Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 2 April 2026, pp: 546-554
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Unsupervised Text Mining of Employee Feedback for Identifying Organizational Strengths and Improvement Areas

Febri Ari Wicaksono[1*], Imam Yuadi[2]

[1]Master of Human Resources Development, Graduate School, Airlangga University, Indonesia
[2]Department of Information and Library Science, Faculty of Social and Political Sciences, Airlangga University, Indonesia

E-Mail: [1]febri.ari.wicaksono-2025@pasca.unair.ac.id, [2]imam.yuadi@fisip.unair.ac.id

Received Dec 27th 2025; Revised Jan 14th 2026; Accepted Feb 27th 2026; Available Online Apr 19th 2026
Corresponding Author: Febri Ari Wicaksono

Abstract

Employee feedback provides rich signals about organizational performance, yet its free-text format makes systematic analysis at scale difficult. This study proposes an unsupervised text mining workflow in Orange Data Mining to extract actionable themes from continuous employee comments by separating two semantic polarities: strength feedback ("What went well?") and improvement feedback ("What could be improved?"). After cleaning and Indonesian-language preprocessing (Sastrawi stemming, custom stopwords), 3,406 strength and 3,172 improvement entries were represented using TF–IDF. Improvement feedback was clustered using K-Means and assessed with silhouette-based validation, while both feedback types were explored using LDA topic modeling supported by topic coherence checks for interpretability. The results reveal recurring organizational themes related to goal execution and performance, supervision, communication/coordination, and motivation, with notable vocabulary overlap between strengths and areas for improvement. Scientifically, this work demonstrates how polarity-aware unsupervised analytics improves interpretability compared to treating feedback as a single corpus, and practically, it provides a scalable way for managers to transform unstructured feedback into structured insights for targeted improvement initiatives.

Keywords: Clustering, Employee Feedback, Organizational Analytics, Text Mining, Topic Modeling

1. INTRODUCTION

Employee feedback is crucial for comprehending organizational performance, employee engagement, and managerial decision-making. Previous research has shown that qualitative employee feedback offers valuable insights that enhance quantitative performance metrics, especially in the realms of performance management and organizational analytics [1], [2], [4]. Nonetheless, this feedback is generally gathered in an unstructured textual format, rendering extensive manual analysis subjective, time-consuming, and hard to scale.

Progress in text mining and unsupervised learning has enabled the automated analysis of large volumes of unstructured organizational data. Text analytics approaches, including as clustering and topic modeling, have been extensively employed to identify hidden patterns and dominant themes within employee-generated content and organizational documentation [3], [6], [7]. Despite these advancements, current research still views employee feedback as a singular construct, potentially overlooking the semantic distinctions between positive feedback and constructive criticism, which may undermine the interpretability and practical applicability of the results [6], [11].

Employee performance feedback is typically gathered through organized performance evaluation cycles facilitated by digital performance management systems. In the examined organization, feedback data were collected during a biannual employee performance evaluation, facilitated by the recent introduction of a performance management system that allows for ongoing qualitative feedback reporting. These systems are acknowledged as crucial facilitators of ongoing performance management and organizational learning [15], [16]. The feedback in this study was derived from the system's continuous feedback module and anonymized prior to analysis.

Continuous feedback data offers both advantages and obstacles. The substantial amount of feedback facilitates comprehensive exploratory analysis; nevertheless, brief textual responses, specialized terminology, and mixed sentiments hinder systematic interpretation [10], [11]. In the absence of suitable preprocessing and analytical frameworks, significant insights may stay obscured inside the data.

Although clustering and topic modeling have been adopted in HR and organizational analytics, a methodological gap remains in how feedback polarity is handled. Treating strength feedback ("What went well?") and improvement feedback ("What could be improved?") as a single corpus can blur intent: reinforcement signals and corrective signals often differ in semantic structure and managerial implications. Polarity-aware analysis is therefore needed to improve interpretability and support more actionable interventions.

To address this gap, this paper proposes a polarity-aware unsupervised text mining workflow that analyzes strength and improvement feedback separately. The workflow combines Indonesian-language preprocessing, clustering for improvement feedback, and topic modeling for both feedback types using a visual analytics pipeline implemented in Orange Data Mining [6], [18].

The study contributes in three distinct ways. Initially, we present a scalable, pragmatic method for assessing extensive employee input using unsupervised text mining [7], [11]. Secondly, we emphasize the analytical advantages of distinguishing between strength-oriented and improvement-oriented feedback rather than treating employee feedback as a single entity [6], [18]. Third, we offer empirical insights into views of organizational performance based on actual feedback gathered during an operational performance management cycle [1], [15].

Research Objective. This study aims to:

1. Develop a polarity-aware unsupervised workflow for continuous employee feedback.
2. Identify dominant improvement themes via clustering and assess cluster quality using silhouette analysis.
3. Extract interpretable themes from strength and improvement feedback using LDA supported by topic coherence.
4. Compare strength and improvement feedback to understand vocabulary overlap and its implications for managerial interpretation.

Research Question**.**

RQ1: What dominant improvement themes emerge from employee feedback using unsupervised clustering, and how reliable are they based on silhouette validation?

RQ2: What latent themes characterize strength and improvement feedback based on LDA topic modeling with coherence-guided interpretability checks?

RQ3: To what extent do strength and improvement feedback share vocabulary, and what does this overlap imply for managerial interpretation?

2. MATERIALS AND METHOD

This research uses an unsupervised text-mining framework to analyze employee input collected from a continuous feedback module of an enterprise performance management system. The methodological framework adheres to a systematic analytical pipeline comprising Business Understanding, Data Understanding, Data Preparation, Text Representation, Unsupervised Analysis, and Visualization & Interpretation, as depicted in Figure 1. This methodology corresponds with recognized text analytics and data mining techniques in organizational and human resources research [1], [2].

2.1. Business Understanding

Employee feedback is an essential qualitative metric for assessing organizational performance and employee engagement. However, this material is generally unstructured, rendering manual analysis inefficient and potentially more biased. This study seeks to reveal concealed patterns and prevailing themes in employee feedback by applying unsupervised learning techniques, which are widely used in organizational text analytics and performance assessment research [3], [4].

The analysis focuses on two feedback categories derived from the feedback form:

1. *What went well?* (strength feedback), and
2. *What could be improved?* (improvement feedback).

By delineating these categories, we employed more focused analytical methodologies that more accurately represent their significances and commercial goals

2.2. Data Understanding

The dataset comprises written feedback provided by employees from a singular organizational unit during a performance review period. The feedback was extracted from the continuous feedback feature of a newly implemented performance management system, where employees provide qualitative comments on performance outcomes.

The dataset initially comprised 4,538 strength feedback entries and 4,222 improvement feedback entries. Preliminary analysis uncovered discrepancies in linguistic usage, colloquial idioms, abbreviations, and unconventional textual formats, which are typical in organizational feedback data [5]. These properties necessitated an extensive text preparation procedure prior to analytical modeling. Because the dataset originates from a single organizational unit, the findings should be interpreted as context-specific insights rather than organization-wide patterns. The study is designed as an exploratory and transferable workflow demonstration; replicating the pipeline across multiple units and time periods is required to test generalizability and stability of themes.

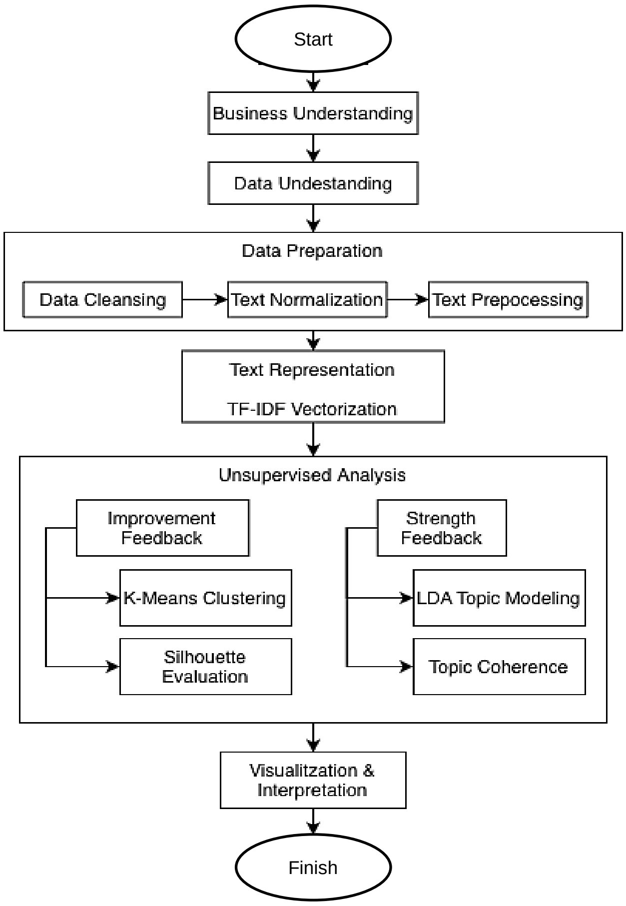

Figure 1. Methodology Workflow

2.3. Data Preparation

Data preparation was conducted to enhance data quality and enabled suitability for computational analysis. It comprised three stages:

1. Data Cleansing
   Data cleansing involved removing duplicate entries, null responses, superfluous punctuation, and extraneous symbols. This phase aims to diminish noise and enhance the reliability of ensuing text representations, as indicated in prior text mining studies [6].

2. Text Normalization
   Text normalization was achieved by converting all text to lowercase, eliminating digits and stopwords, and applying Indonesian stemming with the Sastrawi stemmer. Stemming reduces lexical diversity by correlating inflected words to their base forms, hence enhancing semantic consistency across documents [7].

3. Text Preprocessing
   Tokenization was executed at the lexical level, followed by part-of-speech filtering to retain significant terms. After preprocessing, the final datasets consisted of 3,406 strength feedback records and 3,172 improvement feedback records, showing effective noise reduction while preserving semantic content.

2.4. Text Representation

The Term Frequency–Inverse Document Frequency (TF–IDF) approach was employed to convert textual data into a numerical representation suitable for analysis. TF–IDF is a prevalent method for representing techniques that accentuate distinctive phrases while diminishing the impact of frequently occurring yet less informative terms [8], [9]. This vector format underpins the clustering and topic modeling techniques utilized in this paper.

2.5. Unsupervised Analysis

Due to the exploratory nature of the research and the absence of predefined labels, unsupervised learning methodologies were employed. Diverse approaches were utilized for each feedback type to align with their analytical objectives.

2.5.1 Improvement Feedback: K-Means Clustering

K-Means Clustering was employed to analyze enhancement feedback. This approach groups documents in the TF–IDF vector space based on their semantic similarity. K-Means is frequently used for text clustering due to its efficiency and simplicity [10].

We employed the silhouette coefficient to assess the quality of the clusters. This coefficient indicates the degree of separation between clusters and their interconnectivity. Silhouette analysis is a widely used method for evaluating clustering performance in text mining applications [11]. In the context of short, open-ended feedback, silhouette values can be modest because comments often contain multiple issues and share overlapping vocabulary across themes; therefore, validation is complemented by interpretability checks during cluster labeling

We selected $k = 3$ because it provides a practical balance between cluster interpretability and thematic granularity for organizational action. A smaller k runs the danger of putting different types of improvement issues into groups that are too broad. A bigger k, on the other hand, tends to break up brief feedback into overlapping micro-clusters that are tougher to understand and turn into projects.

2.5.2 Robustness Check: Comparison with Hierarchical Clustering (Ward Linkage)

To address the concern that low silhouette scores may reflect suboptimal separation, we conducted a robustness check by comparing the K-Means ($k = 3$) solution with agglomerative hierarchical clustering applied to the same TF–IDF representation. We calculated a cosine distance matrix and used Ward linkage to do hierarchical clustering. We then sliced the dendrogram into three clusters so we could compare them directly. We used cluster-level mean silhouette values to characterize the quality of the clusters and triangulated the results by assessing the interpretability of the phrases and comments most representative of each cluster. This comparison research support confidence that the prevailing improvement themes are not contingent upon methodology, while recognizing the intrinsic overlap typical of concise organizational books.

2.5.3 Strength Feedback: LDA Topic Modeling

Latent Dirichlet Allocation (LDA) was employed to identify latent topics that signify recurring positive performance themes for strength feedback. LDA conceptualizes each text as a combination of subjects and each topic as a distribution of words. This renders it suitable for exploratory thematic analysis [12], [13].

We employed topic coherence to assess the comprehensibility of the identified subjects. Topic coherence examines the semantic alignment of the primary keywords inside each topic and has been shown to correlate effectively with human judgment [14].

2.6. Visualization and Interpretation

Visualization techniques were utilized to comprehend and validate the analytical data; various visualization methods were applied. Principal Component Analysis (PCA) was utilized to condense high-dimensional TF–IDF vectors into two dimensions for the purpose of cluster visualization (Figure 2). We employed silhouette analysis to assess the quality of the clustering, and due to the substantial size of the dataset, we utilized box plots to illustrate the silhouette values (Figure 3). We utilized word cloud visualizations to illustrate the most significant terms within each topic for strength feedback. This facilitated comprehension of the outcomes of topic modeling (Figure 4).

3. RESULTS AND DISCUSSION

This section presents and discusses the results of the unsupervised text analytics conducted on employee feedback data. Two distinct analytical methods were employed to accommodate the various types of feedback: K-Means clustering for improvement feedback and Latent Dirichlet Allocation (LDA) topic modeling for strength feedback. This division facilitates understanding of personnel issues and effective performance metrics within the organization.

3.1. Data Preparation Outcomes

After cleaning, normalization, and preprocessing, the number of usable feedback items decreased from 4,538 to 3,406 for strength feedback and from 4,222 to 3,172 for improvement feedback. This reduction results from eliminating incomplete, redundant, or unhelpful replies, thereby enhancing the quality of the data for further analysis. Preprocessing is crucial in organizational text analytics, as noisy feedback data might alter the results of clustering and topic extraction [1], [5].

3.2. Improvement Feedback Clustering Analysis

We employed K-means clustering to identify latent patterns in feedback on areas requiring improvement. We categorized the input into three groups after employing Principal Component Analysis (PCA) for vectorization and dimensionality reduction.

3.2.1 Cluster Structure Interpretation

Figure 2 illustrates the clustering outcomes in a two-dimensional PCA projection of the enhancement feedback data. The colors indicate each location's cluster affiliation. The scatter plot clearly illustrates that the three clusters are not in contact with one another. This indicates that the input for enhancement possesses distinct semantic categories. While there is considerable overlap at the cluster boundaries, the general spatial distribution clearly indicates that the improvement themes differ.

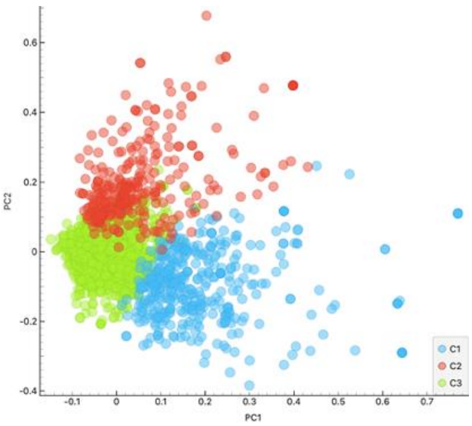

Figure 2. PCA-based scatter plot of improvement feedback clustering results using K-Means, visualized on the first two principal components (PC1 and PC2).

Silhouette analysis was employed to quantitatively assess the quality of the clusters. Figure 3 presents a box plot that encapsulates the silhouette scores for each cluster. This provides insight into the cohesion of the clusters and their separation from one another.

3.2.2 Cluster Quality Evaluation

The silhouette coefficient using cosine distance was implemented as an internal validation metric to assess the validity of K-Means clustering applied to improvement feedback. Due to the substantial quantity of feedback instances (N = 3172), individual silhouette bar plots were not utilized. sometimes prove challenging to comprehend for extensive text volumes. Silhouette values were aggregated using box plots, and mean silhouette scores for each cluster were calculated, providing a clearer representation of clustering efficacy, as depicted in Figure 3. The box figure illustrates the variation in silhouette values across each cluster, including the median and dispersion. This facilitates the comparison of the cohesion of the clusters and their separation from one another. Clusters C1 and C2 exhibit elevated median silhouette values and reduced interquartile ranges, indicating more internal consistency. Cluster C3, conversely, exhibits a lower median and a broader dispersion, indicating a greater overlap with adjacent clusters.

The average silhouette value for each cluster, as shown in the visual summary, is presented in Table 1.

Table 1. Mean Silhouette Score per Cluster (Improvement Feedback)

Cluster	Mean Silhouette Score	Interpretation
C1	0.057	Acceptable cluster cohesion
C2	0.061	Strongest cluster cohesion
C3	-0.008	Overlapping and heterogeneous cluster

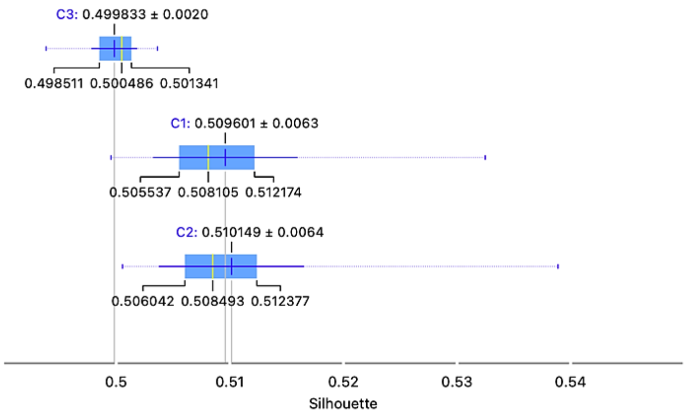

Figure 3. Box plot distribution of silhouette scores for K-Means clustering of improvement feedback

Clusters C1 and C2 exhibited positive mean silhouette ratings of 0.057 and 0.061, respectively; conversely, Cluster C3 recorded a mean silhouette score of -0.008, indicating a highly negative value. The findings indicate that the basic clustering structure is suitable for exploratory research; nevertheless, Cluster C3 reveals a more diverse theme with overlapping semantic content.

This pattern corresponds with the traits of employee enhancement feedback, where process-related, coordination-related, and managerial difficulties often intersect. The diminished silhouette score of Cluster C3 does not indicate methodological failure; rather, it illustrates the complexity of organizational feedback data. Consequently, the three-cluster configuration remains suitable for producing exploratory insights and aiding managerial interpretation.

To further address the concern that modest silhouette values may indicate suboptimal separation, we conducted a robustness check by comparing the primary K-Means solution with an alternative clustering method under the same TF–IDF representation (Table 2).

Table 2. Cluster-Level Mean Silhouette Comparison

Method	Setting	Cluster	Mean Silhouette Score	Interpretation
K-Means	TF-IDF; *k=3*	C1	0.057	Acceptable cluster cohesion
		C2	0.061	Strongest cluster cohesion
		C3	-0.008	Overlapping and heterogeneous cluster
Hierarchical Clustering	TF-IDF + cosine distance; ward linkage; cut = *3*	C1	0.177	Highest cohesion/separation
		C2	0.028	Substantial overlap with neighboring clusters
		C3	-0.014	Overlapping and heterogeneous cluster

We compared K-Means (k = 3) to agglomerative hierarchical clustering with Ward linkage (cut = 3 clusters). The comparison shows that silhouette values stay low for both methods, which is what we would expect for short, open-ended feedback on improvements, since each comment often has more than one issue and uses the same words across themes. Still, hierarchical clustering makes one cluster more cohesive (mean silhouette = 0.177), while the other clusters still show boundary overlap (0.028 and −0.014), which is what you would expect from K-Means (0.057, 0.061, and −0.008). Overall, the comparison results show that the three-cluster setup should be kept for exploratory interpretation. K-Means is still the best model because it is simple and easy to use in practice. The hierarchical baseline gives us greater confidence that the improvement themes we identified aren't linked to a specific method.

3.2.3 Interpretation of Improvement Themes

The identified clusters reveal several areas for improvement expressed by employees. These topics primarily pertain to enhancing procedures, addressing coordination issues, and managing workload distribution while ensuring clarity of expectations for all individuals involved. In contrast to strength feedback, which highlights accomplishments and motivation, improvement feedback primarily focuses on operational deficiencies and areas requiring managerial intervention. The clustering-based approach enables management to prioritize improvement projects by identifying prevailing trends derived from employee feedback.

3.3. Strength Feedback Topic Modeling Analysis

We employed Latent Dirichlet Allocation (LDA) to analyze employee feedback to identify prevailing positive performance themes within the firm. The study of topic coherence revealed that three subjects provide the best interpretable and representative framework of the strength feedback corpus. Table 3 summarizes the extracted subjects and their corresponding keywords.

Table 3. Extracted Latent Topics and Representative Keywords for Strength Feedback

Topic ID	Dominant Theme Description	Representative Keywords
Topic 1	Work Execution and Monitoring	kerja, tahan, jalan, laksana, tingkat, laku, sesuai, monitoring, proses, audit
Topic 2	Team Coordination and Communication	selesai, tugas, kerja, tim, koordinasi, komunikasi, semester, inisiatif, bidang
Topic 3	Performance Achievement and Motivation	capai, target, kerja, sesuai, kpi, semangat, koordinasi, analisa, data, laku

Topic 1 emphasizes the necessity for systematic task execution and operational oversight, indicating that implementation is uniform. Topic 2 explores collaboration, communication, and teamwork, illustrating that individuals coexist harmoniously. Topic 3 addresses achieving objectives, meeting key performance indicators, and maintaining employee motivation. This attribute indicates that the organization possesses a robust culture of productivity.

Figure 4 illustrates the word cloud representation for Topic 3, thus corroborating these conclusions. Phrases like “capai,” “target,” “KPI,” “semangat,” and “analisa” enhance the importance of achievement-oriented terminology in performance assessments. The significance of these phrases reflects employees' recognition of performance outcomes and their motivation.

Figure 4. Dominant Keywords in Strength Feedback Topic 3

3.4. Comparison Between Strength and Improvement Feedback

An analysis of both types of feedback demonstrates a complementary relationship. Improvement feedback identifies operational deficiencies and process-related obstacles, whereas strength feedback underscores accomplishments, collaboration, and performance evaluation. This dual-perspective analysis offers a comprehensive picture of organizational performance, highlighting both areas for enhancement and current strengths. This reinforces the value of polarity-aware analysis: separating the corpora preserves intent while still enabling cross-polarity interpretation

Comparison with Related Findings. The extracted themes goal execution/monitoring, coordination and communication, and performance motivation are consistent with patterns reported in prior HR and employee feedback analytics using topic modeling and related text mining methods [1], [10], [20]. This convergence supports the external plausibility of the thematic interpretation and indicates that operational execution, collaboration, and performance management structures are recurrent dimensions of employee narratives. However, because the corpus reflects one unit’s vocabulary and work context, the cross-polarity overlap observed here should be considered indicative for this setting and validated through replication in other units.

3.5. Managerial and Analytical Implications

The results yield practical implications for organizations utilizing digital performance management solutions such as SAP SuccessFactors PMGM. Improvement clusters facilitate the formulation of targeted actions to address recurring issues, while strength themes help devise strategies to reinforce and acknowledge best practices. Organizations can utilize continuous feedback analytics to transition from manual qualitative evaluations to data-driven performance assessments. The proposed methodology indicates that unsupervised text analytics can enhance continuous feedback systems and improve managers' evidence-based decision-making.

4. CONCLUSION

This study presented an unsupervised text analytics framework for analyzing employee continuous feedback acquired through a digital performance management system, as illustrated by SAP SuccessFactors Performance and Goals Management. The proposed approach effectively discerned notable patterns in unstructured feedback data by combining TF-IDF representation with K-Means clustering and Latent Dirichlet Allocation (LDA), thereby obviating the necessity for predefined labels.

The clustering results, validated by silhouette analysis, indicate that improvement feedback exhibits varying degrees of thematic cohesion, underscoring the complex and interconnected nature of performance-related issues. Topic modeling about strong feedback indicates that positive themes consistently relate to goal attainment, collaboration, and performance evaluation based on data. Together, the findings provide management with actionable signals: what to reinforce (strength themes) and what to address (improvement themes), while also enabling cross-polarity interpretation where vocabulary overlaps.

The data were collected from a single organizational unit, potentially limiting their applicability to other contexts. The study employs TF–IDF characteristics, which may inadequately encapsulate nuanced meanings in succinct sentences. Finally, a lot of the rating depends on things about the work itself, like how well it fits together, how clear it is, and how easy it is to understand. The current goal did not include getting outside help through human coding or connecting to the company's performance indicators.

Future research can extend this work by (1) performing longitudinal analysis across multiple review cycles to assess theme shifts after interventions, (2) expanding coverage to multiple units or organizations to test robustness, and (3) using embedding-based representations to better capture semantics and reduce boundary overlap. Triangulation with expert validation (e.g., HR reviewers) and alignment with quantitative performance metrics could further strengthen external validity and practical deployment.

REFERENCES

[1] P. Thakral, P. R. Srivastava, S. S. Dash, S. M. Jasimuddin, and Z. J. Zhang, "Trends in the thematic landscape of HR analytics research: A structural topic modeling approach," *Management Decision*, vol. 61, no. 12, pp. 3665–3690, 2023, doi: 10.1108/MD-01-2023-0080.

[2] A. Joshi, S. Sekar, and S. Das, "Decoding employee experiences during pandemic through online reviews," *Personnel Review*, vol. 53, no. 1, pp. 288–313, 2024, doi: 10.1108/PR-07-2022-0478.

[3] J. Kim, P.-S. Chang, and S. Yang, "A comparative analysis of job satisfaction prediction models using machine learning: A mixed-method approach," *Data Technologies and Applications*, vol. 59, no. 1, pp. 41–60, 2025, doi: 10.1108/DTA-10-2023-0697.

[4] R. Tripathi, M. Thite, and A. Varma, "Appraising the revamped performance management system in Indian IT multinational enterprises: The employee perspective," *Human Resource Management*, vol. 60, no. 5, pp. 475–493, 2021, doi: 10.1002/hrm.22061.

[5] A. Malik, P. Budhwar, H. Mohan, and N. R. Srikanth, "How does algorithm-based HR predict employees' sentiment? Developing an employee experience model through online reviews," *Industrial and Commercial Training*, vol. 56, no. 4, pp. 273–289, 2024, doi: 10.1108/ICT-08-2023-0060.

[6] X. Liu, W. Lu, S. Liu, and C. Qin, "Discovering trends and journeys in knowledge-based human resource management: Big data smart literature review based on machine learning," *IEEE Access*, vol. 11, pp. 95567–95583, 2023, doi: 10.1109/ACCESS.2023.3296140.

[7] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, "Data-driven strategies in operation management: Mining user-generated content in business models," *Annals of Operations Research*, vol. 333, no. 2, pp. 607–629, 2024, doi: 10.1007/s10479-022-04776-3.

[8] L. Xu, M. Xi, and Y. Zhang, "Quantitative evaluation of policies for combining medical and nursing care based on the LDA–PMC model: A comparative analysis of typical Chinese provinces," *Policy Sciences*, vol. 57, no. 2, pp. 321–345, 2024, doi: 10.1007/s11115-022-00675-0.

[9] S. Hosseini, M. Farahani, and M. Manthouri, "Deep text clustering using stacked autoencoder and k-means," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 38619–38644, 2022, doi: 10.1007/s11042-022-12155-0.

[10] Y. Ding and H. Li, "Uncovering employee insights: Integrative analysis using structural topic modeling and support vector machines," *Journal of Big Data*, vol. 12, no. 1, pp. 1–25, 2025, doi: 10.1186/s40537-025-01100-1.

[11] S. Lee and Y. Choi, "Configurations of resourceful and demanding attributes of organizational culture in US hotels: An innovative approach using topic modeling and fsQCA," *Journal of Innovation & Knowledge*, vol. 9, no. 4, Art. no. 100582, 2024, doi: 10.1016/j.jik.2024.100582.

[12] R. M. Kowalski, M. A. Khanbhai, and P. Aikman, "Patients' written reviews as a resource for public healthcare management: A text mining approach," *Procedia Computer Science*, vol. 112, pp. 263–270, 2017, doi: 10.1016/j.procs.2017.08.275.

[13] M. Mahyoub, "Hierarchical text clustering and categorisation using a semi-supervised framework,"

in *Proc. 12th Int. Conf. Developments in eSystems Engineering (DeSE)*, 2019, doi: 10.1109/DeSE.2019.00037.

[14] S. A. Hasan, R. Wang, and M. G. Hussain, "Clustering analysis of Bangla news articles with TF-IDF and CV using mini-batch k-means and agglomerative clustering," in *Proc. IEEE CyberneticsCom*, 2022, doi: 10.1109/CyberneticsCom55287.2022.9865339.

[15] S. Nadeem and Z. Tayyab, "The interplay between national cultural dimensions and components of a performance management system: A qualitative study from Pakistan," *Canadian Journal of Administrative Sciences*, vol. 38, no. 1, pp. 78–92, 2021, doi: 10.1002/cjas.1587.

[16] M. A. Köseoglu, A. K. F. Wong, and S. S. Kim, "Intellectual structure of the hospitality literature via topic modeling analysis," *Journal of Hospitality & Tourism Research*, vol. 48, no. 4, pp. 679–708, 2024, doi: 10.1177/10963480221118814.

[17] G. R. Akkartal and F. Mızrak, "Operational efficiency and competitiveness in the global logistics industry: An examination of human resources management strategies," in *Strategic Innovations for Dynamic Supply Chains*, Advances in Logistics, Operations, and Management Science. Hershey, PA, USA: IGI Global, 2024, ch. 4, doi: 10.4018/979-8-3693-3575-8.ch004.

[18] X. Mo and Y. Liao, "Analysis of current research in the field of sustainable employment based on latent Dirichlet allocation," *Sustainability*, vol. 16, no. 11, Art. no. 4557, 2024, doi: 10.3390/su16114557.

[19] Y. Guo and O. M. Karatepe, "A 30-year journey of hospitality and tourism research: A comprehensive topic modeling analysis," *International Journal of Contemporary Hospitality Management*, vol. 36, no. 7, pp. 2232–2258, 2024, doi: 10.1108/IJCHM-01-2023-0109.

[20] Y. Guo and O. M. Karatepe, "Is someone listening to me? A topic modeling approach to analyzing open-ended employee feedback in hospitality," *International Journal of Hospitality Management*, vol. 115, Art. no. 104114, 2025, doi: 10.1016/j.ijhm.2025.104114