**Review Article**

**Review of literature on corporate governance and firm performance: A Text mining approach**

**ABSTRACT**

This article synthesises existing literature on “corporate governance” (CG) and its impact on “firms’ performance.” The study followed the PRISMA framework, analysing 442 Scopus-indexed journal documents. The paper critically examines intersection of CG and firm performance through the lens of text mining. We have used Latent Dirichlet Allocation (LDA) algorithm for topic modelling. After topic modelling analysis, from a total of 442 unique documents’ metadata, we identified 10 latent topics in our paper. Out of which we discarded four irrelevant topics from the analysis since they were not associated to Corporate Governance which is our field of study. So, our selected topics were effects of CG on FP, CG disclosure, Agency theory, Ownership Structure, Board Size and Board Diversity. These selected six topics represents a distinct theme with assigned probabilities. The probabilities demonstrate relative importance of each token to its respective topic. This analysis reveals the comprehensive intellectual foundation of the CG field. The study adds to the literature by identifying six relevant topics of academic discussion. This study provides academics and practitioners with a road map for targeted investigation and strategic implementation in enhancing firm performance by highlighting the most important themes in corporate governance research. To our knowledge, very few prior studies has used the LDA approach in reviewing corporate governance and firm performance literature.

Keywords: Corporate Governance, CG, Firm Performance, Text Mining, LDA

**INTRODUCTION**

Corporate Governance describes the relationships between these groups that allow corporate entities to thrive and deal with a wide range of economic, financial, and legal topics (Mallin, 2019). Interest in governance topics has grown substantially since a decade
(Peters, 2001). Corporate governance is described as having legitimacy, accountability, and competence in the realm of policy and delivery of services by simultaneously respecting the law and human rights (Jamali, D et al., 2008), (Tricker, 2015).  However, governance can range from effective to ineffective, shaped by the practices followed and the value it upholds. Corporate governance and firm performance are intricately linked concepts that have gained significant attention from academics, policymakers, and practitioners (Agarwal, M., & Gupta, M.B. 2024). As said above, corporate governance encompasses the framework of rules, practices, and processes through which organizations are directed and controlled, with the primary aim of balancing the interests of stakeholders (Jamali, et al., 2008). Firm performance, on the other hand, reflects the outcomes of these governance mechanisms, often assessed through metrics such as profitability, market valuation, and operational efficiency (Ward et al., 2009). While the theory behind this relationship is well-known, the actual evidence is unclear, showing the need for new methods to better understand their connection.

This review critically examines the intersection of “CG and performance of firms” through the lens of text mining. Text mining, is a data-driven approach that extracts patterns and insights from unstructured textual data, offers the potential to uncover latent themes and sentiments in corporate disclosures, board reports, and regulatory filings (Schöne, 2024).  By analysing large volumes of textual data, researchers can gain a deeper understanding of governance dynamics and their implications for firm performance, capturing nuances that traditional quantitative methods may overlook (Sivarajah et al., 2017) (Fouad, 2024). New innovations and advancement in text mining have opened new avenues for exploring the association of “CG and firm performance”.

**Impact of Corporate Governance on Firm Performance**

Corporate governance significantly impacts firm performance by ensuring effective oversight, reducing agency conflicts, and fostering transparency (Sari, 2023). Sound governance systems, featuring independent boards, performance-linked executive compensation, and high-quality audits, enhance decision-making and align management goals with shareholder interests (Chandrakala et al.,2024), (Ramezanian et al.,2025). Some studies show that firms with good governance frameworks tend to achieve higher profitability, better market valuations, and sustainable growth. (Core et al. 2006)

In the UK and USA there were several companies, which collapsed unexpectedly in the 1980s and 1990s. (Kariyawasam, 2010). Financial reporting irregularities or inadequate internal controls and risk management were analysed as the causes of these corporate failures (Spira& Page, 2003).

In general, corporate governance is seen as playing an important role in regulating the economic growth potential of a country. Proper corporate governance practices reduce risk for investors, attract investment capital and improve performance of companies (Spanos, 2005). However, the effectiveness of governance mechanisms varies across regions, industries, and firm-specific factors, influenced by legal, cultural, and economic contexts (Ernstberger & Grüning, 2013). However, a good governance is essential for long-term business success and for gaining investors’ confidence.

 **Methodology**

This study is a systematic review of literature which follows PRISMA 2020 guidelines, A systematic search was conducted in Scopus, and the selection process involved, identification, screening, eligibility, and inclusion. A metadata of 442 articles was analysed that are extracted from Scopus for a period covering 14 years i.e. 2010-2024. We started from 859 documents; after screening the data we get a corpus of 442 documents for our study. Our goal in this paper is to discover the hidden topics from metadata of documents by applying topic modelling technique. Topic modelling is a method used to find hidden topics in a large set of documents (Egger, 2022). It shows each document as a mix of topics and each topic as a group of related words, helping to uncover patterns in the text (Andres, et al. 2024). While "topics" are often defined in a technical or statistical methods, the final step involves experts’ evaluation to extract and interpret distinct themes. These themes are then given simple, meaningful names that make sense in a real-world context.

For this task, we use the Latent Dirichlet Allocation (LDA) due to its advantages over other topic modelling methods. While approaches like Latent Semantic Analysis (Landauer et al. 2013) classify topics based on word co-occurrence frequencies, and models such as Naive Bayes or Support Vector Machines rely on linking word probabilities to topic probabilities, LDA offers a more comprehensive and refined approach (Maier, et al. 2021). This is explained in the paper further.

We also applied ENTREQ statement in our synthesis to disclose material facts related to study. Table 1 depicted ENTREQ statement which disclose 18 different aspects of study.

Table 1: “ENTREQ” statement

|  |  |  |
| --- | --- | --- |
| No | Item | Guide and description |
| 1 | Aim | 1. To identify the key areas of focus in scholarly works and articles concerning corporate governance?
2. To identify the hidden topics in the corpus.
3. To identify the corporate governance relate to associated topics.
 |
| 2 | Synthesismethodology | Our study is a systematic review of literature following PRISMA guidelines, analysing meta data of 442 articles extracted from Scopus. |
| 3 | Approach tosearching | The data has been taken from Scopus database and a rigorous filtration has been done thereafter. |
| 4 | Inclusion criteria | The study is critical review of meta data of Scopus indexed documents. We applied following filters:1. Time: 2010-2024.
2. Subject area: Business, management, and accounting stream.
3. Document type: Journal articles.
4. Language: English
 |
| 5 | Data sources | Scopus database.  |
| 6 | Electronic Searchstrategy | We have constructed following search string to get the data.“Corporate governance” OR “CG” AND “financial performance’ |
| 7 | Study screeningmethods | Title, abstract and authors keywords are considered while deciding relevance of the document for the study. |
| 8 | Study characteristics | Study in qualitative in nature. Text mining approach has been adopted for analysis.  |
| 9 | Study selectionresults | For selection of relevant study, we critically review the meta data of 859 studies and we removed 417 irrelevant studies.  |
| 10 | Rationale forappraisal |  For the study we have relied on Scopus database due to its credibility across globe. For the selection of topics, we have used latent Dirichlet approach since it is highly versatile and applicable numerous domains and datasets. Also, we have not blindly relied on machine learning models, we have we performed a qualitative check using expert human evaluation to ensure that the identified topics are meaningful. |
| 11 | Appraisal items | To get to the findings of the study we have used PRISMA model. Further we have applied topic modelling analysis to to identify hidden themes in large text sets of data given. Under topic modelling we have used LDA latent Dirichlet allocation approach seeing it’s advantages over others. |
| 12 | Appraisal process | The appraisal process was completed by the researcher herself under the guidance of Ph.D. supervisor. They have pooled their knowledge and experience for the calculation and evaluation of appraisal results. |
| 13 | Appraisal results | In this we have performed a search string using PRISMA. Further we applied LDA for the calculation and evaluation of results. The results were as follows. Our goal was to identify the hidden or unobservable topics in the corpus of documents using topic modelling technique.From a total of 442 unique documents, we have identified 10 latent topics. Each topic represents a distinct theme, and the listed tokens are those most strongly associated with that theme, as indicated by their probabilities. Out of those 10 topics we have discarded 4 of them since it did not associate with our topic corporate governance.  |
| 14 | Data extraction | All the data was extracted using electronic devices that is computer. Various websites were used for the extraction of data. Most data are taken from Scopus database. Many journals, articles and websites were used for data collection. |
| 15 | Software | In our study we applied PRISMA Protocol software for literature reviewing. |
| 16 | Number ofreviewers | For the analysis of data, the researcher along with her research supervisor has completed the review process. |
| 17 | Quotations | Corporate governance and firm performance are intricately linked concepts that have gained significant attention from academics, policymakers, and practitioners in the recent years. And hence, this study is done to critically examine the intersection of corporate governance and firm performance through the lens of text mining. |
| 18 | Synthesis output | From a total of 442 unique documents, we have identified 10 latent topics. Out of those 10 topics we have discarded 4 of them since it did not associate with our topic corporate governance. Our results highlight the topics that characterize different meta-disciplines. The results reveal the full intellectual structure of the corporate governance field. |

*Authors owns compilation*

**Data Collection**

For getting the metadata from the Scopus database, we constructed a search string i.e. “corporate governance” OR “CG” AND “financial performance". We have applied four automated filter related to time, area, document types and language to filtered out irrelevant documents. After extraction of metadata, we removed irrelevant document by manually observation.

**Fig 1-: PRISMA Framework)**

**Identification of studies through structured databases and research registers.**

Records removed before screening: Duplicate records removed: 8

Records removed for less information: 12

Records identified from Scopus

Database: 859

**Identification**

Records screened: 839

Records excluded: 358

**Screening**

Irrelevant Reports excluded: 39

Reports assessed for eligibility: 481

Studies included in review: 442

**Included**

A total of 859 records were initially identified from databases for the review. Before the screening process, 8 duplicate records and 12 records with insufficient information were eliminated. During the manual screening phase, 358 records were removed as they failed to meet the inclusion criteria. Subsequently, 481 reports were assessed for acceptability, out of which 39 were excluded as they were irrelevant. Ultimately, 442 studies were included in the final review, ensuring a systematic and thorough selection process.

**Topic modelling** (Certain repetitive lines have been removed from this section)

Recent progress in machine learning have led to topic modeling methods that can be used as an exploratory technique to identify hidden or latent relationships present in large data sets (Blei, 2012), (Asmussen et al., 2019). It is a discovery method that proves valuable for carrying out progressive and impactful research across disciplines. Our study aims to perform a topic modelling analysis to identify hidden themes in large text sets of corporate governance. Topic modelling is a probabilistic method used to uncover hidden semantic patterns, called topics, in a collection of unstructured documents (Kherwa & Bansal, 2019). These topics capture the essence of the semantic structure of documents (Wang et al., 1998). The method relies on the concept that certain words appear more frequently in a document due to their association with a specific topic. Topic modelling reveals these underlying semantic clusters by identifying groups of words that tend to co-occur in a document (Yi, et al., 2020), (Yang et al., 2018). This process involves calculating the probability distribution of each topic and the topic distribution per document, as well as the topic assignments per word in each document (Steyvers et al., 2024), (Vulić et al., 2015). Many different topic modelling algorithms are available for text mining and natural language processing research, such as Latent Dirichlet allocation (LDA), hierarchical Latent Dirichlet allocation (HLDA), hierarchical Dirichlet process (HDP), non-negative matrix factorization (NMF), Dirichlet multinomial regression (DMR), dynamic topic model (DTM) and correlated topic model (CTM) (Gurcan, & Cagiltay, 2022). (Chauhan et al., 2021). While algorithms such as NMF, CTM and DMR encounter challenges in determining the optimal number of topics through conventional consistency scores, newer models like HDP and HLDA offer automated mechanisms to ascertain the ideal topic count (Ozyurt, et al., 2024) (Jelodar et al., 2019). In contrast, the LDA model permits manual adjustment of the topic quantity (Gan, & Qi, 2021), an iterative process that improves the precision of topic number estimation and their semantic consistency (Liu, J et al., 2015). This flexibility, coupled with robust methods for computing coherence scores, solidifies LDA’s position as a preferred choice across various scholarly fields for its efficacy in semantic content analysis of extensive text collections (Kataishi, 2024), (Xie. et al.2024) (Lin at al., 2017). Therefore, LDA is frequently chosen for many studies and applications.

So, we have used the LDA algorithm for the topic modelling since it has its advantages over other simpler methods. As said above, “Latent Dirichlet Allocation” approach is a statistical method used for topic modelling. LDA helps uncover the hidden themes in a set of documents by analysing patterns in word usage. LDA analysis is useful for uncovering areas of emphasis in the corpus of articles that might not be discovered with regular bibliographic analysis (Kukreja, 2023), (O. Ozyurt et al. 2024). The subject distribution of the articles described using bibliometric analysis is clear and simple. LDA goes beyond citation analysis to uncover topics within the corpus that may be overlooked with traditional methods of analysis (Kushkowski et al. 2020), (Dawar et al. 2024).

The most practical advantage of LDA is its power to handle documents as a blend of multiple topics, with each topic represented as combination of words. Additionally, it is not necessary to label the documents. This makes it highly versatile and suitable to a wide range of domains and datasets. This flexibility is particularly useful in fields like corporate governance where different topics within a document may overlap. Also, LDA is easily scalable, as it handles large-scale datasets efficiently, which makes it valuable to fulfil our task at hand (Zhao et al.,2018), (Andres et al.,2024)

Though LDA provides numerous benefits, it comes with several challenges too. For instance, Identifying the optimal number of topics is not a simple task. To address this, several metrics such as Perplexity or Coherence are suggested in the literature. While increasing the number of topics typically enhances these statistical measures during topic modelling, it also raises processing cost of training the model. More importantly, as the number of topics increases, it becomes more complex for humans to understand their economic significance.

Another challenge with topic modelling is that topics that make sense to machine learning models might not necessarily align with human understanding. To address this, we perform a qualitative check using human expert judgment to ensure that the identified topics are meaningful within the context of existing corporate governance literature. This step helps us verify that the words associated with each topic are appropriately labelled and make sense from a human perspective.

Thus, keeping these challenges in mind we have relied on Latent Dirichlet Allocation model for our review paper. The LDA analysis was done using Python.

**Results of the study**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Topic 1  | Topic 2  | Topic 3  | Topic 4  | Topic 5  |
| Token | Probability | Token | Probability | Token | Probability | Token | Probability | Token | Probability |
| banks | 0.1832 | study | 0.1515 | islamic | 0.2764 | social | 0.1965 | study | 0.1732 |
| firms | 0.1324 | board | 0.1271 | banks | 0.2193 | csr | 0.1527 | firms | 0.145 |
| crisis | 0.1314 | risk | 0.1239 | study | 0.081 | esg | 0.1109 | agency | 0.1342 |
| firm | 0.1278 | management | 0.1142 | ibs | 0.0727 | firm | 0.1033 | family | 0.1101 |
| study | 0.1136 | reporting | 0.0949 | findings | 0.0668 | responsibility | 0.0901 | listed | 0.0871 |
| results | 0.0727 | impact | 0.0864 | bank | 0.066 | cg | 0.0812 | theory | 0.0779 |
| economic | 0.0643 | effect | 0.0786 | research | 0.0651 | study | 0.0688 | quality | 0.0698 |
| banking | 0.0613 | banks | 0.0753 | impact | 0.0534 | companies | 0.0679 | mechanisms | 0.0685 |
| bank | 0.059 | cg | 0.0745 | role | 0.0515 | environmental | 0.0644 | findings | 0.0682 |
| practices | 0.0545 | fp | 0.0736 | conventional | 0.0478 | disclosure | 0.0641 | research | 0.0659 |
|   |   |   |   |   |   |   |   |   |   |
|  Discarded   |  Effects of corporate governance on fp  |  discarded  |  CG disclosure  |  Agency theory  |
| Topic 6  | Topic 7 | Topic 8  | Topic 9  | Topic 10  |
| Token | Probability | Token | Probability | Token | Probability | Token | Probability | Token | Probability |
| study | 0.126 | ownership | 0.16 | financing | 0.1624 | board | 0.1729 | board | 0.293 |
| value | 0.123 | study | 0.152 | banks | 0.1422 | study | 0.1495 | firm | 0.1468 |
| indonesia | 0.1148 | companies | 0.1361 | risk | 0.1284 | firm | 0.1072 | firms | 0.095 |
| good | 0.1039 | board | 0.1245 | research | 0.1128 | size | 0.1031 | diversity | 0.0946 |
| disclosure | 0.1013 | firm | 0.0952 | activism | 0.1126 | firms | 0.1013 | directors | 0.078 |
| company | 0.0965 | firms | 0.0719 | effect | 0.0771 | impact | 0.0878 | study | 0.068 |
| effect | 0.0929 | findings | 0.0674 | innovation | 0.0752 | companies | 0.0749 | ceo | 0.0638 |
| banks | 0.0821 | structure | 0.0652 | islamic | 0.0661 | listed | 0.073 | gender | 0.0593 |
| gcg | 0.0818 | audit | 0.0639 | results | 0.0617 | relationship | 0.0657 | relationship | 0.0586 |
| companies | 0.0777 | research | 0.0638 | esg | 0.0616 | ownership | 0.0645 | structure | 0.0429 |
| Discarded | Ownership structure | Discarded | Board size | Board diversity |

*Authors own compilation*

*Table 2- The table presents the output of a topic modelling analysis*

From a total of 442 unique documents, we have identified 10 latent topics. Out of those 10 topics we have discarded 4 of them since it did not associate with our topic corporate governance. We selected only 10 topics because increasing the number of topics too much would result in the formation of low-frequency topics. which can lead to significant deviations from our research field. Conversely, reducing the number of topics may result in the risk of missing major topics in the research field. The table presents the output of a topic modelling analysis, where tokens are associated with specific topics and assigned probabilities. Each topic represents a distinct theme, and the listed tokens are those most strongly associated with that theme, as indicated by their probabilities (Airoldi et al. 2016). The probabilities reflect the likelihood of each token belonging to the respective topic.

In the study topic 1 focuses on terms like "banks "firms", and "crisis" which does not associate with our topic corporate governance and hence it has been discarded. Similarly, topic 2 includes terms such as "effects", "corporate governance" which could signify research on effects of corporate governance on firm performance.

The purpose of corporate governance is minimizing the gap between various stakeholders that have interests in the company and strengthen the investors’ confidence and reduce the agency costs. Various studies were conducted to examine the impact of corporate governance score on the performance of firms. The results of the studies were mixed in nature (Ahmed Haji et al.,2015). Various studies demonstrated that the governance score is significantly and positively correlated with Tobin’s Q, a key measure of firm performance. Many studies are in the view that better governance induces better performance (Bhagat et al., 2008). Several studies found that better corporate governance is closely linked with better operating performance represented by return on assets (ROA) and market valuation (Tobin Q) (Christensen et al.,2010). Different studies use different indicators such as EVA, financial ratios, and many researchers used Tobin Q, which gives positive as well as negative results, but most studies suggest that corporate governance has a significant impact on a firm's valuation. It implies that firms with stronger corporate governance practices perform better in financial terms as well.

Other topics cover diverse areas. Topic 3 is dominated by tokens such as "islamic" and "banks", possibly exploring Islamic banking which was irrelevant to our study and hence discarded. Topic 4 indicates CG Disclosures. Corporate governance disclosure practices involve the transparent communication of a company's governance mechanisms, policies, and procedures to stakeholders (Fung, 2014). Companies disclose this information through annual reports, proxy statements, and dedicated corporate governance sections on their websites. Enhanced disclosure can increase transparency, potentially leading to greater investor confidence, reduced information asymmetry, and lower capital costs (Madhani, 2008). The aim is to provide Shareholders and other interested parties with a clear knowledge of how the company is managed and controlled, fostering trust and accountability. Effective disclosure practices not only comply with regulatory requirements but also demonstrate a commitment to ethical business conduct and shareholder value creation. As global standards evolve, companies are increasingly expected to provide more comprehensive and accessible governance disclosures, including information on governance, environmental, social, and factors.

Topic 5 likely addresses agency theory and related mechanisms, as suggested by tokens like “study”, “firms”, “agency”, “theory”. Agency theory and corporate governance have a strong relationship. Agency theory is one of the core concepts in corporate governance and deals with the interaction of the shareholders and the managers of a company. Agency theory’s main focus is on inherent conflict between the company’s shareholders, who require the management to maximize corporate earnings, and management, who might decide to act in its self-interest. This misrepresentation can take many forms including overpaying executives, using funds which should be more profitably employed elsewhere etc. To solve these agency problems, there are measures of corporate governance to ensure that managers and shareholders are on the same side of the same fence. The conflicts are managed through internal structures such as board controls, performance-related executive remuneration, and clear reporting. Market for corporate control, legal requirements and mechanisms of institutional investors are equally work as internal mechanisms in holding the management responsive to the shareholders. Altogether, these mechanisms are expected to eliminate agency costs and enhance the efficiency, ethics, and sustainability of corporate governance.

Topic 6 has been discarded since it focuses on the terms like “banks” and “effects”.

Meanwhile, topic **7** centres on "ownership" and "companies" which might deal with ownership structures. When a company issues common stock/equity shares in an initial public offering (IPO), the promoters’ ownership decreases. To maintain control, management adopts various strategies. The intention of these measures is to disconnect voting rights from financial ownership rights. The ownership structure is defined as a control minority structure in which the owner has only a small portion of the cash flow rights but significant voting power (Buchholz & Sandler, 2021), (Cronqvist & Nilsson, 2003).

Shareholding is a significant part of an organization's internal management system. When there is poor leadership by the governing body, additional corporate governance strategies and techniques are implemented to improve oversight and decision-making. The influence of various ownership forms on governance practices is frequently investigated in research studies (Johnson & Greening, 1999). The Impact of various ownership structures on corporate governance has emerged as a significant topic of investigation. Various studies have proved that a good corporate governance helps companies become more efficient, improve access to finance, reduce risk, and avoid substandard governance (Kontogeorga et al., 2022; Mustafa & Morina, 2022; Prasad et al., 2022; Lapina et al., 2016; Raja & Kostyuk, 2015).

Again topic 8 deals with terms like “token”, “innovation”, “Islamic”, which has been discarded.

The table also identifies topics like impact of firm size and board size as suggested by topic 9 and 10 where Topic 9 is suggesting a term board size which has several impacts on firm’s financial performance. Board size means the entire directors on a company's board. It comprises of executive directors as well as non-executive directors. In the theory of corporate governance, the impact of board size on firm performance is backed up by different theories (Jackling, B., & Johl, S. 2009). Board where the number of directors is more in number are supported by the agency theory (Rashid, A. 2015). Whereas, Stewardship theory is in favour of board size with less numbers for the efficient leadership. Agency theory is of the view that where the board size is larger it enhances the firm performance by clearly monitoring by a large group of people. And implies that a larger board size provides a proficiency and varied insights across multiple fields. Thus, agency theory signifies the presence of positive relationship between the board size and firm performance supporting a larger board size (Jackling, B., & Johl, S. 2009). On the other hand, the stewardship theory, supports the negative relationship between the board size and firm performance.

Topic 10 suggests board diversity. Boards of Directors are the pillar for systems that guarantees that the interests of stakeholders and firm are aligned together by monitoring the working of firm as well (Larcker & Tayan, 2020). Board diversity, encompassing differences in gender, ethnicity, experience, and expertise, plays a crucial role in enhancing corporate governance (Katmon et al.,2019), (Rao & Tilt, 2016), (Asri, 2024). A diverse board brings different perspectives, enhances the decision-making, and brings innovation to the company ultimately leading to better financial performance (Dwekat et al.,2025). Research suggests that companies with diverse boards leads to better accountability, ethical decision-making, reducing the likelihood of corporate misconduct (Aziz, 2024). Overall, board diversity is an important factor in improving corporate governance effectiveness and long-term sustainability.

**CONCLUSION**

Our study attempted to uncover the latent themes of the field, as well as to identify the current topics and specific directions this field is taking. This analysis reveals the comprehensive intellectual foundation of the corporate governance field. LDA used in our research uncovers the topics emblematic of various meta-disciplines. Out of the 10 topics that our study found, we selected the ones that were associated to Corporate Governance which is our field of study. The topics that this study uncovers are effects on corporate governance which signifies the effects of CG on firm performance. Our study also revealed topics like CG disclosures which relates with disclosure of CG mechanisms in the companies’ annual reports. The next topic that our study found out is agency theory which is strongly related to corporate governance and deals with the interaction of the shareholders and the managers of a company. One of the topics of study revealed is ownership structures. It is the distribution of equity among shareholders and its impact on control, decision-making, and accountability within a company.

Our study also suggested topics like board size and impact of firm size. These topics are relevant in corporate governance and examining its impact on the performance of finances of firms. These distinct topics relating to corporate governance are identified using LDA. Also, we have depicted ENTREQ statement in our study to disclose material facts related to study. This study provides academics and practitioners with a road map for targeted investigation and strategic implementation in enhancing firm performance by highlighting the most important themes in corporate governance research. Our study adds to the literature by identifying these six relevant topics of academic discussion. To our knowledge, very few studies has previously used LDA approach in “corporate governance” and “firms’ performance” review of literature study.

Disclaimer (Artificial intelligence)

NO generative AI technologies such as Large Language Models and text-to-image generators have been used during the writing or editing of this manuscript.

**References**

Agarwal, M., & Gupta, M.B. (2024). How Corporate Governance Affect the Financial Performance? A Hybrid Review. *International Journal of Management Issues and Research, 13*(2 ), 103-116.

Ahmed Haji, A., & Mubaraq, S. (2015). The implications of the revised code of corporate governance on firm performance: A longitudinal examination of Malaysian listed companies. *Journal of Accounting in Emerging Economies*, *5*(3), 350-380.

Airoldi, E. M., & Bischof, J. M. (2016). Improving and evaluating topic models and other models of text. *Journal of the American Statistical Association*, *111*(516), 1381-1403.

Andres Alonso-Robisco, Javier Bas, Jose Manuel Carbo, Aranzazu de Juan & Jose Manuel Marques (30 Jun 2024): Where and how machine learning plays a role in climate finance research, Journal of Sustainable Finance & Investment, DOI: 10.1080/20430795.2024.2370325

Andres Alonso-Robisco, Javier Bas, Jose Manuel Carbo, Aranzazu de Juan & Jose Manuel Marques (30 Jun 2024): Where and how machine learning plays a role in climate finance research, Journal of Sustainable Finance & Investment, DOI: 10.1080/20430795.2024.2370325

Asmussen, C. B., & Møller, C. (2019). Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1), 1-18.

Asri, M. (2024). Exploring the Impact of Committee Structure and Composition on Corporate Governance Practices. *Advances in Managerial Auditing Research*, *2*(1), 26-39.

Aziz, F. (2024). The Role of Corporate Governance in Preventing Corporate Misconduct. *Liberal Journal of Management & Social Science*, *1*(4), 113-131.

Bebchuk, L., & Cohen, A. (2005). The costs of entrenched boards. Journal of Financial Economics, 78, 409433.

Bevir, M. (2012). *Governance: A very short introduction*. OUP Oxford.

Bhagat, S., & Bolton, B. (2008). Corporate governance and firm performance. *Journal of corporate finance*, *14*(3), 257-273.

Blei, D.M. (2012), “Probabilistic topic models”, Communications of the ACM, Vol. 55 No. 4, pp. 77-84, doi: 10.1145/2133806.2133826

Buchholz, W., & Sandler, T. (2021). Global public goods: A survey. Journal of Economic Literature, 59(2), 488–545. <https://doi.org/10.1257/jel.20191546>

Cadbury, A., 2002. Corporate Governance and Chairmanship: a Personal View. Oxford University Press on Demand.

Chandrakala, M., & Raja Kamal, C. (2024). Observing Shareholder Influence on Stakeholder Interests: An Analysis and Roadmap for Future Inquiry. In *AI in Business: Opportunities and Limitations: Volume 1* (pp. 349-357). Cham: Springer Nature Switzerland.

Chauhan, U., & Shah, A. (2021). Topic modeling using latent Dirichlet allocation: A survey. *ACM Computing Surveys (CSUR)*, *54*(7), 1-35.

Christensen, J., Kent, P., & Stewart, J. (2010). Corporate governance and company performance in Australia. *Australian accounting review*, *20*(4), 372-386.

Core, J. E., Guay, W. R., & Rusticus, T. O. (2006). Does weak governance cause weak stock returns? An examination of firm operating performance and investors' expectations. *the Journal of Finance*, *61*(2), 655-687.

Cronqvist, H., & Nilsson, M. (2003). Agency costs of controlling minority shareholders. *Journal of Financial and Quantitative analysis*, *38*(4), 695-719.

Daud, A., Li, J., Zhou, L., & Muhammad, F. (2010). Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China*, *4*, 280-301.

Dawar, G., Nagariya, R., Bhatia, S., Dhingra, D., Agrawal, M., & Dhaundiyal, P. (2024). Can financial markets help attain carbon goals? Evidence from systematic literature review, bibliometric analysis and topic modelling. *Sustainability Accounting, Management and Policy Journal*.

Dwekat, A., Taweel, A., & Salameh, A. (2025). Boardroom diversity and financial performance in Palestinian banks and insurers. *Discover Sustainability*, *6*(1), 1-18.

Egger, R. (2022).

Ernstberger, J., & Grüning, M. (2013). How do firm-and country-level governance mechanisms affect firms’ disclosure?. *Journal of Accounting and Public Policy*, *32*(3), 50-67.

Fouad, B. (2024). Innovation Accounting's Impact on Organizational Performance and Strategic Decision-Making-A Quantitative Study of Performance Metrics, Challenges and Opportunities.

Fung, B. (2014). The demand and need for transparency and disclosure in corporate governance. *Universal Journal of Management*, *2*(2), 72-80.

Gan, J., & Qi, Y. (2021). Selection of the optimal number of topics for LDA topic model—taking patent policy analysis as an example. *Entropy*, *23*(10), 1301.

Gurcan, F., & Cagiltay, N. E. (2022). Exploratory analysis of topic interests and their evolution in bioinformatics research using semantic text mining and probabilistic topic modeling. *IEEE Access*, *10*, 31480-31493.

Jackling, B., & Johl, S. (2009). Board structure and firm performance: Evidence from India's top companies. *Corporate governance: an international review*, *17*(4), 492-509.

Jamali, D., Safieddine, A. M., & Rabbath, M. (2008). Corporate governance and corporate social responsibility synergies and interrelationships. *Corporate governance: an international review*, *16*(5), 443-459.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, *78*, 15169-15211.

Johnson, R. A., & Greening, D. W. (1999). The effects of corporate governance and institutional ownership types on corporate social performance. *Academy of management journal*, *42*(5), 564-576.

Kataishi, R. (2024). The Technological Trajectory of Semantic Analysis: A Historical-Methodological Review of NLP in Social Sciences. *Available at SSRN 5022988*.

Katmon, N., Mohamad, Z. Z., Norwani, N. M., & Farooque, O. A. (2019). Comprehensive board diversity and quality of corporate social responsibility disclosure: Evidence from an emerging market. *Journal of business ethics*, *157*, 447-481.

Kherwa, P., & Bansal, P. (2019).

Kontogeorga, G. N., Thanasas, G. L., Smaraidos, V. S., & Angelaras, D. N. (2022). Reforming the corporate governance in Greece: After the introduction of Law 4706/2020. Corporate Law & Governance Review, 4(1), 45–53. <https://doi.org/10.22495/clgrv4i1p4>

Kukreja, V. (2023). Recent trends in mathematical expressions recognition: An LDA-based analysis. *Expert Systems with Applications*, *213*, 119028.

Kushkowski, J. D., Shrader, C. B., Anderson, M. H., & White, R. E. (2020). Information flows and topic modeling in corporate governance. *Journal of Documentation*, *76*(6), 1313-1339.

Kushkowski, J. D., Shrader, C. B., Anderson, M. H., & White, R. E. (2020). Information flows and topic modeling in corporate governance. *Journal of Documentation*, *76*(6), 1313-1339.

Kushkowski, J. D., Shrader, C. B., Anderson, M. H., & White, R. E. (2020). Information flows and topic modeling in corporate governance. *Journal of Documentation*, *76*(6), 1313-1339.

Lapina, Yu., Kostyuk, A. N., Braendle, U., & Mozghovyi, Ya. (2016). Shareholders rights and remedies (Comparative law perspective). Corporate Board: Role, Duties and Composition, 12(3), 6–13. https://doi.org /10.22495/cbv12i3art1

Larcker, D., & Tayan, B. (2020). *Corporate governance matters*. FT Press.

Leyel C, Kor JA- nang M. Corporate Governance and Firm Profitability: A Study of Ghana’s Banking Sector. Asian J. Adv. Res. Rep. [Internet]. 2024 Sep. 16 [cited 2025 Jul. 19];18(10):1-15. Available from: https://journalajarr.com/index.php/AJARR/article/view/750

Lin, W. Y., Wang, F., Cheng, M. M., Yeung, S. K., Torr, P. H., Do, M. N., & Lu, J. (2017). CODE: Coherence based decision boundaries for feature correspondence. *IEEE transactions on pattern analysis and machine intelligence*, *40*(1), 34-47.

Liu, J., Shang, J., Wang, C., Ren, X., & Han, J. (2015, May). Mining quality phrases from massive text corpora. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 1729-1744).

Madhani, P. M. (2008). Role of voluntary disclosure and transparency in financial reporting.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Adam, S. (2021). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In *Computational methods for communication science* (pp. 13-38). Routledge.

Mallin, C. A. (2019). *Corporate governance*. Oxford university press.

Mustafa, L., & Morina, M. (2022). Central governance and minority rights: The case of the developing country. Corporate Law & Governance Review, 4(2), 57–67. <https://doi.org/10.22495/clgrv4i2p7>

Ozyurt, O., Kabakus, A. K., & Ayaz, A. Exploring Scholarly Journal Content: Topic Modeling Analysis Using LDA. *Text Mining in Educational Research: Topic Modeling and Latent Dirichlet Allocation*, 71.

Ozyurt, O., Özköse, H., & Ayaz, A. (2024). Evaluating the latest trends of Industry 4.0 based on LDA topic model. *The Journal of Supercomputing*, 1-28.

Ozyurt, O., Özköse, H., & Ayaz, A. (2024). Evaluating the latest trends of Industry 4.0 based on LDA topic model. *The Journal of Supercomputing*, *80*(13), 19003-19030.

Paul, J., Lim, W. M., O’Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR‐4‐SLR). *International Journal of Consumer Studies*, *45*(4), O1-O16.

Peters, B. G. (2001). *The future of governing*. Studies in Government and Public Policy.

Prasad, V. H., Sharma, V. D., Bano, S. S., & Chand, M. N. (2022). Corporate governance and economic performance: A case study of the developing country. Corporate Governance and Organizational Behavior Review, 6(2), 8–18. <https://doi.org/10.22495/cgobrv6i2p1>

Raja, K., & Kostyuk, A. (2015). Perspectives and obstacles of the shareholder activism implementation: A comparative analysis of civil and common law systems. Corporate Ownership & Control, 13(1–5), 520–533. <https://doi.org/10.22495/cocv13i1c5p1>

Ramezanian Bajgiran, N., Margaritis, D., & Salimifar, M. Board Decision Quality, Corporate Governance and Firm Performance: A Study of European Firms. *Corporate Governance and Firm Performance: A Study of European Firms*.

Rao, K., & Tilt, C. (2016). Board composition and corporate social responsibility: The role of diversity, gender, strategy and decision making. *Journal of business ethics*, *138*, 327-347.

Rashid, A. (2015). Revisiting agency theory: Evidence of board independence and agency cost from Bangladesh. *Journal of business ethics*, *130*, 181-198.

References:

Rubin, T. N., Chambers, A., Smyth, P., & Steyvers, M. (2012). Statistical topic models for multi-label document classification. *Machine learning*, *88*, 157-208.

Sari, R. (2023). Enhancing Corporate Governance through Effective Oversight and Accountability. *Advances: Jurnal Ekonomi & Bisnis*, *1*(6), 344-356.

Schöne, M. T. (2024). *Artificial Intelligence and Corporate Reporting: Extracting Information From Unstructured Data With Deep Learning and Natural Language Processing* (Doctoral dissertation, Friedrich-Alexander-Universitaet Erlangen-Nuernberg (Germany)).

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of business research*, *70*, 263-286.

Spira, L. F., & Page, M. (2003). Risk management: The reinvention of internal control and the changing role of internal audit. *Accounting, Auditing & Accountability Journal*, *16*(4), 640-661.

Srivastava, M., 2009. Good governance - concept, meaning and features: A detailed study. Soc. Sci. Res.Network 1–23. <https://doi.org/10.2139/ssrn.1528449>.

Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T. (2004, August). Probabilistic author-topic models for information discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 306-315).

Topic modelling: Modelling hidden semantic structures in textual data. In *Applied data science in tourism: Interdisciplinary approaches, methodologies, and applications* (pp. 375-403). Cham: Springer International Publishing.

Tricker, R. I. (2015). *Corporate governance: Principles, policies, and practices*. Oxford University Press, USA.

Vulić, I., De Smet, W., Tang, J., & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, *51*(1), 111-147.

Wang, W., & Rada, R. (1998). Structured hypertext with domain semantics. *ACM Transactions on Information Systems (TOIS)*, *16*(4), 372-412.

Ward, A. J., Brown, J. A., & Rodriguez, D. (2009). Governance bundles, firm performance, and the substitutability and complementarity of governance mechanisms. *Corporate Governance: An International Review*, *17*(5), 646-660.

Xie, W., Zhao, W., & Xu, Z. (2024). Comparative analysis of multi-platform e-commerce online reviews based LDA and sentiment dictionary. *International Journal of Intelligent Computing and Cybernetics*.

Yang, Y., Wang, F., Zhang, J., Xu, J., & Yu, P. S. (2018). A topic model for co-occurring normal documents and short texts. *World Wide Web*, *21*, 487-513.. Topic modeling: a comprehensive review. *EAI Endorsed transactions on scalable information systems*, *7*(24).

Yi, F., Jiang, B., & Wu, J. (2020). Topic modeling for short texts via word embedding and document correlation. *IEEE Access*, *8*, 30692-30705.

Zhao, B., Zhou, H., Li, G., & Huang, Y. (2018). ZenLDA: Large-scale topic model training on distributed data-parallel platform. *Big Data Mining and Analytics*, *1*(1), 57-74.

Zhao, B., Zhou, H., Li, G., & Huang, Y. (2018). ZenLDA: Large-scale topic model training on distributed data-parallel platform. *Big Data Mining and Analytics*, *1*(1), 57-74.