credit grading") OR ("big data" AND " credit check") OR ("big data" AND "credit history"))	219 documents
Search query after refining	TITLE-ABS-KEY (("big data" AND "credit risk") OR ("big data" AND "credit expos*") OR ("big data" AND " loan risk") OR ("big data" AND "default risk") OR ("big data" AND "credit default") OR ("big data" AND "risk of default") OR ("big data" AND "insolvency risk") OR ("big data" AND "exposure at default") OR ("big data" AND "downgrade risk") OR ("big data" AND "credit rating") OR ("big data" AND "credit score") OR ("big data" AND "credit rate") OR ("big data" AND "credit assessment") OR ("big data" AND "credit evaluation") OR ("big data" AND "credit analysis") OR ("big data" AND "credit valuation") OR ("big data" AND "credit appraisal") OR ("big data" AND "credit grading") OR ("big data" AND " credit check") OR ("big data" AND "credit history")) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cr") OR LIMIT-TO (DOCTYPE, "ch") OR LIMIT-TO (DOCTYPE, "ed") OR LIMIT-TO (DOCTYPE, "re")) AND (LIMIT-TO (LANGUAGE, "English"))	214 documents

Very few empirical findings have tried to assess credit risk assessment and big data using bibliometric analysis. However, this study will help scholars better understand credit risk management and big data as a topic that is genuinely important to today's economic and financial systems. The current investigation will add to the academic literature by shedding light on credit risk assessment and big data by identifying and illustrating key patterns in this domain by utilizing Scopus and VosViewer.

The table above shows that the total number of documents obtained before and after search query refining are 219 and 214, respectively. The research includes both open- and non-open-

Table 2. Most occurrence keywords

Rank	Keyword	Occurrences	Rank	Keyword	Occurrences
1	Big Data	145	11	Credit Risks	20
2	Risk Assessment	80	12	Artificial Intelligence	19
3	Finance	43	13	Forecasting	19
4	Data Mining	35	14	Financial Institution	15
5	Credit Evaluations	27	15	Information Management	15
6	Machine Learning	27	16	Learning Algorithms	15
7	Risk Management	25	17	Neural Networks	15
8	Learning Systems	22	18	Credit Risk Assessment	14
9	Credit Risk	21	19	Credit Risk Management	14
10	Decision Trees	21	20	Credit Ratings	13

access documents. All subject areas are included in the search, mostly limited to articles, reviews, conference proceedings, conference reviews, editorials, and book chapters. The search also includes all source types. Finally, the language of the searched articles is limited to English.

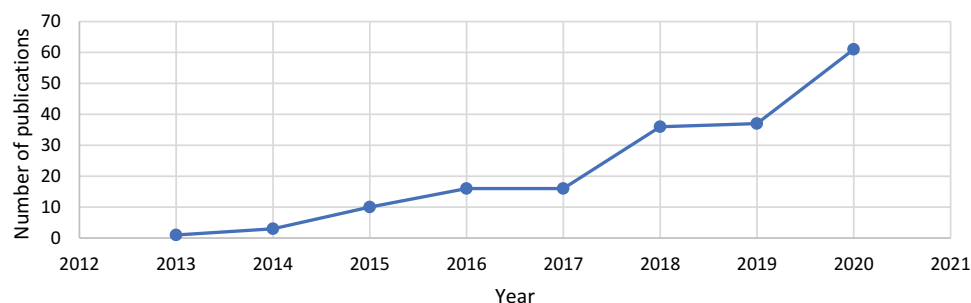
4. Results and discussions

This section presents the Scopus bibliometric findings for the keywords used for studies published between 2012 and 2021 (July).

Following the method system of Apriliyanti and Alon (Apriliyanti & Alon, 2017) and Fetscherin and Heinrich (Fetscherin & Heinrich, 2015), the study also employs a bibliometric software program to ease the process of finding the citation and co-citation links of publications. VOSviewer is used to verify the findings, which may also be used to “construct and visualize bibliometric maps” (Van Eck & Waltman, 2014). VOSviewer offers distance-based representations of bibliometric systems using the VOS (visualization of the similarity) mapping approach (Van Eck & Waltman, 2014).

As Figure 1 shows that during the past nine years (2012–2021 (July)), there has been an increase in the overall quantity of credit risk and big data research publications. In 2013, nearly two publications were published, which is considered insignificant. However, it increased to 10 new publications in 2015, indicating a positive sign for this research topic. Interestingly, 2018, only three years from 2015, has 40 new publications on this topic, which is four times the number of publications in 2015. Moreover, in 2020, 60 new publications were added to the wealth of

Figure 1. Credit risk and big data publications in the last nine years.



knowledge, indicating an upward trend expected to grow further as big data and credit risk assessment is a new emerging topic attracting scholars worldwide.

Hence, this study adds to the classification and synthesis of this huge body of literature, considering the increasing data and literature on credit risk and big data and recommending potential research fields for future study of the subject.

Figure 2 presents the top sponsored institutes/organizations and the number of articles on credit risk assessment and big data published by these institutes. This study will aid potential researchers in discovering organizations interested in the selected topic. Hence, scholars can attempt to receive support from these institutions when they have a great idea but require a lot of professional and financial support. Likewise, postgraduate students can reach those entities if they wish to work and specialize in related fields.

Thus, the figure clearly shows that the greatest number of articles (34) on credit risk assessment and big data is published by the National Natural Science Foundation of China, with a huge difference compared to other entities. However, Chinese sponsors and organizations are the most dominant in credit risk assessment and related research.

Figure 2. Shows the number of articles published by different organizations.

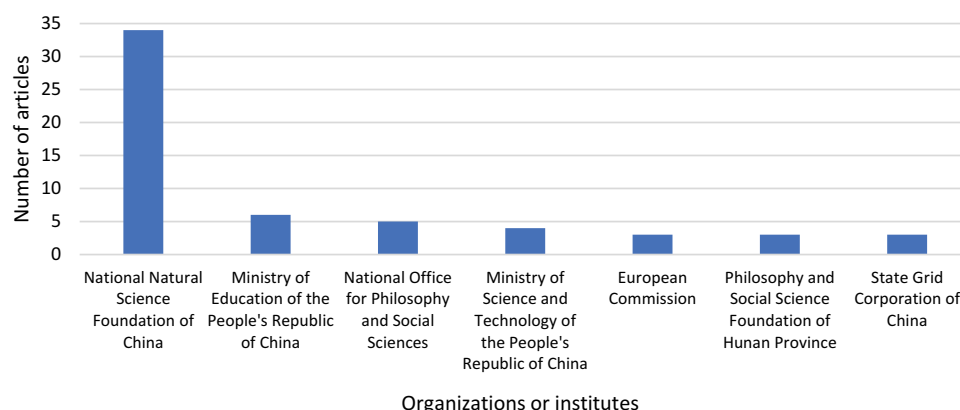


Table 3. Leading sources by citations

Rank	Source	Documents	Citations
1	Electronic Commerce Research and Applications	3	96
2	Policy and Internet	1	66
3	Financial Innovation	3	59
4	Risks	1	47
5	Competition and Change	1	36
6	2016 3rd International Conference on Artificial Intelligence and Pattern Recognition, AIPR 2016	2	29
7	Industrial Management and Data Systems	1	28
8	Journal of Business Research	1	27

Table 3 shows the most-cited journals in all search categories, as it will help future scholars target the correct journals that are specialized in this field and are most interested in the advancements made regarding this topic. Additionally, they can use it to refer to the most cited studies as it may have useful information. The results show that a total of eight journals ranked from the highest number of citations to the lowest. It was found that the journal “Electronic Commerce Research and Applications” has the largest number of citations in the required search category. Three relevant articles with 96 citations are found in the journal. The journal “Policy and Internet” has only one document in the search query, but it is cited 66 times. Similarly, the “Journal of Business Research” ranks lowest, with article citations of 27.

Table 4 presents the journals with the highest number of published articles in the required search categories. The journals are ranked from top to bottom, with the highest number of studies at the top and the lowest number at the bottom. Interestingly, it is found that the greatest number of studies related to credit risk assessment and big data are published in the “Journal of Physics: Conference Series.” This is an open-access Institute of Physics (IOP) Publishing peer-reviewed magazine that gives readers new advances in physics at worldwide conferences. This journal published a total of 15 documents with several citations of 18. Similarly, other journals with several published articles are listed in Table 2.

However, surprisingly, the journals with the highest number of published studies are not mentioned in the table of the top-cited journals, revealing that sometimes quality could outway quantity.

VOS viewer is used with a threshold of 50 documents from seven different journals to map the leading sources of the research articles. A network map of the leading sources is shown in Figure 3. The size of the circle, as in historiographic mapping, determines the journal’s relevance; the larger the circle, the greater the journal’s influence.

Table 5 shows the leading authors in credit risk assessment and big data-related studies by the number of documents and citations. Identifying leading authors in big data and credit risk assessment will open the way for future researchers willing to publish in this regard to communicate and collaborate with influential authors in this area. These researchers could also use the

Table 4. Leading sources by number of documents

Rank	Source	Documents	Citations
1	Journal of Physics: Conference Series	15	18
2	ACM International Conference Proceeding Series	10	7
3	Advances in Intelligent Systems and Computing	10	4
4	Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	7	12
5	Electronic Commerce Research and Applications	3	96
6	Financial Innovation	3	59
7	Information Systems and E-Business Management	3	8

Figure 3. Network mapping of the leading sources.

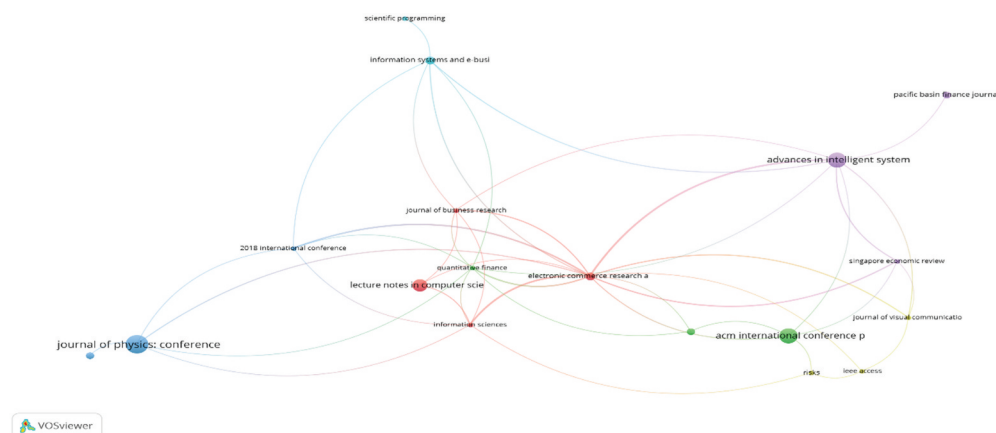


Table 5. Leading authors by the number of articles and citations

Rank	Author	Documents	Citations
1	Zhang Y.	6	32
2	Li X.	4	20
3	Hayashi Y.	3	39
4	Song X.	3	28
5	Wang H.	3	17
6	Yang X.	3	16
7	Yu Y.	3	83
1	Yu Y.	3	83
2	Ma X.	2	77
3	Yang Q.	2	76
4	Niu X.	1	76
5	Sha J.	1	76
6	Wang D.	1	76

wealth of knowledge the leading authors have in this particular field and others. Hence, it was observed that Zhang published the highest number of research articles (six). The total number of citations in these documents is only 32. Similarly, [Table 4](#) also shows that the author Yu Y has the greatest number of research article citations with 83 citations. However, he had only published three articles emphasizing the point that quality is what matters in the end. Other leading authors, by the number of articles and citations, are shown in [Table 5](#). Similarly, [Figure 4](#) shows the

Figure 4. Shows network mapping of the leading authors.

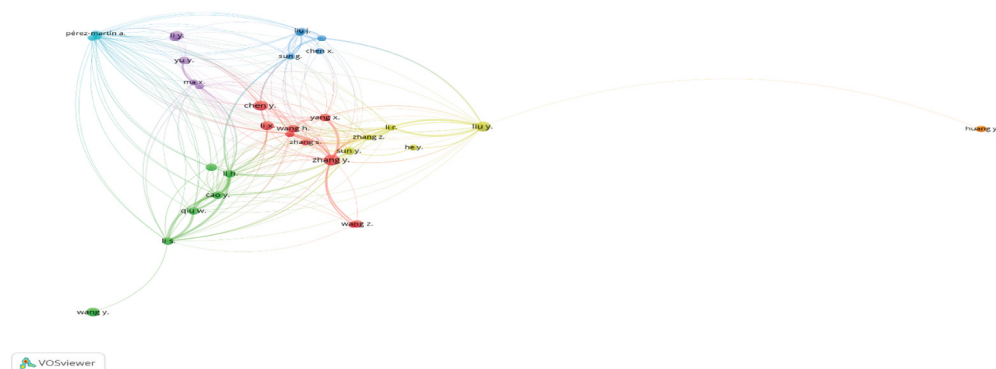


Table 6. Leading affiliations by citations

Rank	Organization	Documents	Citations
1	Asia Australia Business College, Liaoning University	1	76
2	Harbin Institute of Technology	1	76
3	School of Statistics, Dongbei University of Finance and Economics	1	76
4	Department of Communication Studies, University of Michigan	1	66
5	Department of Political Science, University of Michigan	1	66
6	Ford School of Public Policy, University of Michigan	1	66
7	ISR Center for Political Studies, University of Michigan	1	66

VOSviewer network mapping of the leading authors. It also visualizes leading authors by the number of articles and citations.

Table 6 reveals the institutions or affiliations with the highest citations in credit risk assessment and big data research. The table ranks institutes from the highest to lowest number of citations. The most cited research article is published by the Institute of “Asia Australia Business College, Liaoning University.” A research article published by the Asia Australia Business College, Liaoning University, is cited 76 times. This institution can be considered a “center of excellence for prior credit risk assessment and big data research.” While Figure 5 shows the VOSviewer networking map with a threshold of 13 affiliations that show that the most influential institute/affiliation is Henan Key Laboratory of Finance. The majority of major institutions are located in China and the United States.

Figure 5. Shows network mapping of the most influential affiliations.

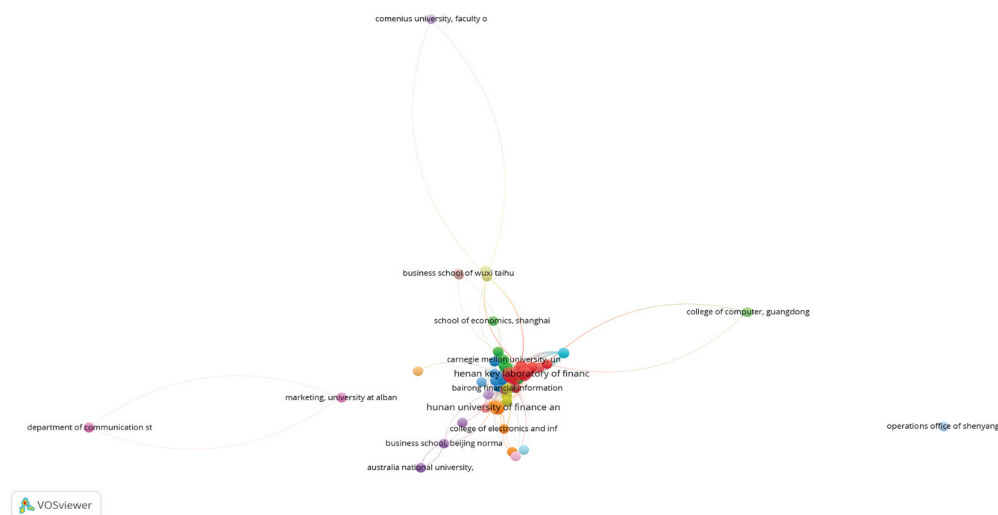


Table 7. Shows the leading countries by number of documents and by citations

Rank	Country	Documents	Citations	Rank	Country	Documents	Citations
1	China	136	348	1	China	136	348
2	United States	22	53	2	Viet Nam	1	173
3	United Kingdom	10	22	3	Hong Kong	4	58
4	India	9	46	4	Italy	4	55
5	Japan	5	40	5	United States	22	53
6	Taiwan	5	48	6	Taiwan	5	48
7	Canada	4	41	7	France	1	47
8	Hong Kong	4	58	8	India	9	46
9	Italy	4	55	9	Canada	4	41

Table 7 presents the leading countries regarding several documents and citations in credit risk assessment and big data research. Highlighting the top countries publishing on big data and credit risk assessment will aid scholars in selecting countries that are currently interested and working. Moreover, it could open the way for new graduates who wish to continue their studies in this area to look for opportunities in these countries, as they already have sufficient knowledge in this area. Therefore, Table 7 shows that China leads in studies published and documents cited. This shows that China's research on credit risk assessment is quite extensive, and China is leading the world in credit risk and big data research, as can also be seen in Figure 6, VOSviewer networking map. The big circle indicates that China is highly influential in credit risk and big data research. Similarly, the United States, the United Kingdom, Vietnam, and India are some other countries alongside China that are influential in credit risk assessment and big data study (Figure 6).

Table 8 presents the most-cited articles related to credit risk assessment and big data research. This type of analysis will give potential researchers influential documents with the most relevant data regarding the topic to gain insights on big data and credit risk assessment and identify where to start. These are the most cited articles by other authors, which will guide other scholars in researching big data and credit risk. Shen et al. (Shen et al., 2018) is the most cited article with 76 citations, revealing that the study has significant and useful information regarding the topic and that 76 other authors can benefit from it until now.

Figure 7 shows the VOSviewer network map of the most cited articles, and it also indicates that Liang et al. (Liang et al., 2018) article is the second with 66 citations. This indicates that this article has a high impact factor and other researchers find it relevant and useful for their research.

5. Limitations and future research

One of the weaknesses of this study is that the Scopus database does not include all relevant articles compared with other scientific databases (such as Google Scholar). It is also possible that many relevant publications could not be incorporated into our study. While the Scopus database is regarded as the most selective, it is intended to concentrate on certified publications that have proven brilliance in quality and impact. Although this technique is being utilized, there is some subjectivity in the assemblage of main research streams since the authors make certain judicial decisions. The current study is also restricted to articles published in English. It is recommended that future researchers include publications and articles written in languages other than English. Similarly, the current research only considered articles published between 2012 and 2021 (July); therefore, it is also recommended to include articles and studies before 2012 to obtain a more comprehensive bibliometric analysis of credit risk assessment and big data.

Figure 6. Network mapping of the most influential countries.

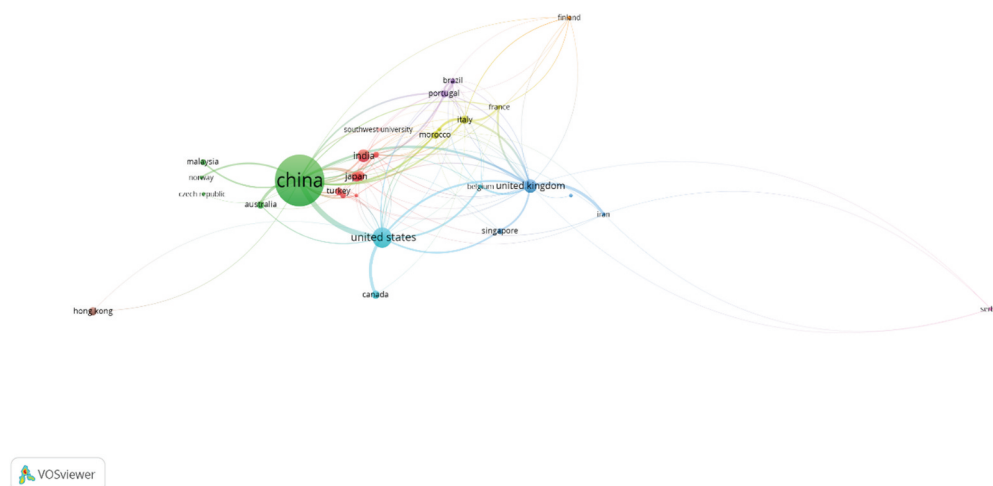


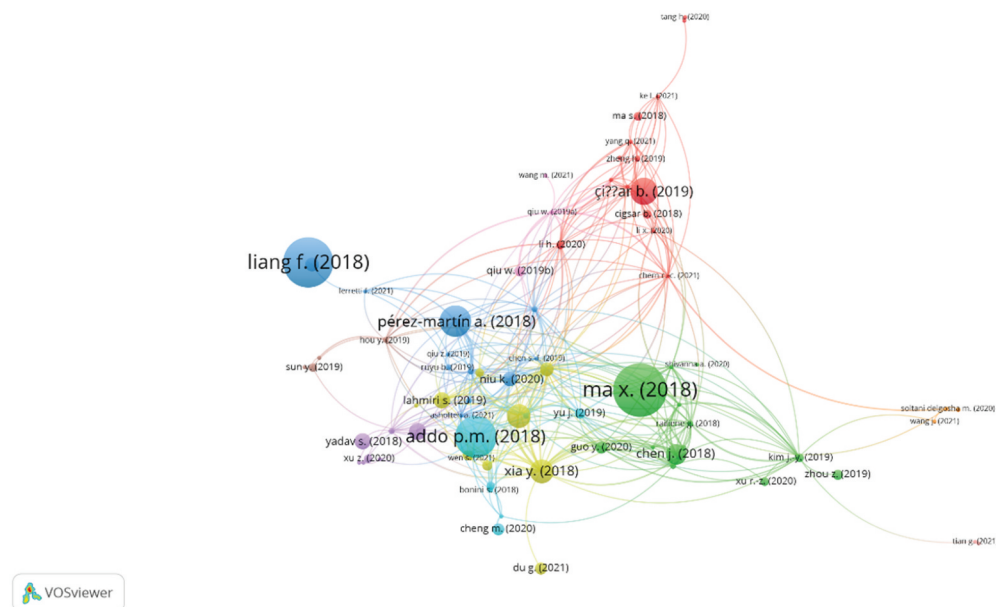
Table 8. Most cited documents

Rank	Document	Citations	Rank	Document	Citations
1	(Shen et al., 2018)	76	11	(Çiğşar & Ünal, 2019)	20
2	(Liang et al., 2018)	66	12	(Hayashi et al., 2016)	18
3	(Addo et al., 2018b)	47	13	(Zhang et al., 2015)	16
4	(Aitken, 2017)	36	14	(Xia et al., 2018)	15
5	(Yan et al., 2015)	30	15	(Zhu et al., 2018)	15
6	(Pérez-Martin et al., 2018b)	27	16	(Saabni, 2016)	15
7	(Favilukis et al., 2015)	27	17	(Botev, 2017)	15
8	(Agarwal, Chomsisengphet, Liu, song & Souleles, 2018)	25	18	(Tsao et al., 2017)	14
9	(Pan et al., 2021)	24	19	(Turkson et al., 2016)	14
10	(Hayashi, 2016)	21	20	(Lin & Shen, 2015)	13

6. Conclusions and recommendations

This study uses bibliometric citation analysis to examine credit risk assessment and big data research during the previous nine years. This study used different search queries and reviewed

Figure 7. Shows network mapping of the most cited documents.



219 documents. Many interesting findings emerged from this study. First, the results suggest that the organization mostly contributing to financing credit risk and big data studies is the National Natural Science Foundation of China. The results also suggest that the greatest number of studies are published in the Journal of Physics: Conference Series, whereas the most cited studies associated with credit risk assessment and big data research, are published in the Electronic Commerce Research and Applications Journal.

The results also show that the greatest number of studies associated with credit risk and big data research was published by Zhang Y., and the author whose articles are most cited is Yu Y. The results show that the leading institute in credit risk and big data study is the Asia Australia Business College, Liaoning University in China. When comparing the number of research articles from different regions and countries, China is the leading country in researching credit risk assessment and big data. While developed countries have dominated credit risk and big data research in the last few decades, China has emerged as the leading credit risk and big data research country in the last nine years. Thus, it can be concluded that Chinese authors, institutes, and organizations have dominated credit risk assessment, big data, and related research in the last few years. It is also concluded from the current study that there are several possibilities to improve the knowledge of credit risk assessment and big data while also advancing theories and their influence on financial investments.

Hence, these results can be used in future research as a guide for potential scholars to identify significant documents, journals, countries, and organizations. Moreover, this study paves the way for future scholars to enhance the practicality of the research studies offered in the literature review section and top document area.

Funding

The authors received no direct funding for this research.

Author details

Haitham Nobanee^{1,2,3}

E-mail: haitham.nobanee@liverpool.ac.ae

ORCID ID: <http://orcid.org/0000-0003-4424-5600>

Hiba Shanti¹

Hind Aldhanhani¹

Abdulrahman Alblooshi¹

Essa Alali¹

¹ Department of Accounting & Finance, College of Business, Abu Dhabi University, Abu Dhabi, United Arab Emirates.

² Oxford Center for Islamic Studies, University of Oxford, Oxford, UK.

³ Faculty of Humanities & Social Sciences, University of Liverpool, Liverpool, UK.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Citation information

Cite this article as: Big data and credit risk assessment: a bibliometric review, current streams, and directions for future research, Haitham Nobanee, Hiba Shanti, Hind Aldhanhani, Abdulrahman Alblooshi & Essa Alali, *Cogent Economics & Finance* (2022), 10: 2132638.

References

- Addo, P., Guegan, D., & Hassani, B. (2018a). Credit risk analysis using machine and deep learning models. *Risks*, 6(2), 38. <https://doi.org/10.3390/risks6020038>
- Addo, P., Guegan, D., & Hassani, B. (2018b). Credit risk analysis using machine and deep learning models. *Risks*, 6(2), 38. <https://doi.org/10.3390/risks6020038>
- Agarwal, S., Chomsisengphet, S., Liu, C., Song, C., & Souleles, N. S. (2018). Benefits of relationship banking: Evidence from consumer credit markets. *Journal of Monetary Economics*, 96, 16–32. <https://doi.org/10.1016/j.jmoneco.2018.02.005>
- Aitken, R. (2017). ‘All data is credit data’: Constituting the unbanked. *Competition & Change*, 21(4), 274–300. <https://doi.org/10.1177/1024529417712830>
- Altman, E. I., & Saunders, A. (1997). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21(11–12), 1721–1742. [https://doi.org/10.1016/S0378-4266\(97\)00011-1](https://doi.org/10.1016/S0378-4266(97)00011-1)
- Anderson, E. J. (2013). *Business risk management: models and analysis*. John Wiley & Sons.
- Anson, M. J. P., Fabozzi, F. J., Choudhry, M., & Chen, -R.-R. (2004). *Credit derivatives: Instruments, applications, and pricing*. John Wiley & Sons.
- Apriliyanti, I. D., & Alon, I. (2017). Bibliometric analysis of absorptive capacity. *International Business Review*, 26(5), 896–907. <https://doi.org/10.1016/j.ibusrev.2017.02.007>
- Benzschawel, T. (2012). Credit risk modelling—facts, theory and applications. *Risk Books*. DuraSpace. <https://lib.hpu.edu.vn/handle/123456789/22628>
- Black, F., & Scholes, M. (1972). The valuation of option contracts and a test of market efficiency. *The Journal of Finance*, 27(2), 399. <https://doi.org/10.2307/2978484>
- Botev, Z. I. (2017). The normal law under linear restrictions: Simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1), 125–148. <https://doi.org/10.1111/rssb.12162>
- Broadus, R. N. (1987). Toward a definition of “bibliometrics”. *Scientometrics*, 12(5–6), 373–379. <https://doi.org/10.1007/BF02016680>
- Caouette, J. B., Altman, E. I., Narayanan, P., & Nimmo, R. (2011). *Managing credit risk: The great challenge for global financial markets*. John Wiley & Sons.
- Çiğşar, B., & Ünal, D. (2019). Comparison of data mining classification algorithms determining the default risk. *Scientific Programming*, 2019, 1–8. <https://doi.org/10.1155/2019/8706505>

- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- Cui, D. (2015). Financial credit risk warning based on big data analysis. *Metallurgical and Mining Industry*, 7(6), 133–141.
- Diaz Weigel, D., & Gemmill, G. (2006). What drives credit risk in emerging markets? The roles of country fundamentals and market co-movements. *Journal of International Money and Finance*, 25(3), 476–502. <https://doi.org/10.1016/j.jimonfin.2006.01.006>
- Duffie, D., & Singleton, K. J. (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies*, 12(4), 687–720. <https://doi.org/10.1093/rfs/12.4.687>
- Du, G., Liu, Z., & Lu, H. (2021). Application of innovative risk early warning mode under big data technology in internet credit financial risk assessment. *Journal of Computational and Applied Mathematics*, 386, 113260. <https://doi.org/10.1016/j.cam.2020.113260>
- Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, 105(3), 1809–1831. <https://doi.org/10.1007/s11192-015-1645-z>
- Favilukis, J., Lin, X., & Zhao, X. (2015). The Elephant in the room: The impact of labor obligations on credit risk. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2648763>
- Fetscherin, M., & Heinrich, D. (2015). Consumer brand relationships research: A bibliometric citation meta-analysis. *Journal of Business Research*, 68(2), 380–390. <https://doi.org/10.1016/j.jbusres.2014.06.010>
- Fetscherin, M., & Usunier, J. (2012). Corporate branding: An interdisciplinary literature review. *European Journal of Marketing*, 46(5), 733–753. <https://doi.org/10.1108/03090561211212494>
- Hayashi, Y. (2016). Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective. *Operations Research Perspectives*, 3, 32–42. <https://doi.org/10.1016/j.orp.2016.08.001>
- Hayashi, Y., Tanaka, Y., Takagi, T., Saito, T., Iiduka, H., Kikuchi, H., Bologna, G., & Mitra, S. (2016). Recursive-rule extraction algorithm with j48graft and applications to generating credit scores. *Journal of Artificial Intelligence and Soft Computing Research*, 6(1), 35–44. <https://doi.org/10.1515/jaiscr-2016-0004>
- Jarrow, R. A., & Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50(1), 53–85. <https://doi.org/10.1111/j.1540-6261.1995.tb05167.x>
- Kealhofer, S. (2003). Quantifying credit risk I: Default prediction. *Financial Analysts Journal*, 59(1), 30–44. <https://doi.org/10.2469/faj.v59.n1.2501>
- Kim, J., & McMillan, S. J. (2008). Evaluation of internet advertising research: A bibliometric analysis of citations from key sources. *Journal of Advertising*, 37(1), 99–112. <https://doi.org/10.2753/JOA0091-3367370108>
- Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). Constructing a data-driven society: China's social credit system as a state surveillance infrastructure: China's social credit system as state surveillance. *Policy & Internet*, 10(4), 415–453. <https://doi.org/10.1002/poi3.183>
- Lin, Y.-M., & Shen, C.-A. (2015). Family firms' credit rating, idiosyncratic risk, and earnings management. *Journal of Business Research*, 68(4), 872–877. <https://doi.org/10.1016/j.jbusres.2014.11.044>
- Md, H., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, 7(1), 21. <https://doi.org/10.1186/s40537-020-00291-z>
- Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1), 141. <https://doi.org/10.2307/3003143>
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2), 449. <https://doi.org/10.2307/2978814>
- Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. *International Journal of Financial Studies*, 9(3), 39. <https://doi.org/10.3390/ijfs9030039>
- Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48(3), 261–297.
- Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 36(2), 3028–3033. <https://doi.org/10.1016/j.eswa.2008.01.018>
- Nobanee, H. (2020). Big data in business: A bibliometric analysis of relevant literature. *Big Data*, 8(6), 459–463. <https://doi.org/10.1089/big.2020.29042.edi>
- Nobanee, H. (2021). A bibliometric review of big data in finance. *Big Data*, 9(2), 73–78. <https://doi.org/10.1089/big.2021.29044.edi>
- Nobanee, H., Alhajjar, M., Alkaabi, M. A., Almemari, M. M., Alhassani, M. A., Alkaabi, N. K., and AlBlooshi, H. H. (2021). A bibliometric analysis of objective and subjective risk. *Risks*, 9(7), 128. <https://doi.org/10.3390/risks9070128>
- Pan, W.-F., Wang, X., Wu, G., & Xu, W. (2021). The COVID-19 pandemic and sovereign credit risk. *China Finance Review International*, 11(3), 287–301. <https://doi.org/10.1108/CFRI-01-2021-0010>
- Pérez-Martín, A., Pérez-Torregrosa, A., & Vaca, M. (2018a). Big data techniques to measure credit banking risk in home equity loans. *Journal of Business Research*, 89, 448–454. <https://doi.org/10.1016/j.jbusres.2018.02.008>
- Pérez-Martín, A., Pérez-Torregrosa, A., & Vaca, M. (2018b). Big data techniques to measure credit banking risk in home equity loans. *Journal of Business Research*, 89, 448–454. <https://doi.org/10.1016/j.jbusres.2018.02.008>
- Saabni, R. (2016). Recognizing handwritten single digits and digit strings using deep architecture of neural networks. 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), 1–6. Lodz: IEEE. <https://doi.org/10.1109/ICAIPR.2016.7585206>
- Saunders, A., & Cornett, M. M. (2017). *Financial institutions management: A risk management approach*. McGraw-Hill Education. <https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=5049851>
- Shen, F., Ma, X., Li, Z., Xu, Z., & Cai, D. (2018). An extended intuitionistic fuzzy TOPSIS method based on a new distance measure with an application to credit risk evaluation. *Information Sciences*, 428, 105–119. <https://doi.org/10.1016/j.ins.2017.10.045>
- Skute, I., Zalewska-Kurek, K., Hatak, I., & de Weerd-Nederhof, P. (2019). Mapping the field: A bibliometric analysis of the literature on university-industry collaborations. *The Journal of Technology Transfer*, 44(3), 916–947. <https://doi.org/10.1007/s10961-017-9637-1>

- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799–813. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:9<799::AID-ASIS9>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASIS9>3.0.CO;2-G)
- Sobehart, J. R., & Keenan, S. C. (2001). A practical review and test of default prediction models. *RMA JOURNAL*, 84(3), 54–59.
- Thomas, L. C., Oliver, R. W., & Hand, D. J. (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, 56(9), 1006–1015. <https://doi.org/10.1057/palgrave.jors.2602018>
- Tsao, Y.-C., Lee, P.-L., Chen, C.-H., & Liao, Z.-W. (2017). Sustainable newsvendor models under trade credit. *Journal of Cleaner Production*, 141(10), 1478–1491. <https://doi.org/10.1016/j.jclepro.2016.09.228>
- Turkson, R. E., Baagyere, E. Y., & Wenya, G. E. (2016). A machine learning approach for predicting bank credit worthiness. 2016 *Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*, 1–7. Lodz: IEEE. <https://doi.org/10.1109/ICAIPR.2016.7585216>
- van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 285–320). Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_13
- Van Raan, A. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *TATuP - Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis*, 12(1), 20–29. <https://doi.org/10.14512/tatup.12.1.20>
- van Raan, A. F. J. (2004). Measuring science: Capita selecta of current main issues. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 19–50). Springer Netherlands. https://doi.org/10.1007/1-4020-2755-9_2
- Verikas, A., Kalsyte, Z., Bacauskiene, M., & Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: A survey. *Soft Computing*, 14(9), 995–1010. <https://doi.org/10.1007/s00500-009-0490-5>
- Wagdi, O., & Tarek, Y. (2022). The of integration big data and artificial neural networks for enhancing credit risk scoring in emerging markets: Evidence from Egypt. *International Journal of Economics and Finance*, 14(2), 32. <https://doi.org/10.5539/ijef.v14n2p32>
- Wang, F., Ding, L., Yu, H., & Zhao, Y. (2020). Big data analytics on enterprise credit risk evaluation of e-Business platform. *Information Systems and E-Business Management*, 18(3), 311–350. <https://doi.org/10.1007/s10257-019-00414-x>
- Xia, Y., Liu, C., Da, B., & Xie, F. (2018). A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Systems with Applications*, 93(1), 182–199. <https://doi.org/10.1016/j.eswa.2017.10.022>
- Yan, J., Yu, W., & Zhao, J. L. (2015). How signaling and search costs affect information asymmetry in P2P lending: The economics of big data. *Financial Innovation*, 1(1), 19. <https://doi.org/10.1186/s40854-015-0018-1>
- Zhang, S., Xiong, W., Ni, W., & Li, X. (2015). Value of big data to finance: Observations on an internet credit service company in China. *Financial Innovation*, 1(1), 17. <https://doi.org/10.1186/s40854-015-0017-2>
- Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms*. Chapman and Hall/CRC. <https://doi.org/10.1201/b12207>
- Zhu, B., Yang, W., Wang, H., & Yuan, Y. (2018). A hybrid deep learning model for consumer credit scoring. 2018 *International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 205–208. Chengdu: IEEE. <https://doi.org/10.1109/ICAIBD.2018.8396195>



© 2022 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

***Cogent Economics & Finance* (ISSN: 2332-2039) is published by Cogent OA, part of Taylor & Francis Group.**

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

