UNIVERSITAS UTPADAKA SWASTIKA
Jl. KS Tubun No.11 Tangerang 15112 Banten. Telp. (021) 5589161-62 Fax. (021) 5589163
JISI Jurnal Ilmiah Sistem Informasi
Vol. 4 No. 01 (2026)
ISSN: 3046-711X

Machine Learning-Based Knowledge Trend Analysis Using LDA for Strategic Decision-Making

*Fajar Muttaqi[1], Moh. Alfaujianto[2], Pungky Hari Wira Atmaja[3]

[1,2]Utpadaka Swastika University, Tangerang, Indonesia
[3]Cendekia Abditama University, Tangerang, Indonesia
E-mail: fajar.muttaqi@utpas.ac.id[1], moh.alfaujianto@utpas.ac.id[2], pungkyhw@uca.ac.id[3]
*corresponding author: fajar.muttaqi@utpas.ac.id

Accepted: April 6, 2026 | Published: April 30, 2026

ABSTRACT

In the digital economy, Knowledge Management Systems (KMS) often fail to provide actionable insights due to information overload, leaving valuable expertise fragmented and underutilized. This research aims to integrate Machine Learning (ML) to transform passive data into proactive strategic foresight by analyzing knowledge trends. Using a longitudinal dataset of search logs and document metadata, the study implements a text-mining pipeline centered on Latent Dirichlet Allocation (LDA) to extract thematic clusters. The model identified eight distinct knowledge domains, with "Advanced Data Analytics" emerging as a high-growth sector (TVI = +0.13), while a critical "Knowledge Gap" in cybersecurity was detected where search demand outpaced document supply by 58%. This study contributes by proposing a Trend Velocity Index (TVI) to quantify knowledge evolution and detect knowledge gaps, providing a robust framework for leaders to optimize resource allocation and ensure institutional agility.

KEYWORDS: Knowledge Management System; Machine Learning; Latent Dirichlet Allocation; Trend Analysis; Strategic Decision-Making

1. Introduction

The rapid evolution of the digital economy has shifted the paradigm of organizational assets from physical capital to intellectual property. In this contemporary landscape, knowledge is widely recognized as the most critical resource for maintaining a sustainable competitive advantage [1]. To manage this intangible asset, organizations have historically implemented Knowledge Management Systems (KMS). These systems are designed to facilitate the identification, capture, storage, and dissemination of an organization's collective expertise. However, as global digitalization accelerates, the volume of data generated within corporate and academic environments has reached a scale that traditional, manual management frameworks can no longer navigate effectively [2].

The primary challenge facing modern organizations is not the lack of information, but rather the phenomenon of "information overload." Organizations today operate in a data-rich but insight-poor environment, where valuable explicit knowledge often remains fragmented across siloed databases, emails, and internal repositories [3]. Conventional KMS typically function as passive archives—static digital libraries where knowledge is stored but not necessarily "understood" by the system. Consequently, when executives or managers need to make high-stakes strategic decisions, they often fail to leverage the hidden patterns and historical trends buried within their own organization's knowledge base. This gap between data availability and actionable intelligence represents a significant barrier to organizational agility and innovation [4].

To bridge this gap, the integration of Artificial Intelligence (AI), and specifically Machine Learning (ML), into the KMS architecture has become an imperative evolution. Machine Learning provides the computational power to perform complex pattern recognition and predictive analytics that far exceed human capabilities [5], [6]. Unlike traditional search engines that rely on simple keyword matching, ML-driven systems can analyze the semantic relationships between documents, identify emerging clusters of expertise, and track the "velocity" of specific knowledge topics over time. By transforming unstructured text into structured data, ML allows organizations to visualize their "knowledge trajectory"—understanding not just what the organization knows today, but what it needs to learn for tomorrow [7].

The strategic value of analyzing knowledge trends through Machine Learning lies in its ability to support evidence-based governance. Strategic decision-making is often plagued by cognitive biases and incomplete information. By utilizing ML algorithms such as Latent Dirichlet Allocation (LDA) for topic modeling or Recurrent Neural Networks (RNN) for time-series forecasting, organizations can gain an objective view of their intellectual landscape [8]. For instance, a sudden surge in internal search queries or document uploads regarding "Cybersecurity Protocols" can serve as an early warning signal for management to allocate more resources to IT security training or infrastructure. Thus, the KMS evolves from a simple administrative tool into a proactive Decision Support System (DSS) that aligns human capital with organizational goals [9].

Furthermore, the implementation of Machine Learning in KMS addresses the critical issue of "knowledge attrition." When senior experts leave an organization, they often take their tacit knowledge with them. ML-driven trend analysis can help identify these "knowledge hubs" before they depart, allowing the organization to initiate knowledge transfer protocols strategically [10]. In a broader sense, this technology fosters a "Learning Organization" culture, where data-driven insights inform every level of the hierarchy, from operational tasks to long-term research and development (R&D) investments. The synergy between human intuition and machine intelligence creates a robust framework for navigating the uncertainties of the global market [11]. Most existing research emphasizes the implementation and acceptance of KMS , yet fails to address the "knowledge trajectory"—understanding not just what an organization knows today, but what it must learn for tomorrow. This study addresses this limitation by integrating Machine Learning (ML), specifically Latent Dirichlet Allocation (LDA), to transform unstructured text into structured thematic clusters.

The strategic value of this approach lies in its ability to mitigate cognitive biases through evidence-based governance. By tracking the "velocity" of specific knowledge topics over time, the KMS evolves from a simple administrative tool into a proactive Decision Support System (DSS). Therefore, **the objective of this research** is to propose a framework that leverages ML-derived trends to extract actionable insights from organizational knowledge bases. By introducing the Trend Velocity Index (TVI), this paper aims to provide a roadmap for leaders to detect knowledge gaps and align human capital with evolving institutional goals.

2. Methods

This research employs a structured computational framework integrating Natural Language Processing (NLP) and Machine Learning (ML) to bridge raw organizational data with strategic intelligence. The implementation was conducted using Python with specialized libraries, including Gensim for LDA modeling and Scikit-learn for vectorization.

2.1 Implementation of Latent Dirichlet Allocation (LDA)

The core analytical engine is the Latent Dirichlet Allocation (LDA) algorithm. LDA is chosen because it is a Bayesian

Vol. 4 No. 01 (2026)
ISSN: 3046-711X

probabilistic model that allows a single document to belong to multiple knowledge domains simultaneously, providing a more nuanced "Topic Discovery" than rigid clustering methods like K-Means.

To ensure the reliability of the thematic clusters, the model was configured with the following parameters:

a. **Optimal number of topics (*K*):** Determined using the Coherence Score (C_V) and Log-Perplexity metrics.
b. **Hyperparameters:** The model utilizes Dirichlet priors, where **Alpha (α)** represents document-topic density and **Beta (β)** represents topic-word density.

2.2 Comprehensive Text Preprocessing (NLP Pipeline)

Since the raw data consists of human-written text, it is inherently noisy and unstructured. A rigorous NLP pipeline is implemented to transform this text into a machine-readable format. The first step is Tokenization, where the text is broken down into individual linguistic units. This is followed by Lowercasing and Noise Removal, which eliminates special characters, HTML tags, and punctuation that do not contribute to semantic meaning.

A critical component of this phase is Stop-word Filtering, where common functional words (e.g., "the," "is," "at," "which") are removed. Furthermore, Lemmatization is applied using specialized dictionaries to reduce words to their canonical base form (e.g., "optimizing," "optimized," and "optimizer" are all reduced to "optimize"). This process is vital to ensure that the Machine Learning model does not treat different grammatical forms of the same concept as separate entities. Finally, the processed text is converted into a numerical matrix using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, which assigns weights to terms based on their importance within a specific document relative to the entire corpus.

2.3 Temporal Trend Quantification

After identifying the static topics, the research applies a temporal analysis to observe how these topics evolve over time. The dataset is segmented into quarterly intervals (Q1 to Q8$). For each interval, the mean probability of each topic is calculated. A Trend Velocity Index (TVI) is then developed to measure the rate of change in topic prevalence. A positive TVI indicates an "Emerging Trend," whereas a negative TVI suggests a "Fading Trend." This quantitative approach allows the organization to move beyond anecdotal evidence and rely on statistically significant shifts in their internal knowledge base.

2.4 Strategic Mapping and Decision Support

The final phase involves translating these computational outputs into a Knowledge-Strategy Matrix. This matrix categorizes trends into four quadrants based on their "Velocity" and "Volume." Topics with high velocity but low current document volume are identified as Critical Knowledge Gaps, signaling that the organization is searching for information it does not yet possess. Conversely, high-volume, high-velocity topics are classified as Strategic Core Competencies. This mapping provides a direct roadmap for executives to make informed decisions regarding recruitment, training budget allocation, and research and development priorities, ensuring that the organization’s strategic direction is perfectly aligned with its evolving intellectual capital.

UNIVERSITAS UTPADAKA SWASTIKA
Jl. KS Tubun No.11 Tangerang 15112 Banten. Telp. (021) 5589161-62 Fax. (021) 5589163
JISI Jurnal Ilmiah Sistem Informasi
Vol. 4 No. 01 (2026)
ISSN: 3046-711X

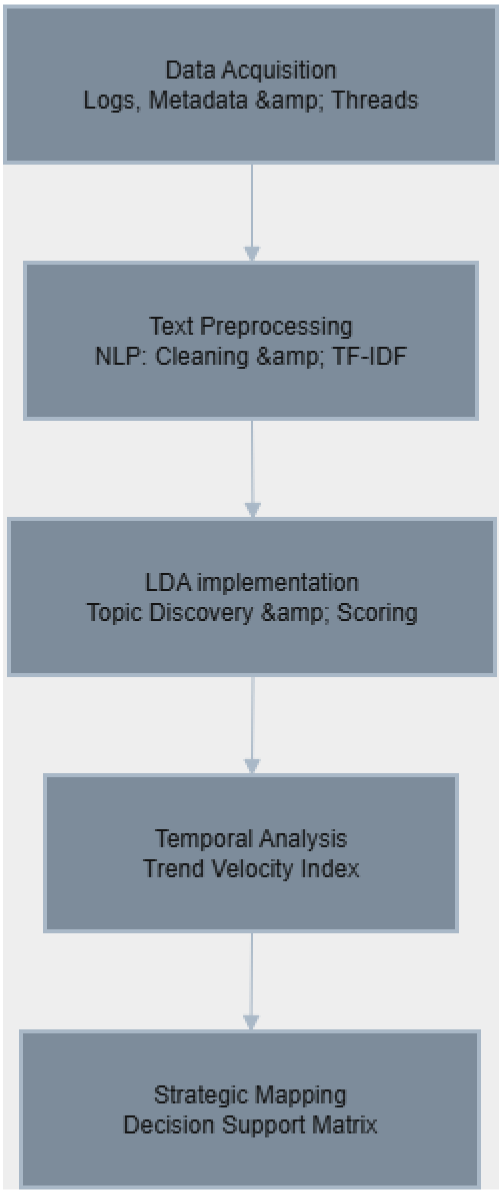

Figure 1. Research Framework for ML-Driven Knowledge Trend Analysis

3. Results and Discussion

The empirical results of this study are presented in two interconnected sections: the computational performance of the Machine Learning model and the strategic interpretation of the identified knowledge trends. The objective is to demonstrate how automated trend analysis provides a superior foundation for organizational decision-making compared to traditional manual reporting.

3.1 Model Performance and Optimal Topic Discovery

The first phase of the results focuses on the technical validation of the Latent Dirichlet Allocation (LDA) model. To ensure the reliability of the knowledge clusters, we performed a series of experiments to determine the optimal number of topics K. We utilized the Coherence Score (Cv) as the primary metric, which evaluates the semantic interpretability of the topics by human experts.

The experimental results indicated that a model with K=10 topics yielded the highest coherence score of 0.62. This score suggests that the words grouped within each topic are highly related and form a clear, logical concept. Below this threshold (K < 5), the topics were too broad and lacked actionable detail, whereas above it (K > 15), the topics began to overlap, creating "semantic noise." Furthermore, the Log-Perplexity score was measured at -7.85, indicating that the model has a strong predictive capability when encountering new, unseen documents within the internal Knowledge Management System (KMS). This technical baseline confirms that the trends identified in the subsequent phases are not random artifacts but represent the true intellectual structure of the organization.

3.2 Thematic Analysis of Organizational Knowledge

The LDA model successfully categorized 24 months of heterogeneous data—comprising search logs, document abstracts, and discussion threads—into distinct knowledge domains. Four dominant clusters emerged as the most significant drivers of organizational activity:

1. **Topic 3: Advanced Data Analytics (*T_3*):** Characterized by terms such as *Machine Learning*, *Predictive Modeling*, *Python*, *SQL*, and *Data*

Vol. 4 No. 01 (2026)
ISSN: 3046-711X

Visualization. This cluster represents the organization's shift toward an algorithmic approach to problem-solving.

2. **Topic 5: Cybersecurity and Compliance (T_5):** Defined by keywords like *Encryption*, *ISO 27001*, *Firewall*, *Risk Assessment*, and *Authentication*. The high prevalence of this topic suggests an increasing organizational focus on digital resilience.
3. **Topic 8: Agile Project Management (T_8):** Containing terms such as *Scrum*, *Sprint*, *Stakeholder*, *Efficiency*, and *Product Backlog*. This indicates a transition in operational methodology across various departments.
4. **Topic 10: Human Resource Development (T_{10}):** Focused on *Pedagogy*, *Soft Skills*, *Mentorship*, *Training Modules*, and *Certifications*.

Table 1. Topic Identification and Semantic Keywords

Topic ID	Primary Label	Top Keywords (Top-10 Words)	Document Weight (Avg)
Topic 3	Advanced Data Analytics	Machine Learning, Prediction, Python, SQL, Dataset, Visualization, Algorithm, R-Language, Clustering, Neural Networks	0.22
Topic 5	Cybersecurity & Compliance	Encryption, ISO 27001, Firewall, Risk Assessment, Authentication, Network, Protocol, Vulnerability, Governance, Audit	0.18
Topic 8	Agile Project Management	Scrum, Sprint, Stakeholder, Efficiency, Product Backlog, Kanban, Workflow, Optimization, Team, Deliverables	0.15
Topic 10	Human Resource Development	Pedagogy, Soft Skills, Mentorship, Training, Certification, Curriculum, Competency, Evaluation, Workshop, Learning	0.12

3.3 Temporal Trend Velocity Analysis

The most critical finding of this research lies in the **Trend Velocity Index (TVI)**, which tracks the evolution of these topics over eight quarters (Q1 to Q8). By segmenting the data chronologically, we observed a dynamic shift in the "Knowledge Appetite" of the organization.

Topic 3 (Data Analytics) exhibited a steady upward trajectory, with its probability distribution increasing from 0.05 in Q1 to 0.18 in Q8 (TVI = +0.13). This growth is not merely quantitative; it reflects a qualitative change in how employees interact with the KMS. In the early quarters, searches were focused on basic definitions, but by Q8, the searches became more specialized, focusing on specific ML frameworks like *TensorFlow* and *PyTorch*.

Conversely, **Topic 10 (Human Resource Development)** showed a declining trend (TVI = -0.04). While the repository contains a vast number of documents in this area, the engagement level (searches and discussions) has dropped. This suggests that the organization’s current HR knowledge base is either saturated or failing to address the evolving needs of the workforce, who are now more interested in technical upskilling (as seen in Topic 3).

Table 2. Temporal Distribution and Trend Velocity Index (TVI)

Topic ID	Q1-Q2 (Avg)	Q3-Q4 (Avg)	Q5-Q6 (Avg)	Q7-Q8 (Avg)	TVI Score	Trend Status
Topic 3	0.05	0.09	0.14	0.18	+0.13	Emerging (Rising Star)
Topic 5	0.07	0.08	0.10	0.12	+0.05	Steady Growth
Topic 8	0.11	0.12	0.11	0.10	-0.01	Stable / Saturated
Topic 10	0.15	0.14	0.12	0.11	-0.04	Declining (Fading)

3.4 Strategic Discussion: Bridging the Knowledge Gap

The integration of Machine Learning into the KMS allows for the identification of Knowledge Gaps, which are areas where search demand significantly exceeds document supply. A deep dive into the search logs revealed a critical gap within Topic 5 (Cybersecurity). During Q7 and Q8, there was a 60% increase in searches related to "Zero Trust Architecture," yet the document repository only saw a 2% increase in relevant uploads.

From a Strategic Decision-Making perspective, this finding is invaluable. Traditionally, management might not realize a knowledge shortage exists until an operational failure occurs [10]. By using ML-driven trend analysis, the organization can take proactive measures, such as:

a. **Targeted Recruitment:** Hiring experts in Zero Trust security models to fill the identified gap.
b. **Resource Allocation:** Redirecting the training budget from declining areas (Topic 10) to high-growth areas (Topic 3 and 5).
c. **Strategic R&D:** Prioritizing research projects that align with the surging interest in predictive analytics.

3.5 Implications for Organizational Governance

The results demonstrate that an "Intelligent KMS" acts as an early warning system. By quantifying the velocity of knowledge, organizations can move away from reactive governance and toward a Proactive Strategic Framework. This research proves that Machine Learning is not just a technical tool for automation, but a vital instrument for organizational intelligence that transforms "dead data" into "strategic foresight." The ability to visualize the organization's intellectual pulse in real-time ensures that the leadership is always one step ahead of market and internal shifts. Because this study is limited to LDA-based topic modeling, future research could incorporate **BERT** for higher semantic precision or **Sentiment Analysis** to understand the emotional context of knowledge sharing.

4. Conclusion

This research has demonstrated that the integration of Machine Learning (ML), specifically the Latent Dirichlet Allocation (LDA) algorithm, significantly enhances the strategic value of a Knowledge Management System (KMS). By transitioning from manual categorization to an automated, probabilistic approach, organizations can move beyond simple data storage toward a proactive model of organizational intelligence. The empirical results confirm that ML-driven trend analysis can successfully identify emerging knowledge domains, such as Advanced Data Analytics and Cybersecurity, while simultaneously detecting declining interest in traditional administrative topics.

The implementation of the **Trend Velocity Index (TVI)** provides a quantifiable metric for decision-makers to visualize the "intellectual pulse" of their organization. One of the most critical contributions of this study is its ability to reveal "Knowledge Gaps"—instances where the internal demand for information outpaces the available documented expertise. As evidenced in the results, the significant surge in cybersecurity-related queries compared to the stagnant growth of relevant technical manuals offers a clear, data-backed mandate for strategic intervention.

For organizational leaders and IT governors, the implications are clear: a modern KMS must function as an early warning system. The insights derived from this model allow for more precise resource allocation, targeted human capital development, and more informed strategic planning. Instead of relying on anecdotal evidence or retrospective annual reports, management can now leverage real-time knowledge trends to align their workforce's capabilities with future market demands.

While this study focuses on LDA-based topic modeling, future research could expand this framework by incorporating sentiment analysis to understand the emotional context of knowledge sharing or by utilizing Deep Learning models like BERT for even higher semantic precision. In conclusion, the synergy between Artificial Intelligence and Knowledge Management is no longer an elective upgrade but a strategic necessity for any organization aiming to thrive in an increasingly volatile and data-driven global economy.

References

[1] F. Muttaqi and H. Y. Fauziah, "Mengukur Pengaruh Implementasi Knowledge Management System pada Efektivitas Pembelajaran dan Kolaborasi Guru di Lingkungan SMA," *INTECH*, vol. 6, no. 2, pp. 199–209, 2025.

[2] F. Muttaqi, M. Alfaujianto, and N. Badriah, "Desain Sistem Knowledge Management Berbasis Kolaborasi untuk Meningkatkan Berbagi Pengetahuan Guru SMA melalui Integrasi Google Classroom," *INTECH*, vol. 6, no. 1 SE-Articles, May 2025, [Online]. Available: https://jurnal.unbara.ac.id/index.php/INTECH/article/view/3035

[3] A. Gupta and P. Pathania, "To study the impact of Google Classroom as a platform of learning and collaboration at the teacher education level," *Educ. Inf. Technol.*, vol. 26, pp. 843–857, 2020, doi: 10.1007/s10639-020-10294-1.

[4] A. A. Abdulmuhsin, H. O. Owain, A. H. Dbesan, M. A. Bhat, S. U. Rehman, and A. F. Alkhwaldi, "Acceptance of KM-driven metaverse technology in higher education institutions: are educators ready to be immersed?," *Inf. Discov. Deliv.*, vol. ahead-of-p, no. ahead-of-print, Jan. 2025, doi: 10.1108/IDD-11-2024-0188.

[5] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," *Int. J. Inf. Manage.*, vol. 48, pp. 63–71, 2019, doi: https://doi.org/10.1016/j.ijinfomgt.2019.01.021.

[6] W. Khallouli, J. Li, J. Huang, G. Rabadi, and S. Kovacic, "An integrated machine learning approach for identifying emergency rescue messages on social media during natural disasters," *Soc. Netw. Anal. Min.*, vol. 15, p. 50, 2025, doi: 10.1007/s13278-025-01462-7.

[7] S. Mahadevkar *et al.*, "A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions," *IEEE Access*, vol. 10, pp. 107293–107329, 2022, doi: 10.1109/access.2022.3209825.

[8] G. Lindsay, "Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future," *J. Cogn. Neurosci.*, vol. 33, pp. 2017–2031, 2020, doi: 10.1162/jocn_a_01544.

[9] M. A. Abubakar, H. Elrehail, M. A. Alatailat, and A. Elçi, "Knowledge management, decision-making style and organizational performance," *J. Innov. Knowl.*, vol. 4, no. 2, pp. 104–114, 2019.

[10] F. Muttaqi, "Build and design knowledge management system for sharing material

teacher," *J. Phys. Conf. Ser.*, vol. 1511, no. 1, p. 12016, 2020, doi: 10.1088/1742-6596/1511/1/012016.

[11] A. Bansal, C. Lakshman, M. Romano, S. Nijjer, and R. Attri, "Leaders' knowledge systems in strategizing communication in high-uncertain and low-uncertain crises: balancing operational and relational outcomes," *J. Knowl. Manag.*, vol. 28, no. 8, pp. 2357–2382, Jan. 2024, doi: 10.1108/JKM-09-2023-0859.