West Science Interdisciplinary Studies
Vol. 4, No. 06, June 2026: pp. 1071-1080

Machine Learning Trends in Improving Startup Business Efficiency: A Bibliometric Review

Loso Judijanto[1], Adam Husain[2]
[1] IPOSS Jakarta, Indonesia and losojudijantobumn@gmail.com
[2] Universitas Kebangsaan Republik Indonesia and adamhusain@ukri.ac.id

ABSTRACT

The paper performs bibliometric analysis of studies on the trends in ML in increasing business efficiency of startups. Based on the data obtained from major scientific databases and analyzed by VOSviewer, the study examines the intellectual structure, thematic development, and global collaboration patterns in the field under investigation. In particular, the results show that the concept of machine learning is the main intellectual node that is related to the concepts like startups, investments, entrepreneurship, decision-making, forecasting, and data innovations. As for temporal dynamics, there is an evolution of basic statistical and predictive modeling methods to more sophisticated tools, which involve artificial intelligence-based decision support systems, innovation ecosystems, and scalability of startups. Finally, citation analysis detects key papers devoted to implementation of AI, success predictions of startups, and intelligent systems for validation of business models. Global collaboration patterns demonstrate a semi-centralized structure, with the US and India being the main nodes in the network.

Keywords: *Machine Learning, Startup Efficiency, Bibliometric Analysis, Entrepreneurship, VOSviewer*

1. INTRODUCTION

The fast development of information technologies has brought about a revolution in modern business environments, especially those involving startups. Startups function in volatile and uncertain environments with scarce resources, fierce competition, and need for innovation. In order to remain viable and develop sustainably, startups should be able to constantly enhance their performance, streamline decision-making procedures, and apply data-driven approaches [1], [2]. In recent times, machine learning (ML), one of the leading subfields of artificial intelligence (AI), has become an innovative technology that has the potential to increase the performance of businesses thanks to prediction, automation, pattern recognition, and intelligent decision making. With the rise in accessibility of computational power, cloud computing, and big data technologies, ML has been increasingly used in many business areas helping startups gain a competitive advantage in spite of scarce resources [3], [4].

The use of machine learning techniques provides numerous benefits for startups through ensuring greater efficiency of business processes and cost reduction. Machine learning enables businesses to detect hidden patterns, predict consumer behavior, manage inventory better, automate customer services, and enhance marketing efforts due to processing of both structured and unstructured big data sets. Recommendation engines, fraud detection tools, customer segmentation solutions, demand forecasting methods, and predictive maintenance software are some applications of machine learning used by technology-oriented startups. These opportunities enable organizations to make quicker and more accurate decisions and reduce the risk of errors and inefficiency of processes. Therefore, machine learning is a valuable resource for entrepreneurs [5].

Machine learning techniques are very useful for startups in terms of increased efficiency of business processes and decreased costs. Through using machine learning, companies are able to find patterns, predict behavior of consumers, control inventories, automate customer services and

increase their marketing activities by processing structured and unstructured big data. Examples of machine learning techniques applied by startups are recommendation engines, fraud detection systems, customer segmentation systems, demand forecasting approaches and predictive maintenance software. Such opportunities allow firms to take faster and more accurate decisions and avoid any risks of mistakes and inefficiency of operations. Consequently, machine learning becomes a valuable tool for entrepreneurs [6]–[8].

Even though the number of research papers examining the use of machine learning algorithms in businesses continues to increase, the existing studies are highly dispersed in multiple scientific areas and lines of research. While many studies discuss the application of machine learning within a certain industry, technology, or area of use, they lack a systematic review of the development process of the field and its intellectual structure. Thus, it may be difficult for researchers and practitioners to comprehend the process of development of machine learning research dedicated to the topic of startups' business efficiency, the most dominant themes in the area, and emerging themes that will affect the further development of the field. The rapidly changing nature of the technology makes it difficult to track the development trends through traditional narrative reviews of scientific publications.

Bibliometric analysis has proven to be a useful methodology for studying the development of scientific knowledge and researching the trends in a certain field. Bibliometric studies use quantitative methods to analyze the publication data, citation practices, collaboration networks, and keyword relations in order to determine the intellectual structure, prominent authors, thematic development, and emerging research areas of a field. In relation to machine learning and startup business efficiency, bibliometric analysis allows one to map the knowledge base and detect leading research clusters and development of academic interests in a certain timeframe. This is important both from the standpoint of further theoretical development and practical application by startups and other actors in the field [9].

While machine learning has been recognized as one of the most essential technologies to improve startup business efficiency, the existing research on this topic still lacks a structured analysis of its intellectual structure and evolution. The quick growth of publications dealing with the application of machine learning in such fields as entrepreneurship, innovation, operational efficiency, and startup management has created a fragmented knowledge base that makes it challenging to distinguish prevailing research topics, important papers, emerging trends, and possible research gaps. Moreover, no general opinion has been formed about the evolution of machine learning research in the field of startups and possible directions for future research in this area. Hence, a systematic bibliometric review is required to map the evolution of this research domain and identify the key domains of knowledge in this area, as well as possible opportunities for further research. This research is aimed at conducting a global review of the research in the area of machine learning in improving startup business efficiency through bibliometric analysis.

2. METHODS

The current paper uses the bibliometric analysis technique to analyze the development of scientific studies in the field of machine learning trends in improving the efficiency of startups' business. Bibliometric analysis is known as a very efficient approach to describing the structure of scientific knowledge, identifying important works and finding new topics of research in a certain scientific area. Data for the current analysis were collected using the Scopus database, which is regarded as one of the largest and most comprehensive abstract and citation databases of scholarly

literature. The Scopus database was chosen because of the wide range of journals, conference papers and other scholarly publications covered by it. The process of collecting data included searching for relevant papers by means of the keywords "machine learning", "startup", "business efficiency", "entrepreneurship" and others. Only those papers that were listed in Scopus and corresponded to the inclusion criteria were included in the analysis. The bibliographic data obtained were exported in CSV format.

Bibliometric data were analyzed using the software called VOSviewer, which is specifically intended for constructing and visualizing bibliometric networks. VOSviewer was used to conduct the analysis of keyword co-occurrences, co-authorship, citations and thematic maps, in order to identify the intellectual structure of the research field. The software produced various types of visualizations, such as network visualizations, overlay visualizations and density visualizations, which helped identify major research clusters, relations between the key concepts, prominent scholars and trends in research. Network visualizations were used to demonstrate relations between common keywords, while overlay visualizations showed the evolution of the research topic over time.

3. RESULT AND DISCUSSION

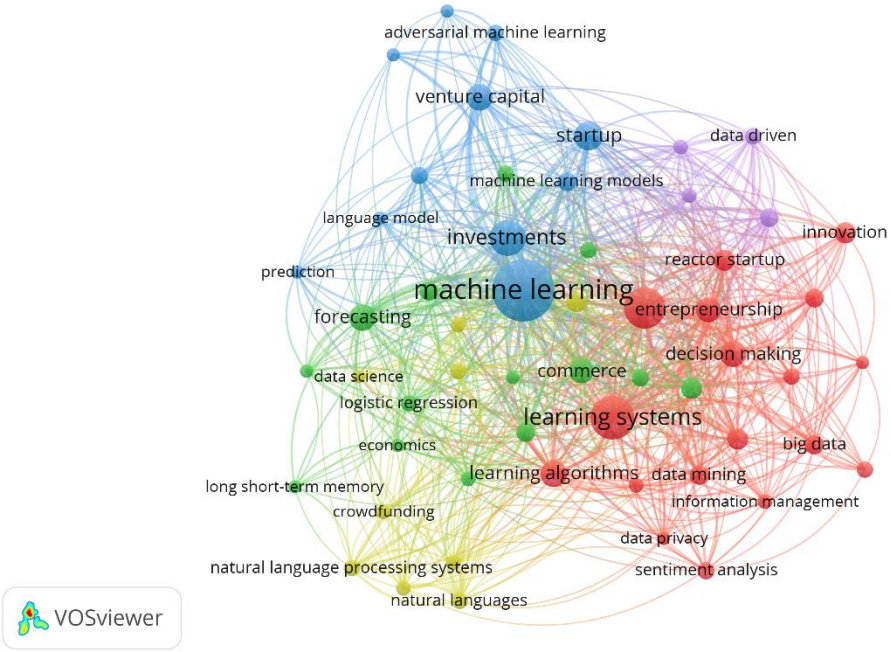

Figure 1. Network Visualization
Source: Data Analysis Result, 2026

As can be seen from Figure 1, machine learning serves as the core intellectual hub of the entire research field. Due to its big node size and high connectivity, it can be assumed that it is the keyword that occurs most often and co-occurs most often. This placement proves the fact that machine learning is viewed not as an ancillary method, but as the fundamental paradigm that links many application fields including investment, entrepreneurship, learning, and decision-making. There emerges another distinctive cluster (in blue) that centers on startups, venture capital, investments, and machine learning models. The formation of such a cluster clearly points to the financial aspects of the research topic. It can be observed that machine learning models are widely considered in terms of capital allocation, evaluation of startups, and systems for investment decision-making. The use of the terms "venture capital" and "adversarial machine learning" suggests a new area of concern – not only finding opportunities but managing risks in the field of digital startups.

The cluster associated with the color red relates to topics such as entrepreneurship, decision making, innovation, big data, and information management, which denote the managerial application of machine learning technology. Innovation acts as a bridging term for the relation between technological capability and organizational performance. It should be noted that the coexistence of "decision making" and "big data" denotes the growing importance of data-driven decision-making processes among young ventures.

The green cluster represents the more technical/operational aspect, which contains keywords like learning systems, learning algorithms, forecasting, data science, logistic regression, LSTM, economics, and crowdfunding. This is an indication that there is a lot of research conducted on the methodology and predictive models in the machine learning field. Forecasting and econometrics methodologies are indications of the utilization of ML in enhancing predictive efficiencies, while crowdfunding and economic terminologies show the use of ML in raising resources and making financial predictions. The yellow cluster consisting of NLP algorithms, sentiment analysis, and natural languages is an example of machine learning application in unstructured data. This particular cluster is relatively more specialized but is becoming increasingly significant, pointing to a trend in startups in the direction of behavioral and sentiment analysis of users and markets through text analytics.

3.1 Overlay Visualization

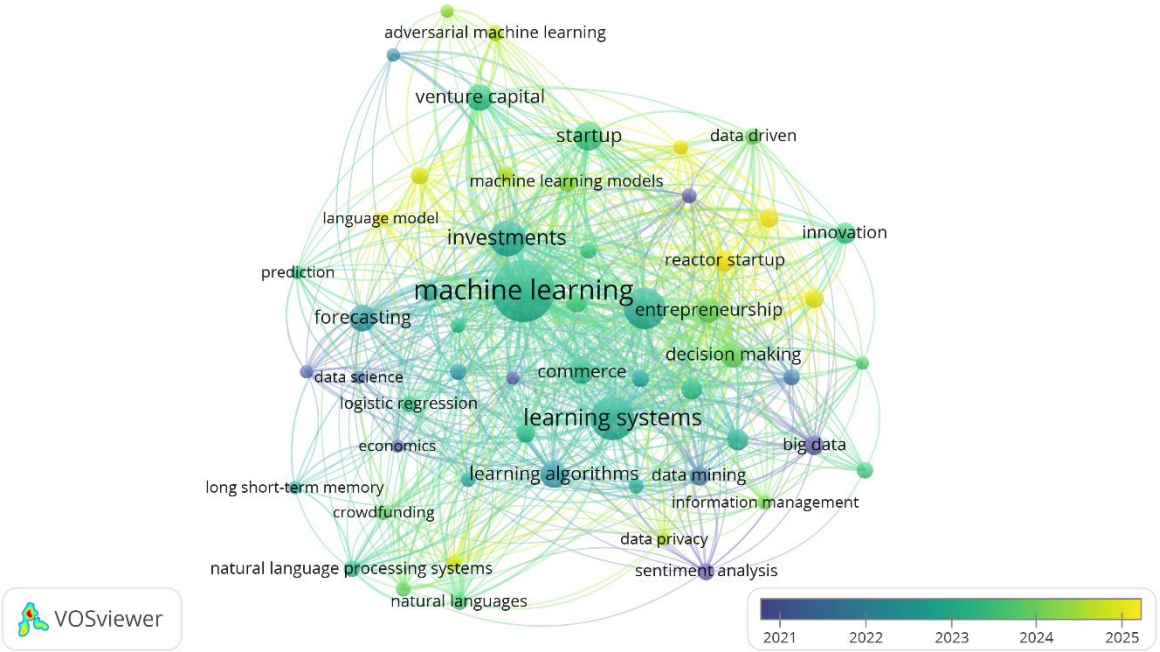

Figure 2. Overlay Visualization

Source: Data Analysis Result, 2026

As seen in Figure 2, Machine Learning is still the key conceptual node, having the biggest node and most interconnected one in all clusters. It acts as an integration layer between investment-related concepts (venture capital, startup, forecasting), methodological ones (learning algorithms, data science, logistic regression), and applications (entrepreneurship, decision-making, big data, and information management). The high level of interconnectedness implies that the field is multidisciplinary, and Machine Learning plays the role of an integrating concept rather than a separate methodological approach. Considering the timeline aspect (overlay 2021-2025), it is clear that the trend moves from basic techniques to application and innovation-oriented areas. The initial phase (blue, 2021-2022) focuses on more classical analytic and statistical techniques such as forecasting, econometrics, logistic regression, and venture capital-startup relationships. The middle

phase (green, 2023) highlights consolidation in terms of "learning systems," "data science," and enterprise-oriented applications. Newer innovations (yellow, 2024-2025) show the trend towards data driven innovation, modeling and decision intelligence for entrepreneurship. Innovation, data driven, machine learning models, language models, entrepreneurship are some of the terms used here, showing increased focus on models.

3.2 Citation Analysis

Table 1. The Most Impactful Literatures

Citations	Authors and year	Title
154	[10]	AI adoption in America: Who, what, and where
123	[11]	Detecting indicators for startup business success: Sentiment analysis using text data mining
110	[12]	A machine learning, bias-free approach for predicting business success using Crunchbase data
110	[13]	Examining the impact of artificial intelligence (Ai)-assisted social media marketing on the performance of small and medium enterprises: Toward effective business management in the saudi arabian context
103	[14]	AI Startup Business Models: Key Characteristics and Directions for Entrepreneurship Research
99	[15]	Artificial intelligence and policy: quo vadis?
87	[16]	Design principles for a hybrid intelligence decision support system for business model validation
79	[17]	The Data Science Handbook
49	[18]	Web-based startup success prediction
46	[19]	Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business

Source: Scopus, 2026

Most influential literature in this field consists of studies which combine machine learning, artificial intelligence, and startups'/businesses' performance analysis and have a high citation rate due to both their applied relevance for industry as well as their methodology. The highest rated study [10] focuses on the large scale usage of artificial intelligence in the USA. This paper is very important because it provides insights on the macro level of how people adopt AI, where it is used, and how it impacts organizations. In turn, [11] and [12] analyze how startups can succeed based on sentiment analysis and machine learning of structured datasets such as Crunchbase. Another important body of work deals with the application of AI technology to support business models and decision-making processes. Here, contributions by [14] and [16] can be singled out as providing frameworks for the design of hybrid intelligence and AI-supported validation of business models. Such contributions move the conversation from modeling per se to decision augmentation and organizational design, thus suggesting that AI is increasingly becoming a part of decision-making infrastructures instead of remaining merely an analytic instrument. Another contribution in this area is the work by [15], which focuses on policy considerations. The remaining articles that are highly cited offer the methodological and practical foundation for business analytics based on data, and include [20], analyzing the influence of AI-aided social media marketing in SME performance; [17] that is used as a fundamental work in data science practice. The previously mentioned articles by [18] and [19] represent the first attempts of machine learning application in predicting success and segmentation, respectively.

3.3 Density Visualization

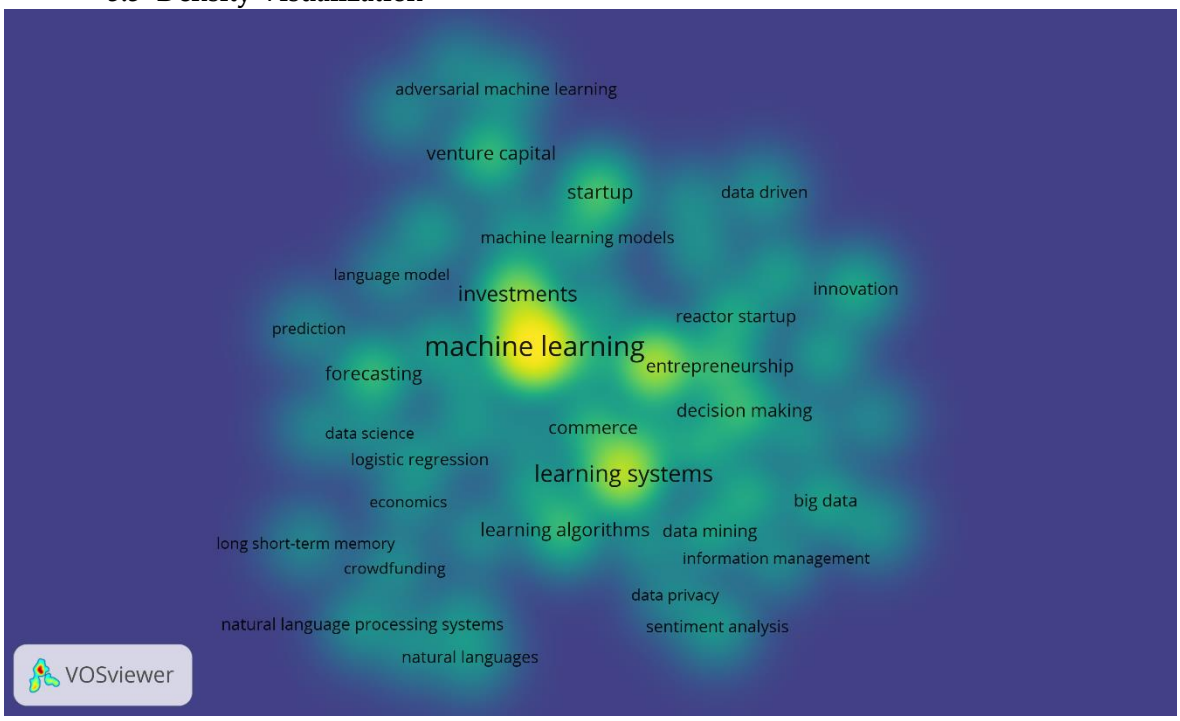

Figure 3. Density Visualization

Source: Data Analysis Result, 2026

Figure 3 indicates that machine learning is the most dominant and centrally concentrated research theme, represented by the highest intensity (yellow zone) in the network. This suggests a strong frequency of co-occurrence and conceptual centrality across the literature. Surrounding this core, moderately dense areas such as investments, learning systems, entrepreneurship, and forecasting show that the field is structurally organized around the application of machine learning to both financial and operational decision-making contexts. The gradual diffusion from the center toward blue regions reflects decreasing thematic concentration, where more specialized topics (e.g., adversarial machine learning, crowdfunding, natural language processing) appear less frequently but remain connected to the core structure. From a structural perspective, the map shows that the field is highly clustered around applied business and data-driven innovation domains rather than purely theoretical machine learning development. Areas such as big data, decision making, data mining, and information management form a contiguous high-density band, indicating strong integration between analytics infrastructure and managerial decision processes. Peripheral but still connected topics such as sentiment analysis, NLP systems, and data privacy suggest emerging but less saturated research directions.

3.4 Co-Authorship Analysis

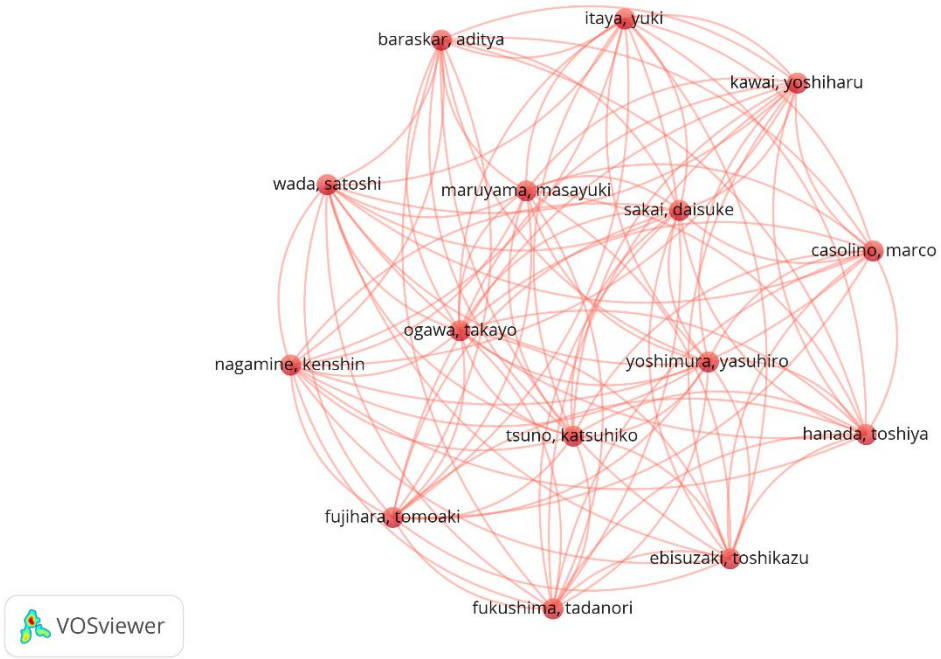

Figure 4. Author Visualization
Source: Data Analysis Result, 2026

Figure 4 depicts a highly interconnected co-authorship structure dominated by a single dense collaboration cluster, where all authors are strongly linked with minimal fragmentation into separate subgroups. This indicates a relatively cohesive research community in which knowledge production is concentrated and collaborative rather than dispersed. Key nodes such as Maruyama Masayuki, Tsuno Katsuhiko, Yoshimura Yasuhiro, and Sakai Daisuke appear more central, suggesting they function as connective actors bridging multiple collaboration pathways within the network. The absence of distinct clusters or isolated components implies that the field is characterized by intensive intra-group collaboration and repeated co-authorship patterns, rather than competing or independent research schools. Such a structure is typical in specialized or emerging research domains where a core group of scholars repeatedly co-publishes and develops a shared research agenda.

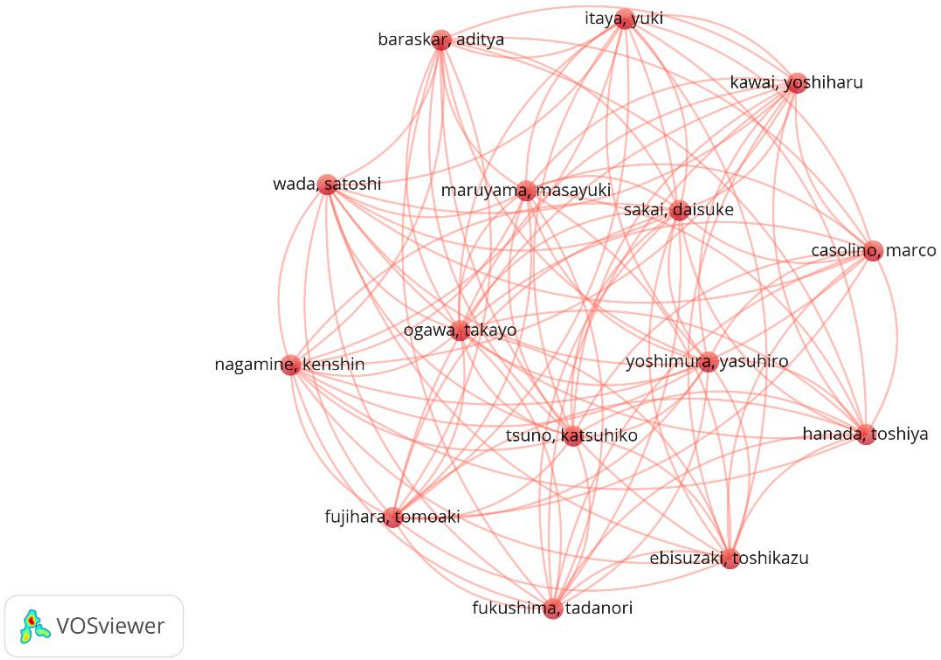

Figure 5. Institution Visualization
Source: Data Analysis Result, 2026

The network presents a densely connected co-authorship structure dominated by a single integrated research community, where all authors are strongly interlinked without clear fragmentation into distinct clusters. This indicates that the field is characterized by sustained collaboration among a relatively stable group of researchers rather than competing schools of thought or isolated research teams. The uniformly distributed connections suggest frequent co-publication and knowledge sharing across the entire network. Within this structure, several authors—such as Maruyama Masayuki, Tsuno Katsuhiko, Yoshimura Yasuhiro, and Sakai Daisuke—appear more central due to their higher connectivity, functioning as key nodes that bridge multiple collaboration pathways. However, the overall network remains highly cohesive, meaning influence is not strongly centralized in a single dominant author but rather distributed across multiple core contributors.

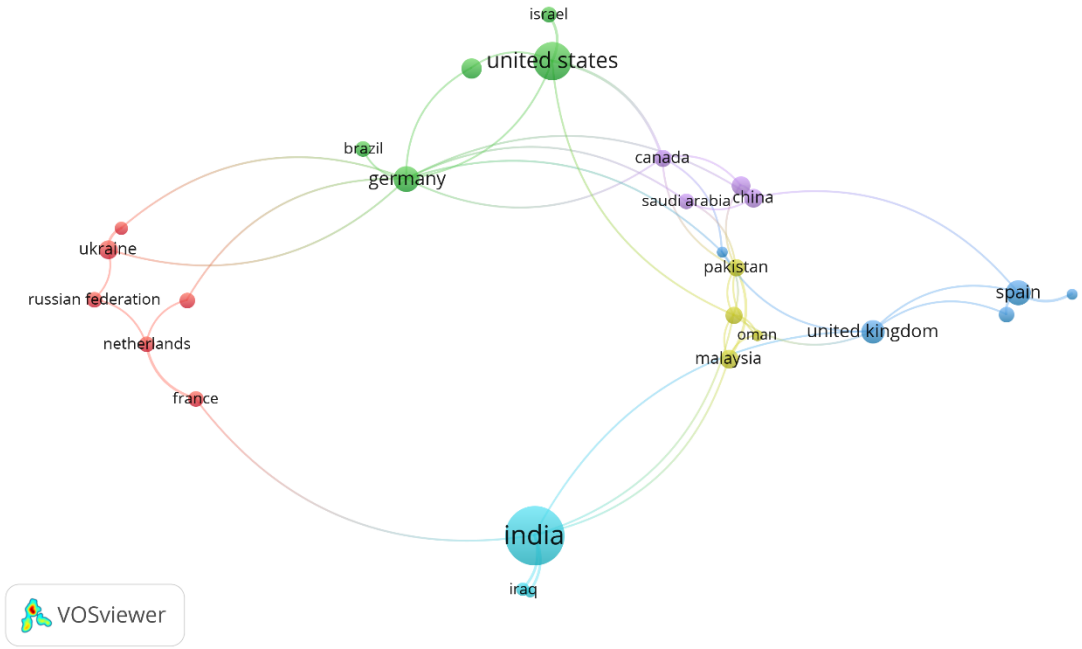

Figure 6. Country Visualization
Source: Data Analysis Result, 2026

Figure 6 illustrates a global collaboration structure in machine learning/startup-related research with clear regional clustering and uneven centrality. The United States and India emerge as the most prominent hubs, with the United States occupying a central bridging position connecting multiple regions (notably Germany, Brazil, and Israel), while India forms a second major hub linking South Asia and parts of the Middle East (including Malaysia, Oman, Pakistan, and Iraq). This indicates a dual-polar structure where both Western and South Asian actors play dominant integrative roles in global knowledge production. A second pattern is the formation of regional sub-clusters with strong intra-regional collaboration. Europe forms a relatively cohesive bloc involving France, Netherlands, Russia, and Ukraine, while another European-Western cluster links the United Kingdom with Spain. On the Asian side, China appears in a separate cluster connected with Saudi Arabia and Canada, suggesting cross-regional but selective collaboration pathways. These clusters indicate that while international cooperation exists, it is often structured through regional alliances rather than fully globalized interaction networks.

CONCLUSION

Bibliometric analysis of machine learning trends in making the business operations of startups efficient shows a fast-growing field with well-integrated intellectual structure. Intellectual structure is highly centralized towards the concept of machine learning, connecting various themes like investments, entrepreneurship, decision-making, forecasting, and learning systems. Keyword evolution and density show a clear shift from fundamental and statistical methods for prediction to advanced data-driven innovations, especially in startup ecosystems. Relevant literature also highlights that the trend of research is moving from methodology to practical application of AI systems in enhancing performance and business model validation of startups. In terms of international collaboration networks, there is evidence of a semi-centralized network of machine learning knowledge producers, where major bridge-nodes are represented by key states like the United States and India. In spite of the fact that intra-regional collaborations remain strong in some cases (like Europe and some Asian countries), the entire network shows signs of growing but rather fragmented international collaboration. All things considered, it can be stated that machine learning

knowledge is becoming an integral part of startup research and practice, but the landscape of global knowledge production is not evenly distributed and is limited to a certain number of states.

REFERENCES

[1] A. S. Munna and M. S. I. Shaikh, "Impact of digital marketing tools and methods on SMEs and proposed recommendations," *Contemporary Approaches of Digital Marketing and the Role of Machine Intelligence*. IGI Global, University of Sunderland, London, United Kingdom, pp. 173–191, 2023. doi: 10.4018/978-1-6684-7735-9.ch008.

[2] S. R. Cholil, R. Gernowo, C. E. Widodo, A. Wibowo, B. Warsito, and A. M. Hirzan, "Predicting Startup Success Using Tree-Based Machine Learning Algorithms," *Rev. Inform. Teor. e Apl.*, vol. 31, no. 1, pp. 50–59, 2024, doi: 10.22456/2175-2745.133375.

[3] S. A. Nisar, R. Arulkumar, B. Jones, and D. Banerjee, "A Hierarchical Business Model for Translating Digital Twins: A Case Study for Health Care Applications From Research to Market," *ASME Open J. Eng.*, vol. 5, 2026, doi: 10.1115/1.4064961.

[4] F. J., Ś. M., B. A., and H. S., Eds., "22nd International Conference on Enterprise Information Systems, ICEIS 2020," *Lect. Notes Bus. Inf. Process.*, vol. 417, 2021, [Online]. Available: https://www.scopus.com/pages/publications/85106414853?origin=resultslist

[5] A. Redchuk and F. W. Mateo, "New Business Models on Artificial Intelligence—The Case of the Optimization of a Blast Furnace in the Steel Industry by a Machine Learning Solution," *Appl. Syst. Innov.*, vol. 5, no. 1, 2022, doi: 10.3390/asi5010006.

[6] D. Gangwani, X. Zhu, and B. Furht, "Exploring investor-business-market interplay for business success prediction," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00723-6.

[7] N. S. Abd, O. S. Atiyah, M. T. Ahmed, and A. Bakhit, "Digital Marketing Data Classification by Using Machine Learning Algorithms," *Iraqi J. Electr. Electron. Eng.*, vol. 20, no. 1, pp. 245–256, 2024, doi: 10.37917/ijeee.20.1.23.

[8] A. Rawat, S. Tyagi, A. Sharma, and K. Ahalawat, "Machine Learning Insights into the Influence of Business Incubators on Startup Success," in *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking, CICTN 2025*, School of Management, Graphic Era Hill University, Dehradun, India: Institute of Electrical and Electronics Engineers Inc., 2025, pp. 164–168. doi: 10.1109/CICTN64563.2025.10932614.

[9] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021.

[10] K. McElheran *et al.*, "AI adoption in America: Who, what, and where," *J. Econ. Manag. Strateg.*, vol. 33, no. 2, pp. 375–415, 2024, doi: 10.1111/jems.12576.

[11] J. R. Saura, P. Palos-Sanchez, and A. Grilo, "Detecting indicators for startup business success: Sentiment analysis using text data mining," *Sustain.*, vol. 11, no. 3, 2019, doi: 10.3390/su11030917.

[12] K. Żbikowski and P. Antosiuk, "A machine learning, bias-free approach for predicting business success using Crunchbase data," *Inf. Process. Manag.*, vol. 58, no. 4, 2021, doi: 10.1016/j.ipm.2021.102555.

[13] W. Basri, "Examining the impact of artificial intelligence (Ai)-assisted social media marketing on the performance of small and medium enterprises: Toward effective business management in the saudi arabian context," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 142–152, 2020, doi: 10.2991/ijcis.d.200127.002.

[14] M. Weber, M. Beutter, J. Weking, M. Böhm, and H. Krcmar, "AI Startup Business Models: Key Characteristics and Directions for Entrepreneurship Research," *Bus. Inf. Syst. Eng.*, vol. 64, no. 1, pp. 91–109, 2022, doi: 10.1007/s12599-021-00732-w.

[15] A. Lauterbach, "Artificial intelligence and policy: quo vadis?," *Digit. Policy, Regul. Gov.* , vol. 21, no. 3, pp. 238–263, 2019, doi: 10.1108/DPRG-09-2018-0054.

[16] D. Dellermann, N. Lipusch, P. Ebel, and J. M. Leimeister, "Design principles for a hybrid intelligence decision support system for business model validation," *Electron. Mark.*, vol. 29, no. 3, pp. 423–441, 2019, doi: 10.1007/s12525-018-0309-2.

[17] F. Cady, *The Data Science Handbook*. Wiley Blackwell, 2017. doi: 10.1002/9781119092919.

[18] B. Sharchilev, D. Ozornin, M. Roizner, P. Serdyukov, A. Rumyantsev, and M. De Rijke, "Web-based startup success prediction," in *International Conference on Information and Knowledge Management, Proceedings*, P. N., C. S., W. H., A. J., A. R., L. A., C. A., Z. M., S. D., B. A., and S. A., Eds., Yandex Moscow, Russian Federation: Association for Computing Machinery, 2018, pp. 2283–2292. doi: 10.1145/3269206.3272011.

[19] D. Kamthania, A. Pahwa, and S. S. Madhavan, "Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business," *J. Comput. Inf. Technol.*, vol. 26, no. 1, pp. 57–68, 2018, doi: 10.20532/cit.2018.1003863.

[20] S. B. Assegaff and S. O. Pranoto, "Price determines customer loyalty in ride-hailing services," *Am. J. Humanit. Soc. Sci. Res.*, vol. 3, pp. 453–463, 2020.