

RESEARCH

Open Access



Uncovering employee insights: integrative analysis using structural topic modeling and support vector machines

Kai Ding¹, Ruihong Li², Zeyu Li^{3*} and Shangui Hu¹

*Correspondence:
andyli_macao@hotmail.com

¹ Ningbo University of Finance and Economics, Ningbo, China
² Hengshui University, Hengshui, China

³ Universiti Putra Malaysia, Serdang, Malaysia

Abstract

Online platforms that enable employees to voluntarily share their opinions and experiences about current and former employers present a valuable data source for investigating worker satisfaction. This user-generated feedback has the potential to provide insights that surpass the limitations of traditional survey methodologies. This study proposes a novel approach by integrating Structural Topic Modeling (STM) analysis with Support Vector Machine (SVM) techniques to scrutinize the robustness of STM findings, particularly concerning the relative significance of extracted topics. This research reveals several insightful observations based on analyzing employee reviews of a large Chinese tech company. Notably, the findings highlight the importance of intangible aspects within the work environment, such as cultural conflicts, leadership style, and perceived fairness, as significant contributors to satisfaction and dissatisfaction among the company employees. Furthermore, this study reveals inconsistencies with prior research on two significant aspects. First, while work-life balance is typically linked to job dissatisfaction, this study suggests that the negative consequences of work-life balance factors can be mitigated by favorable performance in some job satisfaction-related aspects. Second, while monetary rewards undoubtedly exert a considerable influence, they may fail to ensure employee satisfaction when other key aspects of the work experience are underperforming. This research not only contributes to the body of organizational research but also offers practical implications for enhancing employee satisfaction within global tech enterprises.

Keywords: Online employee reviews, Job satisfaction, Text mining, Machine learning, Topic modeling

Introduction

Employee satisfaction plays a key role in driving long-term organizational success and competitiveness, which is often seen as a cornerstone of a company's overall performance [40]. When employees are satisfied, they tend to be more engaged, productive, and loyal, which translates into better outcomes for the organization, including sustainable growth and profitability [16, 81, 90]. On the other hand, dissatisfaction can have serious consequences, such as reduced job performance, lower morale, and

higher turnover rates [35]. This leads to significant financial costs for organizations, including lost productivity, recruitment, and training expenses [6, 62, 64, 71].

To monitor and address the factors influencing employee satisfaction, traditional methods like structured surveys and interviews are commonly used. While these approaches can be valuable, they are often limited by issues such as response bias, survey fatigue, and their inability to fully capture the complexity of the employee experience [45, 83]. In contrast, employee-generated insights on online platforms offer an alternative data source that can provide richer, more authentic insights into employees' experiences. These anonymous and voluntary reviews reflect a more candid and unfiltered perspective of the workplace environment [20], making them a valuable data source for exploring employee satisfaction in academic research (e.g., [41, 44, 79]). While online employee reviews offer opportunities to identify emergent topics and insights that traditional survey methods might overlook [24], their unstructured nature presents significant analytical challenges. These challenges necessitate sophisticated techniques to extract meaningful insights from the data [65]. To address this, this study aims to advance current research methodologies for analyzing unstructured big data by focusing on employee reviews from a large tech company based in China.

The technology industry, characterized by rapid innovation and intense competition for top talent [10], faces unique challenges in maintaining employee satisfaction [19]. High turnover rates are a persistent issue in this sector, with employees frequently leaving companies due to factors such as burnout, inadequate work-life balance, limited career advancement opportunities, and dissatisfaction with workplace culture [7, 59]. According to LinkedIn's 2020 Workforce Report, the technology sector has some of the highest turnover rates across all industries, with an average annual attrition rate of 13.2%. This turnover incurs direct costs related to recruitment and training and disrupts innovation and knowledge transfer within organizations [80].

While employee satisfaction is a significant concern across the technology industry, Chinese tech companies face additional challenges as they expand globally. Many of these companies have distinct corporate cultures [92], and their international growth has created difficulties for employees from diverse cultural backgrounds in fully integrating into the company culture. This phenomenon can hinder the long-term development of Chinese tech companies, as international employees play a crucial role in technology transfer, fostering cross-cultural collaboration, and driving innovation [53, 72]. As such, improving the satisfaction of international employees has become increasingly vital. Although existing research has extensively examined employee satisfaction within Chinese tech companies, it has predominantly focused on domestic employees (e.g., [17, 18, 52, 75]). This creates a notable gap in understanding the unique experiences and satisfaction drivers of international employees within these organizations. To address this gap, we focus on a large Chinese technology company that has achieved significant growth through the contributions of its international workforce operating worldwide. The specific objectives of this study are as follows:

Objective 1: To identify the sources of satisfaction among the company's international employees.

Objective 2: To identify the sources of dissatisfaction among the company's international employees.

Objective 3: To assess the relative impact of different satisfaction attributes on employee satisfaction levels across the company's international employees.

In this study, we gained more profound insights into the experiences of international employees across different countries at the selected company by analyzing reviews posted on Glassdoor. Glassdoor is recognized for providing valuable insights from employees across the globe [30, 69], making it a suitable platform for our research. We employed Structural Topic Modeling (STM) to analyze the unstructured data from these reviews. STM is advantageous because it allows for the inclusion of additional data (covariates) in the analysis [68], offering a deeper understanding of the text compared to traditional models like Latent Dirichlet Allocation (LDA). We analyzed positive and negative reviews separately, allowing us to differentiate factors that contribute to employee satisfaction or dissatisfaction. By incorporating review ratings into our analysis, we were able to link the prevalence of specific topics with their respective ratings, providing more precise insights into what influences employee perceptions at the company. We also used Support Vector Machine (SVM) to further validate findings from the topic modeling. This method helped confirm the relationships between text data and review ratings, enhancing the reliability of our analysis.

The findings of this study contribute to the broader understanding of employee satisfaction within a Chinese multinational technology corporation. By applying advanced text analytics techniques, we have added new insights into how employees from different cultural backgrounds perceive their work environment. Methodologically, integrating SVM with topic modeling has proven to enhance the robustness of our analysis. From a practical standpoint, the insights gained can assist HR practitioners and organizational leaders at Chinese technology companies in developing engagement strategies that cater to the diverse needs of their international workforce. Such tailored strategies will likely create a more supportive work environment, improving employee satisfaction and overall productivity.

Literature review

Employee satisfaction in tech companies

In the technology industry, employee satisfaction can significantly impact a company's long-term development. Previous research has established a positive correlation between employee satisfaction and organizational performance, such as reducing turnover rates [90], enhancing productivity [16], and increasing profitability [81]. It is often observed that compensation and benefits packages significantly impact the satisfaction levels of tech professionals. Various studies have indicated that monetary benefits, including salaries, bonuses, and different forms of financial incentives, are strongly linked to increased employee satisfaction [27, 34]. This correlation emphasizes the importance of having competitive and well-structured compensation schemes in attracting top talent and retaining them by providing a sense of appreciation and financial security. Apart from monetary compensation, other factors such as organizational culture, career growth opportunities, and work-life balance are also crucial in influencing the satisfaction levels

of tech professionals [85, 93]. Nair and Salleh [60] found that the type of organizational culture within a technology firm significantly influences employees' satisfaction. Specifically, team-oriented, diverse, and transparent cultures are associated with higher satisfaction levels among employees from tech companies.

Moreover, Moro et al. [59] suggest positive co-worker relationships contribute to a supportive environment and enhance job satisfaction. Similarly, Schneckenberg [73] found that empowerment in decision-making and open communication from leadership also contribute to cultural satisfaction among employees of high-tech companies. The technological experience of employees is found to be another important environmental factor. Lee and Kim [48] emphasize the crucial connections between technology, organizational structure, and climate. For instance, access to advanced technological infrastructure boosts efficiency and creates an engaging work atmosphere, further contributing to staff contentment [14]. Besides the work environment, professional growth opportunities are influential at technology companies. According to studies, access to training programs for advancement is positively related to satisfaction [31, 61]. The rapid pace of change within technical disciplines means ongoing learning and career progression are priorities when tech professionals evaluate employers [63]. Lastly, the ability to effectively balance work and personal life is an evolving focus area. Accommodating individual needs for work-life balance and implementing family-friendly policies represent critical components in creating satisfied teams of technical professionals [11, 46].

Covid-19 and employee satisfaction

Recent research has revealed significant shifts in employee satisfaction due to the impact of the COVID-19 pandemic. The pandemic transformed the work environment, with many employees transitioning to remote or hybrid work arrangements [76]. The adoption of flexible work models has been particularly prevalent in technology companies [8], however, these changes have had varying effects on employee satisfaction. Bocean et al. [13] found that employee satisfaction improved for some workers, as they enjoyed better work-life balance due to reduced commuting time and greater flexibility. On the other hand, remote work also led to feelings of isolation and burnout for others, particularly as long working hours became normalized [22]. Toscano and Zappalà [82] identified social isolation—an inherent aspect of remote work during the pandemic—as a significant contributor to increased stress, which, in turn, negatively affected employees' perceptions of their productivity and job satisfaction. Carraher-Wolverton [15] further emphasized that while many employees appreciated the convenience of remote work, those in highly collaborative roles faced challenges that undermined their job satisfaction. Moreover, communication was also highlighted as an important factor affecting employee satisfaction. Lee et al. [49, 50] found that transparent internal communication during COVID-19 is essential, as organizations that provide clear, honest, and timely information significantly support employees' intrinsic needs. Men et al. [56] observed that supervisory communication can significantly impact employee satisfaction and trust, mainly through three forms of motivating language: meaning-making language, empathetic language, and direction-giving language. Furthermore, the pandemic heightened employees' awareness of health, which further influenced their overall satisfaction. Gorgenyi-Hegyes et al. [33] found that healthcare support from employers during the

pandemic had a positive impact on employees' well-being, satisfaction, and loyalty. Similarly, Mihalache and Mihalache [57] observed that, in the post-pandemic era, employees continued to value employer-provided health support, which contributed to shifts in their affective commitment to the organization and overall job-related well-being.

Online employee reviews

With the rise of websites like Glassdoor and Indeed, online employee reviews are becoming valuable data sources for studying job satisfaction. Researchers have started leveraging these authentic reviews for insights into employee attitudes, supplementing traditional satisfaction surveys that have been widely applied in previous research. A primary advantage of online reviews is that they offer unfiltered feedback reflecting employees' true perspectives, shared voluntarily outside company channels [70]. On the other hand, satisfaction surveys may suffer response bias if employees feel pressured by management, while online forums' anonymity and public nature encourage candid expressions of sentiment. Consequently, previous studies demonstrate stronger correlations between review platform ratings and business performance than typically seen in survey-based research [70, 73]. Review data also provides dynamic longitudinal tracking of satisfaction over time. Xia et al. [87] analyzed hospitality employees' reviews on Glassdoor to identify the factors that lead to hospitality employees' negative experiences and changes in these factors caused by COVID-19 from 2020 to 2022. This time-series observation enables more detailed insights compared to periodic surveys. Furthermore, web-scraped review content offers unique advantages for text and sentiment analysis, particularly due to the rich, unstructured nature of the data, which is often far more abundant than open-ended responses in surveys.

Text data analytics

Topic modeling is a machine learning-based text mining technique that can aid scholars in automatically extracting hidden topics from large volumes of text documents. This approach is distinct from several other popular text analysis methods in the field. For instance, unlike semantic network analysis, which focuses on identifying the direct relationships between concepts within a text [91], topic modeling aims to uncover latent topics that are not explicitly stated. Similarly, corpus linguistics differs from topic modeling as it concentrates on the quantitative analysis of word frequency within texts rather than grouping documents into topic-based clusters [39]. Compared to these methods, text clustering techniques are similar to topic modeling, both aiming to categorize documents based on patterns in the content [2]. However, the nature of their outputs is different. Text clustering involves assigning documents to groups without further exploring the underlying topics or providing labels for these clusters. By contrast, topic modeling not only groups documents according to thematic content but also uncovers the latent topics that define these groups, thereby providing a richer layer of interpretability.

Latent Dirichlet Allocation (LDA) and STM have become two popular methods of topic modeling for extracting coherent topics from large collections of textual documents. Both methods excel at drawing out major topics from extensive datasets. A key difference is that STM allows for the direct incorporation of metadata, such as author attributes and timestamps, into the modeling process, enabling it to more precisely

match discovered topics with variables of interest and enhancing the relevance and applicability of the findings [63, 80]. In contrast, LDA lacks the native capability to use external data to guide topic discovery, limiting its ability to provide context-specific insights.

Despite STM's widespread application in academic research, limitations remain in determining the relative importance of the extracted topics. Many studies use the raw proportions of topics as proxies for priority, with more prevalent topics considered more important [25, 74]. This approach assumes that higher frequencies alone indicate greater influence or relevance to the commentators. However, Xu [89] challenged this assumption, finding that the most frequently mentioned topics are not necessarily the most impactful on user satisfaction or experience. Simply capturing common topics does not establish a direct link between the determined topics and the actual sentiments expressed by the commentators.

To overcome the limitations mentioned above in using STM, researchers have proposed integrating sentiment scoring with topic modeling to better understand the relationship between extracted topics and user ratings [5, 24]. By combining numeric scores with topic proportions, studies can more effectively assess the impact of specific topics on overall user satisfaction and identify areas in need of improvement. This approach allows for a deeper analysis that goes beyond identifying prevalent topics to focus on how these topics influence user perceptions. Leveraging the strengths of this method, this study employs STM to link topics with employee ratings, thereby identifying which topics are more prevalent in feedback from employees with higher or lower ratings. This method provides insights into aspects related to employee satisfaction or dissatisfaction, enabling organizations to prioritize their efforts in addressing the most critical issues. To further enhance the robustness of the research findings, this study also employs SVM text regression analysis to validate the results obtained from STM. By applying SVM text regression, this study can identify specific textual features that have predictive power for different ratings, offering a solution for verifying the robustness of the results derived from STM.

Methodology

Data collection and preprocessing

We collected the employees' review dataset from Glassdoor (<https://www.glassdoor.com/Community/index.htm>), a widely used platform for publishing reviews about numerous organizations in academic research. This dataset contains several attributes, including "Job Position," "Overall Rating," "Pros," "Cons," and "Management Suggestions." These data were extracted through a Python script designed for web scraping. Prior to conducting the analysis, an initial step involved filtering out all reviews not written in English to ensure language consistency throughout the dataset. Figure 1 shows the proposed research framework developed for this study.

After gathering the data, the next step was to move into the data preprocessing phase. This phase was guided by the methodologies outlined by Guo et al. (2017) and Ding et al. [24], involving several steps designed to prepare the dataset for analysis. First, the text was tokenized to break down the reviews into manageable units. Subsequent steps included the removal of numbers and punctuation, which are non-lexical characters that

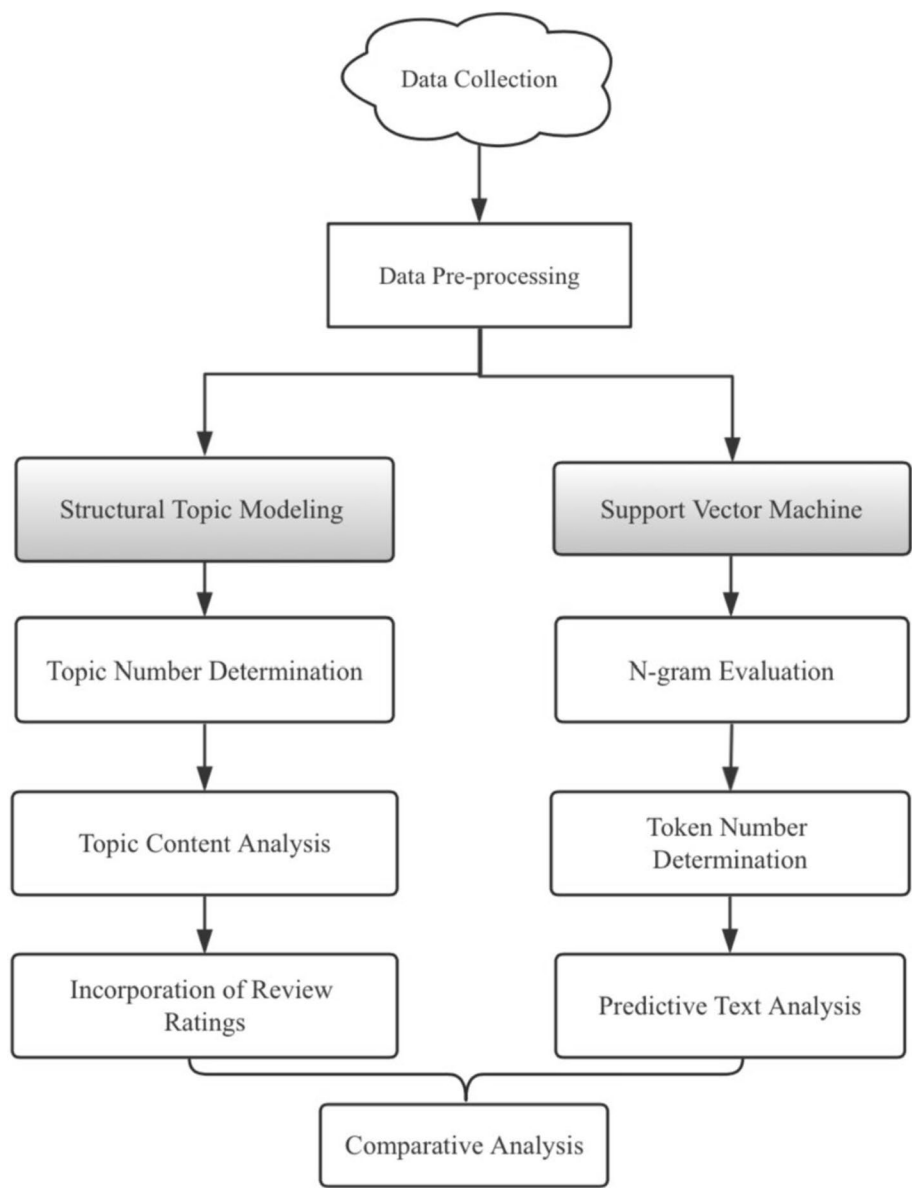


Fig. 1 Proposed research framework

do not contribute to semantic analysis. To refine the dataset further, common stop words were excluded using the SMART stop word list, which can eliminate words with minimal meaningful content. This removal process was extended to include context-specific stop words, such as specific department names and employee names within the company, aimed at reducing potential biases in the subsequent analysis. Lastly, the process moved to lemmatization, standardizing word forms to their base or dictionary form for more accurate analysis. For instance, the words “jumping,” “jumps,” and “jumped” were all normalized to the root word “jump.” Furthermore, to ensure a balanced dataset, any reviews that were missing content in either the “Pros” or “Cons” sections were removed from the dataset. After refinement, the average number of terms per Pro review was 8, while for Con reviews, it was 9. The majority of employees posting comments on Glassdoor are

from technical, engineering, and managerial positions. The most technical positions are “Software Engineer” (620 comments) and “Senior Software Engineer” (242 comments). Other notable technical roles include “Engineer” (178 comments), “Network Engineer” (157 comments), “Wireless Engineer” (70 comments), and “Senior Engineer” (96 comments). Within the managerial category, “Project Manager” (219 comments), “Product Manager” (150 comments), and “Account Manager” (101 comments) are prominent. The final corpus, comprising 19,186 online reviews, was then prepared for the subsequent analytical procedures.

Rating distribution

According to Fig. 2, employee satisfaction ratings demonstrate a distinct pattern: a score of 4 is the most frequently reported, indicating a generally positive perception among the workforce. This rating is closely followed by scores of 3 and 5, each of which surpasses 2000 instances. In contrast, the lower ratings of 1 and 2 are significantly less common, with each category not reaching 1000 occurrences. This distribution suggests that while most employees report high satisfaction levels, there exists a smaller, yet notable, segment of the workforce that expresses dissatisfaction with their experiences at the company.

Determining the topic number

We adopt a two-pronged analytical approach that combines macro-level and micro-level analysis to determine the optimal number of topics for a given corpus. At the macro level, we use an evaluation framework that assesses the performance of various topic model configurations, with the number of topics (K) ranging from 5 to 40. This framework employs four statistical metrics: held-out likelihood, residuals analysis,

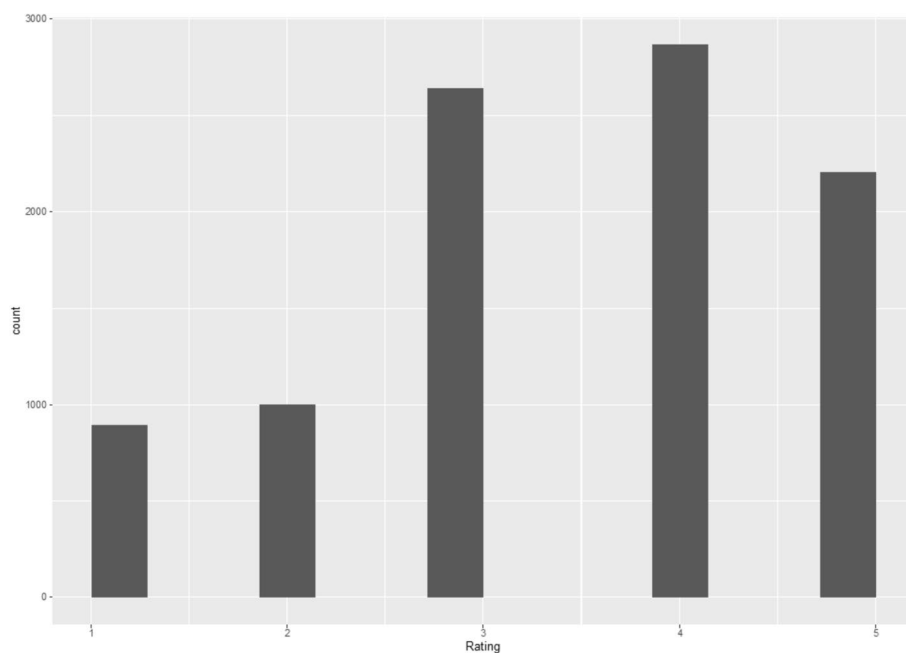


Fig. 2 Review rating distribution

semantic coherence, and the lower bound of the variational likelihood. Figure 3 illustrates that the sensitivity of the metrics increases in topic models with 20 to 25 topics for both Pro and Con reviews. The line with hollow circles represents Pro reviews, while the line with solid circles represents Con reviews. Based on this observation, this range has been selected as the optimal candidate model for further analysis.

We focus on a subset of potential topic model solutions to examine the micro-level of our analysis. Our evaluation is guided by two critical metrics: semantic coherence and topic exclusivity. Semantic coherence is an evaluative metric that has its theoretical underpinnings in the concept of pointwise mutual information. It posits that a semantically coherent model is one in which the words statistically most likely to be associated with a given topic will tend to frequently co-occur within the same document [9]. Empirical evidence suggests that topic models demonstrating superior clustering efficacy are characterized by heightened semantic consistency within individual topics and a pronounced semantic demarcation between disparate topics [3]. Supporting this notion, the work of Minmo et al. [58] illustrates a positive correlation between semantic coherence scores and human evaluations of topic pertinence and quality. The mathematical formulation for semantic coherence (C_k) for a given topic (k) is as follows:

$$C_k = \sum_{i=1}^M \sum_{j=1}^{i-1} \ln \left(\frac{D(v_i, v_j)}{D(v_i)D(v_j)} \right) \quad (1)$$

where $D(v_i, v_j)$ symbolizes the document frequency of the co-occurrence of words (v_i) and (v_j), and (M) represents the number of most probable words within the topic (k).

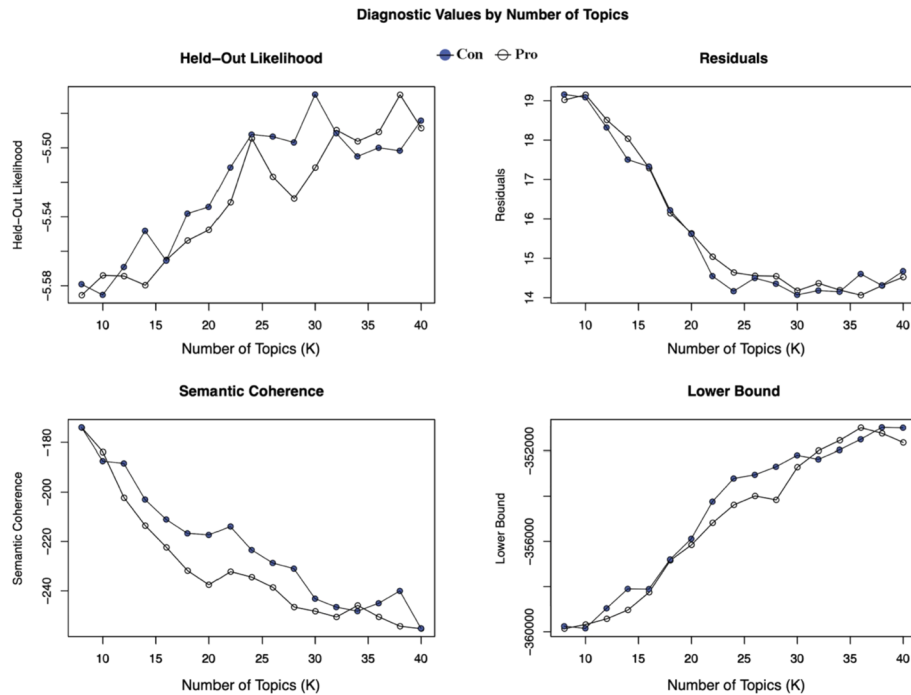


Fig. 3 Statistical metrics of different topic solutions

Semantic coherence plays a crucial role in optimizing topic models, significantly enhancing the quality of topic mining and aligning it more closely with human subjective evaluations [51]. However, relying on semantic coherence as the sole metric to determine the ideal number of topics might lead to oversaturated topic representations with high frequency but less discriminative words. To address this issue, Airoldi and Bischof [3] proposed a metric known as the FREX score (Frequency and Exclusivity of Terms). FREX aims to correct potential biases by focusing on the uniqueness of words across various topics rather than their mere frequency. The calculation of the FREX score considers both the frequency of terms and their exclusivity across different topics, which can help differentiate topics. The FREX value is modeled as follows:

$$FREX_{k,v} = \left(\frac{\omega}{F(\beta_{k,v} / \sum_{j=1}^K \beta_{j,v})} \right) + \left(\frac{1 - \omega}{F(\beta_{k,v})} \right)^{-1} \quad (2)$$

where k is the k -th topic, v is the term under consideration, β is the word distribution of the k -th topic, ω is the prior used to impose exclusivity, and the weight ω is set to be 0.7 by default.

Figure 4 shows that, for both Pro and Con reviews, no topic achieved superior performance across both metrics. However, the topic model with 22 topics demonstrated a more balanced performance in terms of semantic coherence and topic exclusivity in both Pro and Con reviews. Even though statistical metrics can guide the selection of an appropriate number of topics, the process still requires subjective evaluation based on the research objectives and corpus characteristics. After qualitatively comparing models with 23 and 24 topics, we found that the 22-topic model had fewer overlapping topics and better interpretability. Therefore, 22 topics were

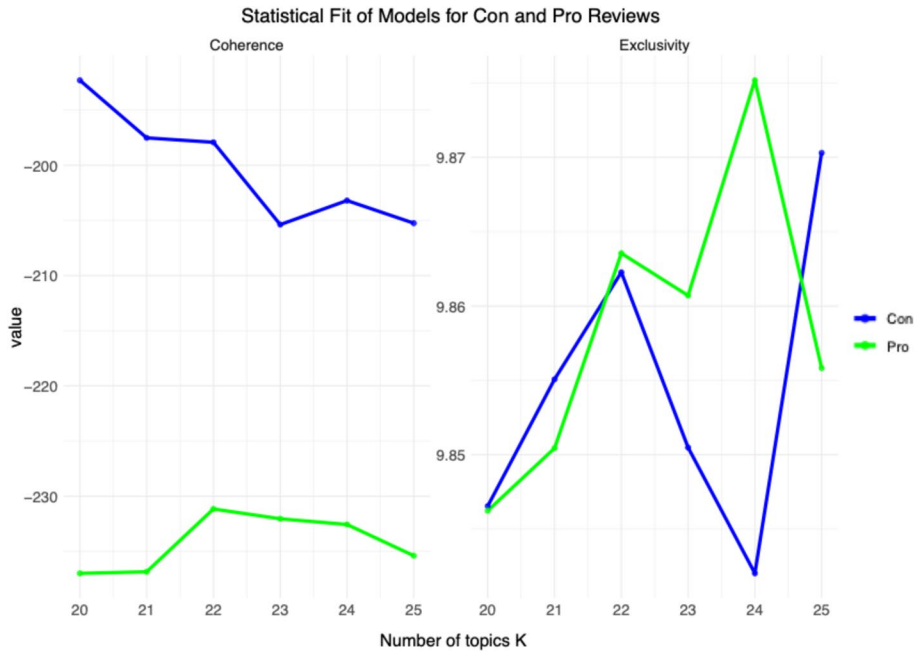


Fig. 4 Average semantic coherence and exclusivity comparison

selected for both sets of reviews. Following the code provided by Roberts et al. [67], the subsequent analysis was performed using the `stm()` function in R programming with $K=22$ topics, chosen based on model fit criteria. The model was run for a maximum of 75 iterations. Employee satisfaction ratings were included as a covariate in the prevalence formula to assess their impact on topic prevalence. The “Spectral” initialization method was used for stability, and a random seed of 706 was set to ensure reproducibility.

Findings

Topic labeling

The process of labeling each topic involves analyzing both the top words and representative sentences associated with that topic. Typically, the most utilized words for labeling are the ones with the highest probability of occurrence, as shown in Figs. 5 and 6, which frequently occur within their respective topics and hold substantial thematic value. However, relying solely on high-probability words can sometimes lead to an overemphasis on common terms, which may obscure the meanings of the topics and complicate their interpretation. To mitigate this issue, FREX (frequent and exclusivity) words are also considered when the highest probability words do not adequately capture the essence of the topic. Representative sentences are examined to understand the context in which the top words are used, aiding in refining the selection of topic labels.

The finalized topic labels are listed in Table 1. These topics have been organized according to the Employee Satisfaction Index (ESI) model proposed by Hsu and Wang [36]. This framework is specifically developed for technology companies, which aligns well with the context of the company as a high-tech firm. It is evident from the analysis that the distribution of topics derived from Pro and Con reviews across different satisfaction dimensions shows significant variation. Figure 7 presents the expected

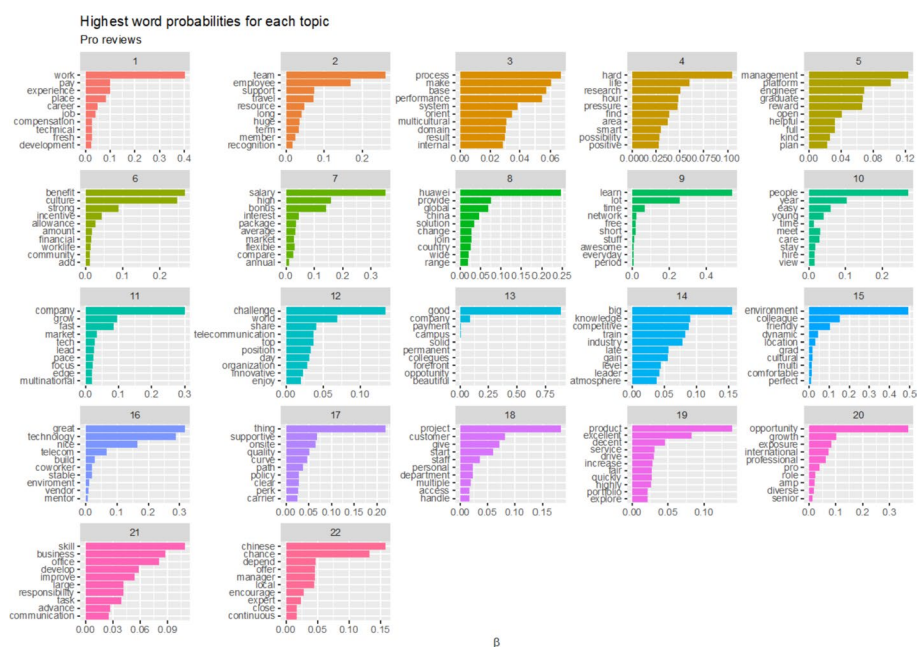


Fig. 5 Word probabilities of topics extracted from Pro reviews

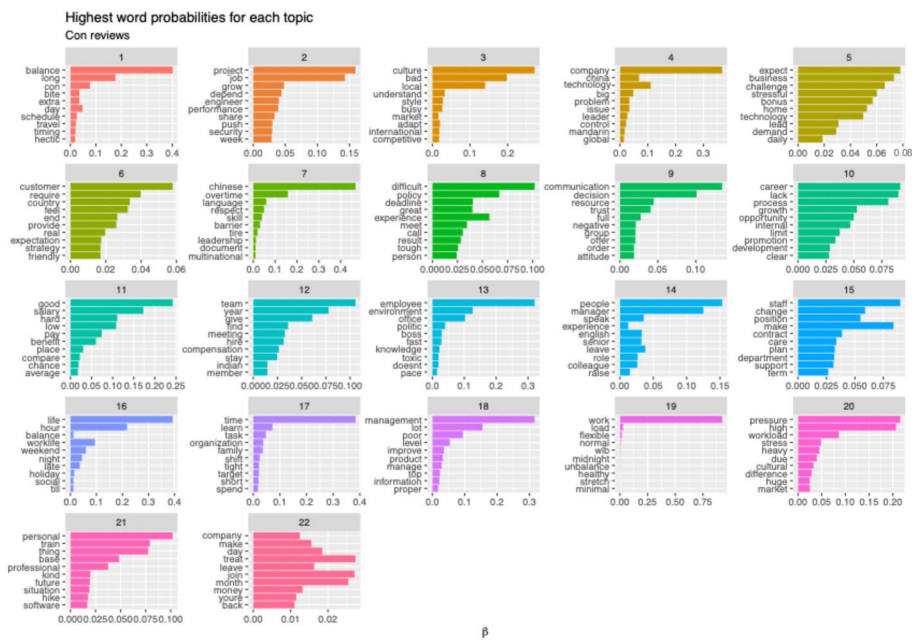


Fig. 6 Word probabilities of topics extracted from Con reviews

Table 1 Satisfaction attributes classification

Dimension	Pro	Con
Extrinsic rewards	Work conditions, Company incentives, Compensation and Benefits, Competitive salary, Good payment	Work-life imbalance, Workplace environment concerns
Supervisory support	Employee support, Clear business process, Supportive company policies	Poorly Fitted Leadership Styles, Language barrier, Poor communication, Poor management, Ineffective workforce planning, time constraint
Fairness	Differential Treatment of Employees	Pay equity, Unequal treatment, Promotion inequities in top Management, Employee mistreatment
Autonomy		Inflexible policies, Long Working hours, Work inflexibility
Corporate image	Global Presence, Fast growth of the company, competitive advantage of the company, customer-focused culture, Product and service	
Affinity	Employee characteristics, Friendly co-workers,	Cultural adaptation difficulties
Development	Positive career Development, Learning and Development Opportunities, Training programs, International growth opportunities, Skill development	Job growth difficulties, Limited growth opportunities, Training issues
Others	Positive work environment	Demanding customers, Heavy workload, High work pressure

proportion of topics extracted from both Pro and Con reviews. It shows that discussions on company benefits are more prevalent in Pro reviews, while job-related challenges are more frequently discussed in Con reviews.

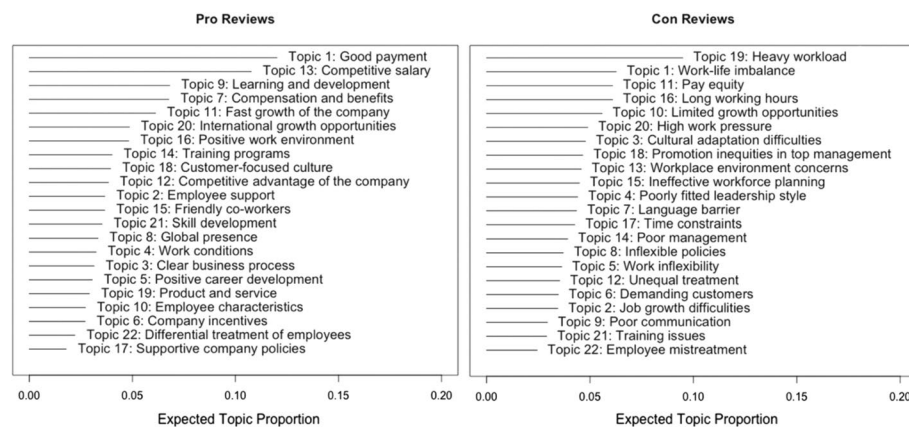


Fig. 7 Topic ranking

Topic interpretation

Extrinsic rewards

In the dimension of “Extrinsic rewards,” the company employees’ Pro reviews indicate a different focus of appreciation for the company’s extrinsic rewards. Pro reviews commend the work conditions at the company, emphasizing well-equipped workspaces and flexible lunch breaks, although there are hints at the need for better work-life balance. Company incentives appear to be an important source of satisfaction, with employees valuing transparent and merit-based rewards, while the compensation and benefits package is seen as competitive, including fair salaries and graduate starting salaries. Specifically, payment satisfaction is frequently mentioned, suggesting that employees feel they are remunerated adequately for their work.

In the con reviews, “work-life imbalance” points to employee reviews that frequently cite excessive working hours and intense pressure to perform, which can lead to stress. “Workplace environment concerns” suggest that employees find the physical and social aspects of their work setting to be suboptimal, including the political nature of the office environment, the treatment of employees, especially those relating to diversity, and possibly international or Asian employees, and concerns with how the company handles inclusivity and international employees. The con reviews point to dysfunction in non-monetary aspects, potentially undermining well-being. Frequent references to excessive hours and intense performance norms indicate an imbalance in prioritizing outputs over private life. Additional concerns center on unfavorable psychosocial attributes of the work environment. Dissatisfaction with political office dynamics and unequal treatment of marginalized identities, such as international employees.

Supervisory support

The dimension of “Supervisory support” at the company includes both positive and negative feedback from employees. On the positive side, employees appreciate the robust support from supervisors, which aids in their professional growth. The clarity of business processes, communicated effectively by supervisors, ensures a transparent understanding of tasks and projects, reducing potential confusion in daily activities. Additionally,

supportive policies such as social programs for new joiners, expatriate compensation packages, and opportunities for learning and skill development are highly valued.

However, there are notable criticisms regarding supervisory support. These include leadership styles that are poorly matched with the needs of a diverse international workforce, leading to language barriers and communication issues between predominantly Chinese management and non-Chinese staff. Employees also expressed frustration with what they perceive as poor management practices, including unfair treatment, a lack of support for development, and ineffective workforce planning that results in skill-job mismatches and staff shortages. Moreover, the imposition of stringent time constraints by supervisors is reported to create a high-pressure work environment, further highlighting the critical areas for improvement in supervisory support at the company.

Fairness

Topics under the dimension of “Fairness” indicate the complexity of the company’s workplace dynamics. Specifically, topics extracted from Pro reviews show the perceived preferential treatment Chinese employees receive in terms of promotions and opportunities compared to international colleagues. Con reviews more explicitly point to fairness issues. Employees express dissatisfaction with a lack of pay equity, implying remuneration does not align with perceptions of compensation deserved given roles and efforts. Some topics also suggest instances of favoritism or discrimination based on unspecified demographic or personal attributes. Promotion inequities at top management levels are highlighted, with international staff facing difficulties advancing despite solid performance over many years. Reports of employee mistreatment, including a range of negative behaviors. Specifically, reviews allege disrespect towards international employees and termination of employment without proper justification.

Autonomy

Topics extracted from Con reviews revealed that the company culture might be overly strict and harmful to employee independence and well-being. The representative reviews show that employees are unhappy with inflexible company policies that limit their freedom and adaptability in their work. Additionally, consistent complaints about long working hours indicate a work environment that encourages excessive overtime, which can lead to burnout. This inflexibility is further exacerbated by a lack of flexible work arrangements, such as remote work options or adaptable scheduling, that could help promote a better work-life balance for employees.

Corporate image

In the dimension of “Corporate image,” the topics extracted from Pro reviews reflect a strong sense of pride in the company, such as its global presence, fast growth, and competitive market position. Employees are proud to be part of an extensive multinational corporation that offers diverse cultural and professional experiences and opportunities for advancement within an aggressively evolving industry. They also express their satisfaction with the company’s focus on maintaining a competitive advantage through innovation and a customer-centric ethos. Additionally, there is an explicit endorsement of

the quality of the company's products and services, which may reflect employees' enthusiasm for the brand.

Affinity

In the "Affinity" dimension of Pro reviews from the company employees, the workplace is described as one where individuals are valued for their professional attributes, and friendly, supportive, collegial relationships are prevalent. Positive employee characteristics such as professionalism, dedication, and expertise are recognized, and the emphasis on friendly colleagues suggests that the company fosters a warm and collaborative environment. However, in Con reviews, the topic of "cultural adaptation difficulties" highlights issues with integrating into the company's work culture.

Development

In the "Development" dimension, topics extracted from Pro reviews by the company employees describe the company's substantial investment in its staff's professional growth and support. Employees acknowledge the numerous opportunities for career and personal development, such as upward mobility and internal promotions supported by the company. They mentioned that the company provides ample learning resources and structured training programs, investing in skill enhancement to ensure employees are equipped to meet current and future industry challenges. Moreover, the company's global operations offer valuable international growth opportunities to its staff, allowing for diverse experiences and cross-cultural skill development.

However, topics extracted from the Con reviews indicate a lack of employment growth opportunities, limited prospects for advancement, and insufficient training and development support. Employees feel that their ability to develop careers, expand skills, and fully realize their potential is restricted within the company. The main reasons for employment growth difficulties are job descriptions that do not match reality, frequent changes in work arrangements, and a lot of uncertainty. The topic of "limited growth opportunities" revolves around concerns about career development, internal opportunities, and the limitations of the company's organizational structure. Additionally, some employees complain about the lack of necessary training programs, poorly designed training materials, and many materials written in Chinese, making it difficult to understand the content.

Others

The topic of "positive work environment" extracted from Pro reviews suggests a focus on the company's positive work environment within the technology sector. Employees highlighted the multicultural team and partnerships across different regions, a stable work culture that fosters learning, and the ability to scale operations globally. The topics extracted from the company employees' Con reviews highlight significant company workforce challenges. These challenges include coping with demanding customers, managing a heavy workload, and dealing with high work pressure. The analysis suggests that employees may be grappling with meeting customer expectations, possibly due to the competitive nature of the industry or unique client requirements. Some employees express dissatisfaction in their reviews by highlighting that the company prioritizes

customer trust over employee concerns during conflicts. Additionally, employees report feeling overwhelmed by their workload, which may contribute to increased work pressure.

Topic distribution analysis

In the context of high employee satisfaction ratings shown in Fig. 8, certain topics stand out as particularly influential in shaping a cohesive narrative that places great emphasis on the social, developmental, and cultural aspects of the workplace. Specifically, the prevalence of topics such as “employee support,” “positive career development,” “friendly co-workers,” and a “positive work environment” suggests that these elements are deeply valued by employees. Of these topics, “employee support” is particularly noteworthy for its emphasis on the provision of resources and assistance to employees. This focus on nurturing employees closely aligns with the nurturing aspect of a positive work environment. Additionally, the presence of “friendly co-workers” is significant in that it not only enhances day-to-day job satisfaction but also fosters a sense of community and belonging among employees. Meanwhile, “positive career development” offers a forward-looking element, indicating that employees perceive a clear trajectory for growth and advancement within the organization. This emphasis on long-term career prospects is complementary to the more immediate benefits of a supportive and congenial workplace. While these topics overlap in their contribution to overall employee contentment, they differ in their primary focus: “employee support” and a “positive work environment” often refer to the present conditions and infrastructural backing, while “positive

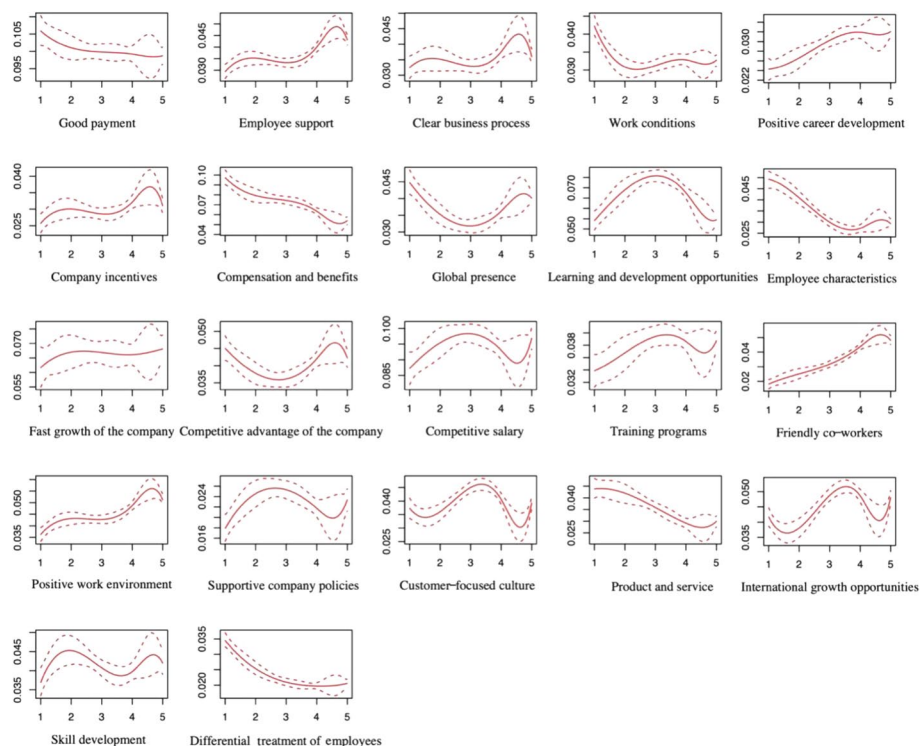


Fig. 8 Topic distribution by ratings (Pro reviews)

career development” points to future-oriented aspirations and the potential for personal progression within the company.

As for topics discussed more in lower satisfaction ratings, including “good payment,” “work conditions,” “compensation and benefits,” “global presence,” “employee characteristics,” “product and service quality,” and “differential treatment of employees.” Despite traditionally being significant factors in job satisfaction, “good payment” and “compensation and benefits” appear insufficient to offset employee discontent. This indicates that non-monetary issues can significantly impact employees’ evaluation of satisfaction in the company. Similarly, “global presence” and “product and service quality” represent the company’s external successes and market reputation that employees acknowledge but do not directly translate into positive job sentiment, suggesting a possible disconnect between company achievements and individual employee experiences. The topic of “employee characteristics” suggests a potentially skilled and diverse workforce, but the perception of “differential treatment” indicates that any positive aspects of employee dynamics can be influenced by a sense of unfairness or inequality within the organization. This highlights the importance of ensuring equitable treatment and fostering a culture of inclusivity, where all employees feel valued and respected regardless of their characteristics.

Findings shown in Fig. 9 highlight the prevalence of topics such as “work-life imbalance,” “job growth difficulties,” “time constraints,” “heavy workload,” “high work pressure,”

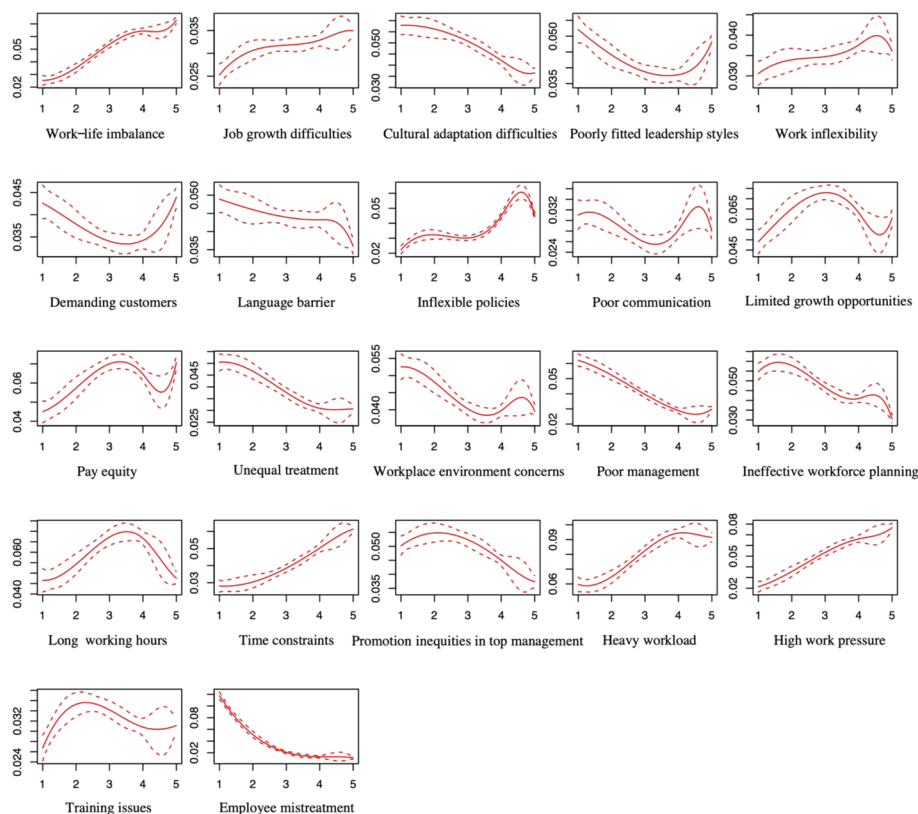


Fig. 9 Topic distribution by ratings (Con reviews)

and “inflexible policies” in highly satisfied employee reviews at the company. The juxtaposition of these topics presents a complex picture, as these topics all describe aspects of a work environment that can lead to burnout and dissatisfaction, yet their frequent mention in positive reviews suggests that some employees perceive them as integral to professional success and personal growth or are outweighed by other job benefits. These topics each touch upon a unique facet of the work experience that can significantly shape an employee’s perception of their job. The fact that these topics are cited in highly positive reviews may point to a culture that equates high demands with high rewards or to a subset of employees who thrive under pressure and view these challenges as a necessary component of a rewarding career at a leading tech company. While certain work-related stressors may be considered undesirable, some employees may perceive them as a means to achieve personal and professional growth, particularly in a highly competitive and fast-paced industry.

A substantial portion of the lowest satisfaction ratings among the company employees reflects issues with “cultural adaptation difficulties,” “poorly fitted leadership styles,” “language barriers,” “unequal treatment,” “poor management,” “ineffective workforce planning,” and “employee mistreatment.” These issues exist at both macro and micro levels within the organization. Macro issues, such as ineffective workforce planning and poor management, reflect systemic organizational deficiencies. On the micro level, individual experiences, such as mistreatment and language difficulties, highlight the personal challenges faced by employees. The prevalence of these issues among the lowest satisfaction ratings suggests a significant focus on the company’s management teams and raises concerns that encompass cultural barriers and communication issues, as well as larger organizational management and leadership practices.

Text prediction using SVM regression

The dataset of the company employee reviews was prepared for analysis by partitioning it into a training and test set, with an 80/20 split, respectively. The Term Frequency-Inverse Document Frequency (TF-IDF) technique was then utilized to filter out commonly occurring words that contribute minimal substantive meaning to the analysis. TF-IDF is a widely used method in natural language processing (NLP) applications, as it evaluates a word’s significance in a document relative to a collection of documents or corpus.

The TF-IDF score is derived from two metrics: term frequency and inverse document frequency (IDF). Term frequency (TF) computes how often a given word (term t) appears in a document (d), normalized by the document’s word count, as indicated in Eq. (3) (Jurafsky and Martin, 2000):

$$tf(t, d) = \frac{\text{count of } t \text{ in } d}{\text{number of words in } d} \quad (3)$$

In the case of examining word counts within employee reviews, the inverse document frequency (IDF) gauges the significance of a word across the entire corpus. Common terms receive a lower IDF. The calculation involves dividing the total number of documents by the number of documents containing the term (document frequency) and then

taking the logarithm of this quotient to prevent division by zero issues and moderate the influence of high frequencies:

$$idf(t) = \log\left(\frac{N}{df + 1}\right) \quad (4)$$

To deduce the TF-IDF score for a particular term, we multiply its TF and IDF values:

$$tf - idf = tf(t, d) \cdot \log\left(\frac{N}{df + 1}\right) \quad (5)$$

N-gram evaluation

We employed n-gram models to conduct a linguistic analysis. N-grams, which are contiguous sequences of words from a given text, include unigrams (single words), bigrams (pairs of words), and trigrams (triplets of words). This methodology facilitated an exploration of how varying levels of contextual information impact the predictive efficacy of the model [47]. To assess the models, we utilized the Root Mean Squared Error (RMSE). RMSE is particularly appropriate for evaluating SVM regression models because it quantifies the average magnitude of the prediction errors, thereby providing a robust indicator of the model's precision. As shown in Fig. 10, the bigram model demonstrated the lowest RMSE. Consequently, we opted to proceed with the bigram configuration for more detailed analysis.

Determining the number of tokens

To enhance the model's performance, it is important to adjust the token count, given that the size of the vocabulary is intrinsically linked to the model's precision. We employed three evaluation metrics to ascertain the optimal token quantity, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and RMSE. Through methodical experimentation, it was established that a vocabulary size comprising 250 tokens for both Pro and Con reviews yielded the most balanced efficacy across these statistical indicators. This token count marked the point where the model exhibited superior proficiency, effectively aligning with key indicators of predictive accuracy. Based on the convergence of these findings shown in Fig. 11, we ultimately selected a 250-token configuration for our model.

SVM model building

Following the guidelines outlined by Hvitfeldt and Silge [37], the model-building process began with data preprocessing, which included tokenizing employee reviews into n-grams ($n=2$). Token filtering was set to 250 to ensure that the bigrams both had predictive power and a certain frequency. This filtering approach enhances qualitative assessment, as low-frequency terms may possess predictive value but offer limited interpretability [88]. In addition, a limitation of SVM regression is its tendency to become computationally expensive and less efficient when handling large datasets with high-dimensional features [86]. We then applied TF-IDF transformation, followed by normalization of the predictors. To evaluate the model's performance, we implemented k-fold cross-validation using the "cv" method, dividing the training dataset into 10 folds. This

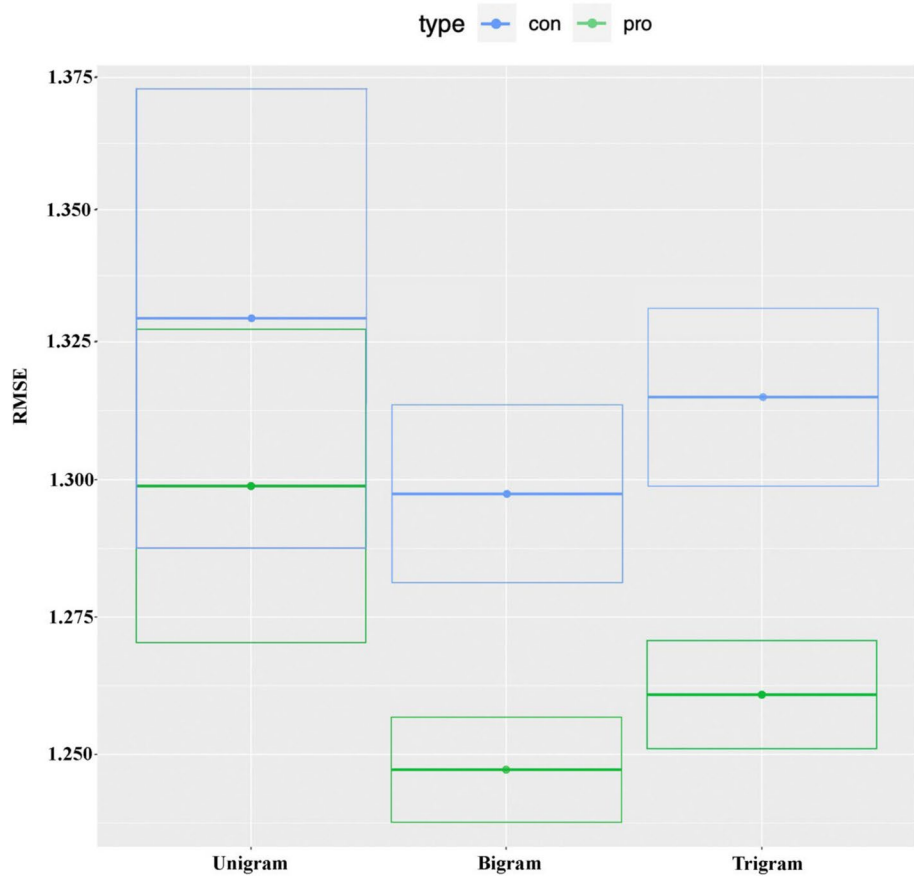


Fig. 10 N-gram performance of reviews

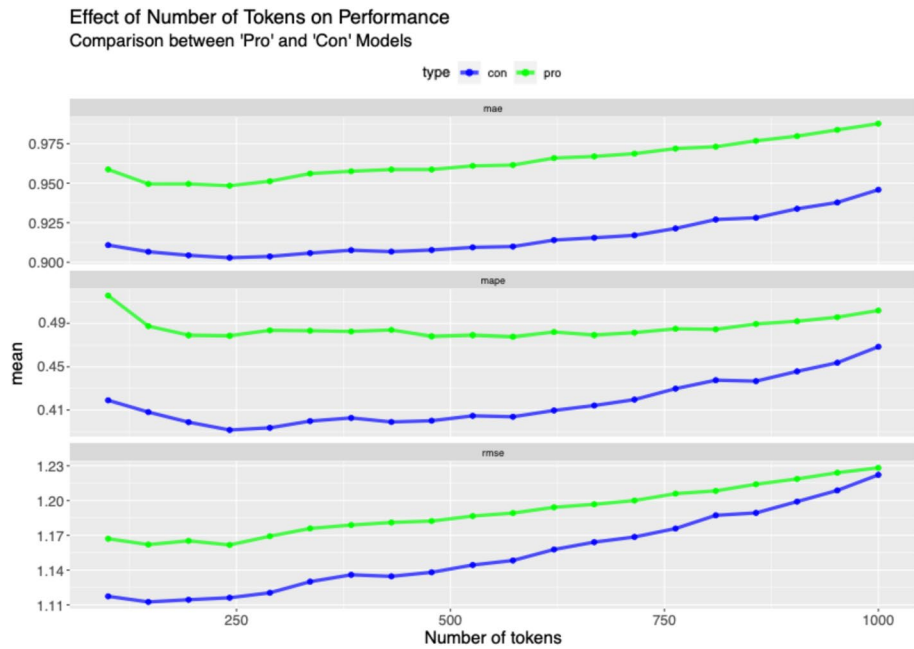


Fig. 11 Performance of different token numbers

approach helps assess the model's generalizability and mitigate overfitting. After training the model across these folds, we performed the final fitting using the entire training dataset and evaluated its predictive performance on the test set. The performance of the final model for Pro reviews yielded an RMSE of 1.2 and an MAE of 0.992, while for Con reviews, the RMSE was 1.19 and the MAE was 0.969.

Interpretation of SVM results

Figures 12 and 13 provide a comprehensive graphical representation of the fifty most influential bigrams that influence the levels of job satisfaction and dissatisfaction among international employees of the company. In these figures, the bigrams are represented by blue and red bars, with the length of each bar directly proportional to the predictive strength of that bigram in determining employee satisfaction outcomes. The predictive bigram analysis in Fig. 12 reveals a strong correlation between high job satisfaction and a supportive, dynamic work environment. This finding highlights the significance of a corporate culture that fosters effective leadership, teamwork, and professional development. When compared with the results of STM, it becomes evident that “positive career development” is frequently discussed in reviews with higher satisfaction ratings, with a comparatively much higher estimated proportion of 0.03 in ratings 4 and 5. The associated bigrams have positive estimated values, such as “improve skill” (0.048), “train development” (0.047), and “learn environment” (0.046), which also show a strong predictive relationship with employee satisfaction. This emphasizes that career advancement and skill development opportunities are critical factors in driving higher levels of satisfaction. Furthermore, “positive work environment” also emerges as an important factor associated with higher satisfaction, with an estimated proportion of 0.038 in rating 5. The predictive bigrams related to this topic, including “good environment” (0.083), “work environment” (0.063), and “highly motivated” (0.058), demonstrate the significant role that a positive and motivating work atmosphere plays in fostering employee contentment.

In contrast, employees expressing lower satisfaction scores frequently cite compensation-related concerns. Contrary to initial expectations, material rewards, such as compensation and benefits, appear to have a less significant impact on overall employee satisfaction within the company. Specifically, the topic of “good payment” is most strongly associated with lower satisfaction ratings, with an estimated proportion of 0.107 in rating 1, the highest among all topics. Similarly, the topic “compensation and benefits” is linked with an estimated proportion of 0.098 in rating 1. These findings are consistent with the results from SVM analysis, which identifies “salary time” as the most significant predictor of lower satisfaction ratings, with a predictive value of -0.061. The bigrams “salary international” and “salary compare” also show negative associations with satisfaction, with predictive values of -0.052 and -0.049, respectively. This suggests that employees' dissatisfaction is strongly linked to perceptions of being underpaid or unfavorable salary comparisons, particularly on an international scale. Additionally, some employees view the company primarily as an opportunity for skill acquisition rather than as a long-term career destination, as reflected in bigrams like “possibility learn” and “good CV.” Despite recognizing the company's leadership in the technology sector, employees express substantial dissatisfaction with certain cultural and managerial aspects. The

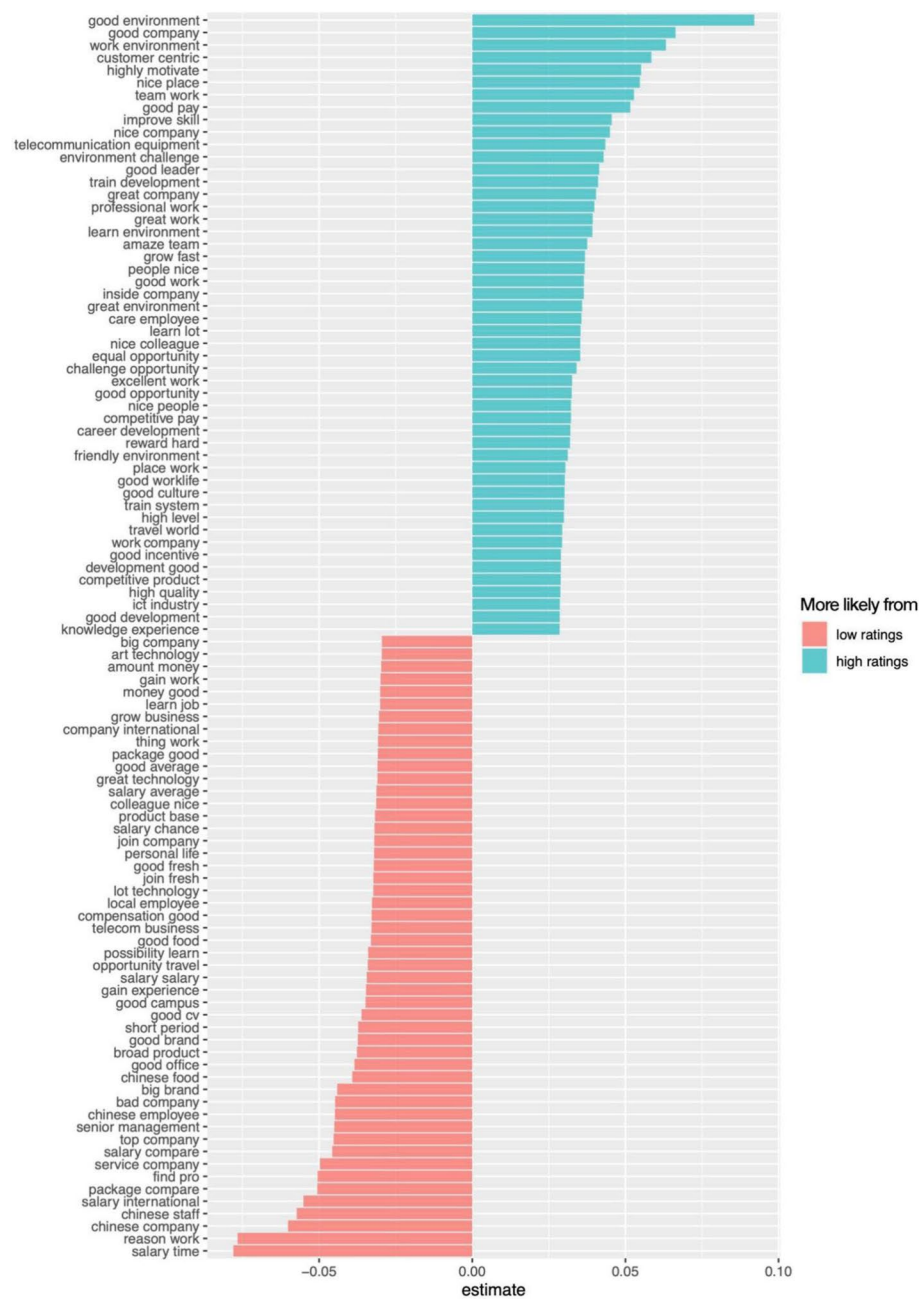


Fig. 12 Texts with strong predictive power (Pro reviews)

predictive bigrams highlight this dichotomy, revealing that while employees appreciate the company's technological acumen and professional growth opportunities, these benefits are insufficient to mitigate the adverse effects of inadequate compensation, cultural challenges, and senior management practices.

Figure 13 illustrates certain negative aspects that emerge in higher satisfaction ratings, although they may not significantly affect overall employee satisfaction. For example, employees express concerns about high turnover, indicating potential issues with how the company manages organizational changes and treats its staff. Additionally,

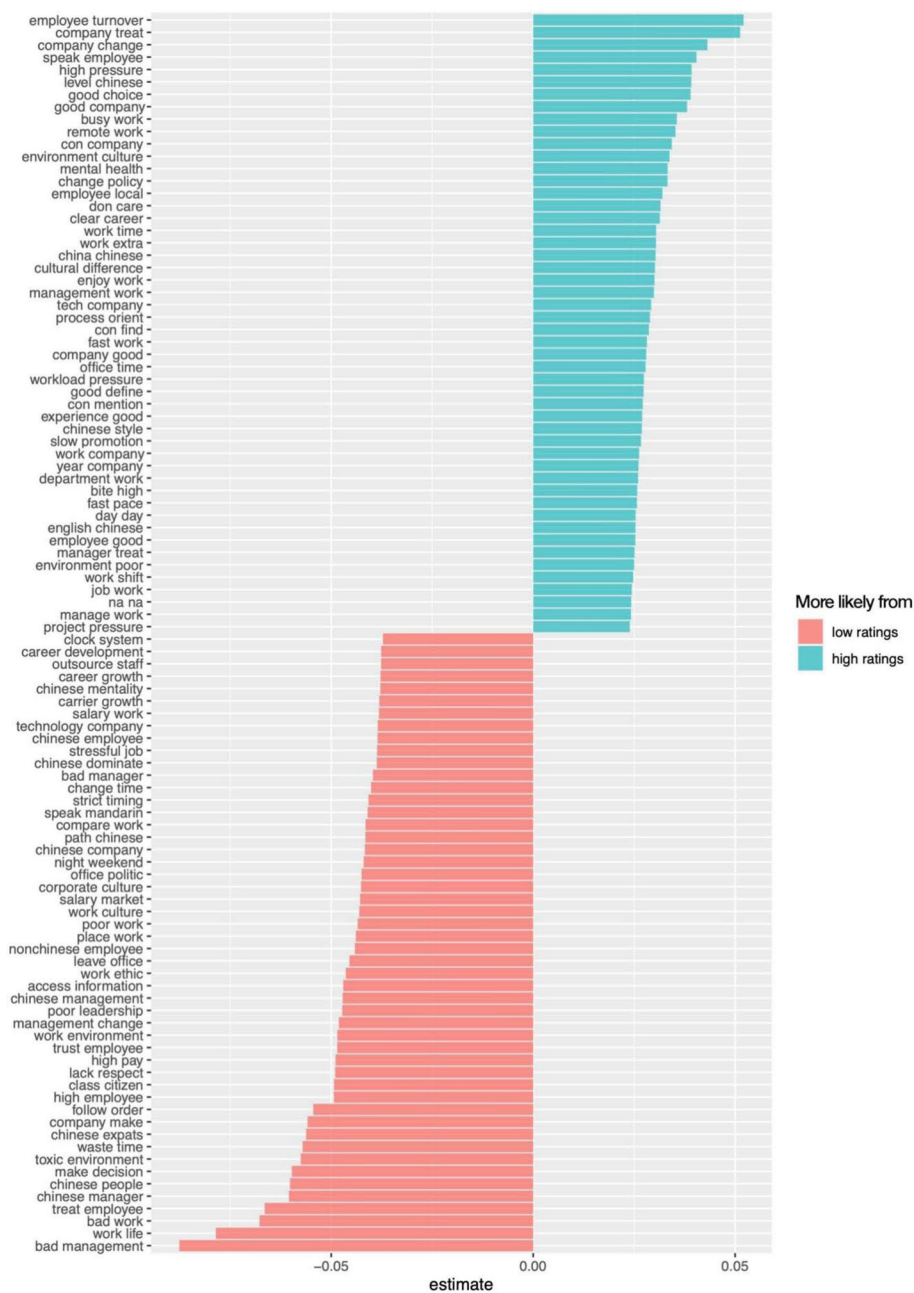


Fig. 13 Texts with strong predictive power (Con reviews)

high-pressure workloads, which can lead to work-life imbalance and negatively affect mental health, are also highlighted. These findings align with the results from STM, where topics such as “work-life imbalance” (0.072 in rating 5), “job growth difficulties” (0.035 in rating 5), and “high work pressure” (0.078 in rating 5) show increasing proportions as satisfaction ratings rise. This may suggest that employees with higher levels of job satisfaction might view these challenges as integral components of their professional development journey. The related bigrams, such as “high pressure” (0.033), “busy work” (0.031), and “mental health” (0.029), exhibit a relatively more substantial predictive

power in relation to employee satisfaction. Despite these pressures, employees seem to value certain aspects of their work environment and management, though these are tempered by the stress of fast-paced work and performance demands.

On the other hand, aspects associated with lower satisfaction ratings highlight several significant concerns. Career progression is particularly problematic, with perceptions of bias related to “Chinese management” contributing to feelings of exclusion among non-Chinese employees. In addition, management practices are frequently criticized, with terms like “bad manager,” “poor leadership,” and “lack of respect” signaling a disconnect between employees and their supervisors. Despite relatively high salaries, employees express a pervasive dissatisfaction with the work atmosphere, which is often described as a “toxic environment” marked by “office politics” and a negative “corporate culture.” This dissatisfaction is further compounded by the demands of overtime, which negatively impact work-life balance. This is reflected in the higher prevalence of the topic “cultural adaptation difficulties” in lower satisfaction ratings, with an estimated proportion of 0.053 in rating 1. Communication issues, such as the need to “speak Mandarin” and the perception of opaque decision-making processes, also exacerbate dissatisfaction. These concerns are evident in topics like “language barrier,” which shows an estimated proportion of 0.047 in rating 1, compared to 0.036 in rating 5, highlighting the negative impact of language and communication challenges on employee satisfaction.

Discussion

Key findings

The study provides a comprehensive understanding of the key factors driving employee satisfaction and dissatisfaction at a large Chinese tech company. Through the novel application of STM and SVM, the research has uncovered several valuable insights, which are discussed below.

The attributes categorized within the company’s “extrinsic rewards” dimension, such as “work conditions,” “compensation and benefits,” and “competitive salaries,” confirm the influence of tangible rewards on employees’ evaluation of job satisfaction. These rewards are found to be important for attracting and retaining talent in a competitive field where such factors can directly influence employee motivation and engagement [27, 34]. However, the presence of “work-life imbalance” and “workplace environment concerns” highlighted in the Con reviews suggests that while extrinsic rewards are crucial, they may not suffice to ensure overall employee satisfaction if other aspects of the work experience are compromised. Attributes in the dimension of “Supervisory support” emphasize the importance of high-quality relationships between supervisors and subordinates. However, the topics extracted from Con reviews highlight the challenges the company faces with leadership and communication barriers. This tension between the positive attributes identified in the Pro reviews and the challenges highlighted in the Con reviews reflects the complex reality of managing a diverse global workforce [1].

As for topics related to “Fairness,” the presence of concerns related to differential treatment of employees, pay equity, and promotion inequities in the Con reviews give rise to the importance of fairness in a company with a diversified workforce. Previous studies also highlighted the importance of addressing fairness concerns to maintain a positive and satisfied workforce [21, 23, 30]. In the dimension of “Autonomy,” the lack

of autonomy, as indicated by “inflexible policies,” “long working hours,” and “work inflexibility” in the Con reviews, is noteworthy. The pursuit of flexible working has become increasingly important in recent years, particularly for tech professionals. Researchers have shown that flexible work arrangements can lead to increased job satisfaction, work-life balance, and productivity, [35, 52]. However, the tension between the company’s rigorous and structured work environment and the tech industry’s emphasis on creativity and flexibility illustrates the ongoing debate between the need for organizational control and the pursuit of innovation, as these two priorities can often be in conflict. [29].

The dimensions of “Corporate Image” and “Affinity” highlight the importance of employees’ satisfaction derived from their association with a reputable and successful organization, as well as from positive relationships with their co-workers. This sense of belonging and commitment is fostered by the attributes within these dimensions. On the other hand, the “Development” dimension emphasizes the significance of positive career development, learning opportunities, and skill development, as evident in the Pro reviews. These attributes are in line with the fast-paced nature of the high-tech industry, where continuous learning is not only essential but also a prerequisite for growth [54].

In terms of the relative importance of topics to the company’s employee satisfaction, both the SVM and STM analyses consistently emphasize the significant role of the work environment as a key source of satisfaction. This emphasis aligns with the findings of Google’s Project Aristotle, which identified psychological safety, dependability, and structure clarity as crucial elements of effective teams [28]. Contrary to numerous studies that highlight the importance of financial incentives in job satisfaction [42, 43], the findings suggest that monetary benefits do not significantly influence employee satisfaction at the company. This may be attributed to the context of high-tech firms, which are often assumed to use high salaries and bonuses as primary motivators. One intriguing aspect of the findings is the positive association between the challenges of work-life balance, such as tight schedules and high pressure, and higher satisfaction ratings. This counterintuitive result may reflect a unique characteristic of high-tech company cultures, where engagement with challenging and meaningful work is highly valued [4]. Employees in such environments may derive satisfaction from overcoming difficulties and achieving their goals, thus viewing the demands placed upon them more positively. Lastly, the analysis further highlights inequitable treatment, particularly concerning promotion opportunities for non-Chinese staff and the poor implementation of cultural localization, as a significant source of employee dissatisfaction at the company. This issue is particularly salient in a global corporation where equitable practices across diverse cultural backgrounds are expected to be a cornerstone of corporate ethics and employee engagement strategies.

Implications

The findings of our study have several implications for academic researchers and business managers in HR departments. Firstly, our study questions the widely accepted notion that financial rewards are a universal driver of job satisfaction. Traditional theories and previous research often highlight financial compensation as a critical motivator across various sectors. However, our investigation indicates that the link between financial incentives and job satisfaction in high-tech environments is weaker than typically

assumed. Our findings suggest that, within the high-tech sector, financial aspects might play a minor role in influencing job satisfaction, as once believed. These findings could steer future research toward developing more tailored motivational theories that reflect the unique dynamics of high-tech employees. Secondly, this research challenges the prevailing notion that work-life imbalance invariably leads to dissatisfaction. While many sectors, such as those outlined by Bhende et al. [12], consider poor work-life balance a significant detractor from employee well-being and satisfaction, our findings indicate that this may only sometimes be the case within the high-tech industry. This divergence could be attributed to distinct industry characteristics. For example, competitive remuneration structures in the high-tech sector might help employees feel adequately compensated, which could mitigate the negative impacts of work-life imbalances on job satisfaction. Moreover, the high levels of self-fulfillment and meaningful work typically found in innovative tech environments may also alter traditional views on work-life balance. However, these hypotheses remain untested empirically.

From a methodological perspective, applying advanced NLP techniques in this study, specifically STM and SVM, to analyze large-scale employee review data illustrates the vast potential of computational methods in uncovering detailed insights into employee perceptions and experiences. Integrating these analytical tools enhances the robustness and validity of the findings, offering a more comprehensive understanding of organizational dynamics. In addition, the results demonstrate that SVM regression provides valuable micro-level insights that complement topic modeling, revealing aspects that may not be fully captured through topic modeling alone. This methodological advancement provides references for future research in organizational studies.

As demonstrated, the quality of the work environment significantly impacts employee satisfaction, suggesting that business managers should prioritize creating supportive, psychologically safe, and well-structured workspaces. This alignment with insights from Google's Project Aristotle confirms the universal applicability of these factors across leading innovative environments. It provides a clear pathway for companies looking to bolster team performance and satisfaction. Moreover, this research challenges the prevailing assumption that financial incentives are the primary drivers of employee motivation in high-tech contexts. Considering this, the company's HR teams are encouraged to reassess their compensation structures and explore alternative motivational strategies. Emphasizing non-monetary factors such as purpose, growth opportunities, and work-life balance could enhance employee motivation and satisfaction. Additionally, this study highlights specific cultural integration challenges within the company, particularly regarding the treatment of non-Chinese employees. This finding suggests a critical need to enhance its cultural competence and inclusivity. Implementing sophisticated talent management practices that acknowledge and bridge cultural differences will improve internal team dynamics and position these firms for tremendous success in international markets. Lastly, the counterintuitive finding regarding work-life balance challenges suggests that the company should adopt an appropriate approach. Instead of solely focusing on reducing work hours, high-tech companies should ensure that work is perceived as meaningful and rewarding. Providing adequate support resources and systems to help employees manage stress and prevent burnout is crucial.

Limitations and future research

The focus of this study is on the company due to its distinct corporate culture within the high-tech industry, which inevitably limits the generalizability of its findings across other companies in the sector. To enhance the broader applicability of these insights, future research should aim to include a more diverse range of companies that share similar characteristics within the high-tech industry. Expanding the scope of the study to include multiple organizations will enable researchers to identify common patterns and variations in the drivers of employee satisfaction. Additionally, future studies are encouraged to integrate some relevant variables, such as the timing of review postings and the specific roles of the reviewers, into the analytical frameworks. Expanding the scope of variables considered will enable researchers to explore how satisfaction attributes vary not only between companies but also across different segments within the same company. While this study uncovers relationships that diverge from established research findings, the underlying causal factors driving these relationships remain ambiguous. Consequently, future research should investigate the conditions under which factors typically detrimental to work-life balance may not adversely impact satisfaction. Additionally, it would be beneficial to explore the circumstances under which monetary incentives fail to significantly enhance satisfaction.

Author contributions

Conceptualization, R.L. and S.H.; data curation, Z.L.; formal analysis, Z.L. and K.D.; investigation, R.L. and S.H.; methodology, Z.L. and K.D.; resources, R.L. and S.H.; software, K.D.; validation, R.L.; visualization, Z.L.; writing—original draft, Z.L. and K.D.; writing—review and editing, Z.L., K.D., R.L., and S.H. All authors have read and agreed to the published version of the manuscript.

Funding

The authors received no specific funding for this work.

Availability of data and materials

The data used during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 May 2024 Accepted: 11 February 2025

Published online: 20 February 2025

References

1. Agrawal V. Managing the diversified team: challenges and strategies for improving performance. *Team Performance Management: An International Journal*. 2012;18:384–400.
2. Ahmed MH, Tiun S, Omar N, Sani NS. Short text clustering algorithms, application and challenges: A survey. *Appl Sci*. 2022;13(1):342.
3. Airoldi EM, Bischof JM. Improving and evaluating topic models and other models of text. *J Am Stat Assoc*. 2016;111(516):1381–403. <https://doi.org/10.1080/01621459.2015.1051182>.
4. Amabile TM, Kramer SJ. The power of small wins. *Harv Bus Rev*. 2011;89(5):70–80.
5. Amplayo RK, Lee S, Song M. Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Inf Sci*. 2018;454:200–15. <https://doi.org/10.1016/j.ins.2018.04.079>.
6. Azeem MU, Bajwa SU, Shahzad K, Aslam H. Psychological contract violation and turnover intention: The role of job dissatisfaction and work disengagement. *Employee Relations: Int J*. 2020;42(6):1291–308.
7. Baron JN, Hannan MT, Burton MD. Labor pains: change in organizational models and employee turnover in young, high-tech firms. *Am J Sociol*. 2001;106(4):960–1012.

8. Battisti E, Alfiero S, Leonidou E. Remote working and digital transformation during the COVID-19 pandemic: Economic–financial impacts and psychological drivers for employees. *J Bus Res.* 2022;150:38–50.
9. Belwal RC, Rai S, Gupta A. Text summarization using topic-based vector space model and semantic measure. *Inf Process Manage.* 2021;58(3): 102536.
10. Bezdrob M, Šunje A. Transient nature of the employees' job satisfaction: the case of the IT industry in Bosnia and Herzegovina. *Eur Res Manag Bus Econ.* 2021;27(2): 100141.
11. Bharathi SV, Mala EP. A study on the determinants of work–life balance of women employees in information technology companies in India. *Glob Bus Rev.* 2016;17(3):665–83. <https://doi.org/10.1177/0972150916630847>.
12. Bhende P, Mekoth N, Ingalthalli V, Reddy YV. Quality of work life and work–life balance. *J Human Values.* 2020;26(3):256–65. <https://doi.org/10.1177/0971685820939380>.
13. Bocean CG, Popescu L, Varzaru AA, Avram CD, Iancu A. Work-life balance and employee satisfaction during COVID-19 pandemic. *Sustainability.* 2023;15(15):11631.
14. Burnett JR, Lisk TC. The future of employee engagement: real-time monitoring and digital tools for engaging a workforce. In: *International perspectives on employee engagement*. England: Routledge; 2021.
15. Carraher-Wolverton C. The co-evolution of remote work and expectations in a COVID-19 world utilizing an expectation disconfirmation theory lens. *J Syst Inf Technol.* 2022;24(1):55–69.
16. Chang ML, Cheng CF. How balance theory explains high-tech professionals' solutions of enhancing job satisfaction. *J Bus Res.* 2014;67(9):2008–18. <https://doi.org/10.1016/j.jbusres.2013.10.010>.
17. Che Y, Zhu J, Huang H. How does employee–organization relationship affect work engagement and work well-being of knowledge-based employees? *Front Psychol.* 2022;13: 814324. <https://doi.org/10.3389/fpsyg.2022.814324>.
18. Cheng CF, Lai MK, Wu WY. Exploring the impact of innovation strategy on R&D employees' job satisfaction: a mathematical model and empirical research. *Technovation.* 2010;30(7–8):459–70.
19. Chien CF, Chen LF. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Syst Appl.* 2008;34(1):280–90. <https://doi.org/10.1016/j.eswa.2006.09.003>.
20. Chou HH, Fang SC, Yeh TK. The effects of facades of conformity on employee voice and job satisfaction: The mediating role of emotional exhaustion. *Manag Decis.* 2020;58(3):495–509. <https://doi.org/10.1108/MD-04-2019-0492>.
21. Colquitt JA, Conlon DE, Wesson MJ, Porter COLH, Ng KY. Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *J Appl Psychol.* 2001;86(3):425–45. <https://doi.org/10.1037/0021-9010.86.3.425>.
22. Costin A, Roman AF, Balica RS. Remote work burnout, professional job stress, and employee emotional exhaustion during the COVID-19 pandemic. *Front Psychol.* 2023;14:1193854.
23. Cui Y, Jiao H. Organizational justice and management trustworthiness during organizational Change: interactions of benevolence, integrity, and managerial approaches. *Inf Process Manage.* 2019;56(4):1526–42.
24. Ding K, Choo WC, Ng KY, Ng SI. Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation. *Int J Hosp Manag.* 2020;91:102676.
25. Ding K, Choo WC, Ng KY, Ng SI, Song P. Exploring sources of satisfaction and dissatisfaction in airbnb accommodation using unsupervised and supervised topic modeling. *Front Psychol.* 2021;12(April):1–19. <https://doi.org/10.3389/fpsyg.2021.659481>.
26. Ding K, Niu Y, Chong W. The evolution of airbnb research : a systematic literature review using structural topic modeling. *Heliyon.* 2023;9(6): e17090. <https://doi.org/10.1016/j.heliyon.2023.e17090>.
27. Dorner M, Fryges H, Schopen K. Wages in high-tech start-ups—Do academic spin-offs pay a wage premium? *Res Policy.* 2017;46(1):1–18.
28. Duhigg C. What google learned from its quest to build the perfect team. *NY Times Mag.* 2016;26(2016):2016.
29. Englehardt CS, Simmons PR. Organizational flexibility for a changing world. *Leadersh Org Dev J.* 2002;23(3):113–21.
30. Filbeck G, Zhao X, Warnaka M. Glassdoor's best places to work internationally: are they best for shareholders? *Int J Financ Econ.* 2022;27(4):4350–63. <https://doi.org/10.1002/ijfe.2375>.
31. Garg P, Rastogi R. New model of job design: motivating employees' performance. *J Manag Dev.* 2006;25(6):572–87. <https://doi.org/10.1108/02621710610670137>.
32. Getnet B, Jebena T, Tsegaye A. The effect of employees' fairness perception on their satisfaction towards the performance appraisal practices. *Int J Manag Commerce Innov.* 2014;2(1):174–210.
33. Gorgenyi-Hegyes E, Nathan RJ, Fekete-Farkas M. Workplace health promotion, employee well-being and loyalty during COVID-19 Pandemic—Large scale empirical evidence from Hungary. *Economies.* 2021;9(2):55.
34. Gupta N, Shaw JD. Employee compensation: The neglected area of HRM research. *Hum Resour Manag Rev.* 2014;24:1–4.
35. Hom PW, Lee TW, Shaw JD, Hausknecht JP. One hundred years of employee turnover theory and research. *J Appl Psychol.* 2017;102(3):530.
36. Hsu S-H, Wang Y-C. The development and empirical validation of the employee satisfaction index model. *Total Qual Manag Bus Excell.* 2008;19(4):353–66. <https://doi.org/10.1080/14783360701595052>.
37. Hvitfeldt E, Silge J. *Supervised machine learning for text analysis* in R. Boca Raton: Chapman and Hall/CRC; 2021.
38. Jackson LT, Fransman EI. Flexi work, financial well-being, work–life balance and their effects on subjective experiences of productivity and job satisfaction of females in an institution of higher learning. *South Afr J Econ Manag Sci.* 2018;21(1):1–13. <https://doi.org/10.4102/sajems.v21i1.1487>.
39. Jacobs T, Tschötschel R. Topic models meet discourse analysis: a quantitative tool for a qualitative approach. *Int J Soc Res Methodol.* 2019;22(5):469–85. <https://doi.org/10.1080/13645579.2019.1576317>.
40. Jehanzeb K, Bashir NA. Training and development program and its benefits to employee and organization: a conceptual study. *Eur J bus manag.* 2013;5(2):2222–839.
41. Joshi A, Sekar S, Das S. Decoding employee experiences during pandemic through online employee reviews: insights to organizations. *Pers Rev.* 2024;53(1):288–313.
42. Judge TA, Weiss HM, Kammeyer-Mueller JD, Hulin CL. Job attitudes, job satisfaction, and job affect: a century of continuity and of change. *J Appl Psychol.* 2017;102(3):356–74. <https://doi.org/10.1037/apl0000181>.

43. Judge TA, Zhang SC, Glerum DR. Job satisfaction. In: Essentials of job attitudes and other workplace psychological constructs. England: Routledge; 2020.
44. Jung Y, Suh Y. Mining the voice of employees: a text mining approach to identifying and analyzing job satisfaction factors from online employee reviews. *Decis Support Syst*. 2019;123: 113074.
45. Kam CCS, Meyer JP. How careless responding and acquiescence response bias can influence construct dimensionality: the case of job satisfaction. *Organ Res Methods*. 2015;18(3):512–41. <https://doi.org/10.1177/1094428115571894>.
46. Kantén P. Family friendly policies in organizations and their effects on work-life balance, work alienation and life satisfaction. *Int J Bus Technol*. 2014;2(2):18–30.
47. Kumar S, Zymbler M. A machine learning approach to analyze customer satisfaction from airline tweets. *J Big Data*. 2019;6(1):1–16.
48. Lee M, Kim B. Effect of employee experience on organizational commitment: case of South Korea. *Behav Sci*. 2023;13(7):521.
49. Lee Y, Ng S, Shevlin T, Venkat A. The effects of tax avoidance news on employee perceptions of managers and firms evidence from glassdoor.com ratings. *Account Rev*. 2021;96(3):343–72.
50. Lee Y, Tao W, Li JYQ, Sun R. Enhancing employees' knowledge sharing through diversity-oriented leadership and strategic internal communication during the COVID-19 outbreak. *J Knowl Manag*. 2021;25(6):1526–49.
51. Ling Y, Cai F, Hu X, Liu J, Chen W, Chen H. Context-controlled topic-aware neural response generation for open-domain dialog systems. *Inf Process Manage*. 2021;58(1): 102392.
52. Liu HY. The role of the state in influencing work conditions in China's internet industry: Policy, evidence, and implications for industrial relations. *J Ind Relat*. 2023;65(1):3–21.
53. Liu X, Lu J, Filatotchev I, Buck T, Wright M. Returnee entrepreneurs, knowledge spillovers and innovation in high-tech firms in emerging economies. *J Int Bus Stud*. 2010;41:1183–97.
54. London M, Diamante T. Technology-focused expansive professionals: developing continuous learning in the high-technology sector. *Hum Resour Dev Rev*. 2002;1(4):500–24. <https://doi.org/10.1016/j.ipm.2020.102392>.
55. McNall LA, Masuda AD, Nicklin JM. Flexible work arrangements, job satisfaction, and turnover intentions: The mediating role of work-to-family enrichment. *J Psychol*. 2009;144(1):61–81.
56. Men LR, Qin YS, Jin J. Fostering employee trust via effective supervisory communication during the COVID-19 pandemic: Through the lens of motivating language theory. *Int J Bus Commun*. 2022;59(2):193–218.
57. Mihalache M, Mihalache OR. How workplace support for the COVID-19 pandemic and personality traits affect changes in employees' affective commitment to the organization and job-related well-being. *Hum Resour Manage*. 2022;61(3):295–314.
58. Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. Optimizing semantic coherence in topic models. In Proceedings of the 2011 conference on empirical methods in natural language processing; 2011. pp. 262–272.
59. Moro S, Ramos RF, Rita P. What drives job satisfaction in IT companies? *Int J Product Perform Manag*. 2021;70(2):391–407. <https://doi.org/10.1108/IJPPM-03-2019-0124>.
60. Nair MS, Salleh R. Linking trust and turnover intention: a conceptual framework. *Global Bus Manag Res*. 2017;9(1s):281–8.
61. O'Connor E, Crowley-Henry M. Exploring the relationship between exclusive talent management, perceived organizational justice and employee engagement: Bridging the literature. *J Bus Ethics*. 2019;156:903–17.
62. Pandey P, Singh S, Pathak P. An exploratory study on factors contributing to job dissatisfaction of retail employees in India. *J Retail Consum Serv*. 2021;61: 102571.
63. Pastrana B, Tobón S. Emerging needs of human talent training in leading information technology companies, a socio-formative analysis. *World Rev Sci Technol Sustain Dev*. 2020;16(4):303–19.
64. Prottas DJ. Relationships among employee perception of their manager's behavioral integrity, moral distress, and employee attitudes and well-being. *J Bus Ethics*. 2013;113:51–60.
65. Razali NAM, Malizan NA, Hasbullah NA, Wook M, Zainuddin NM, Ishak KK, Sukardi S. Opinion mining for national security: techniques, domain applications, challenges and research opportunities. *J Big Data*. 2021;8:1–46.
66. Roberts ME, Stewart BM, Airolidi EM. A model of text for experimentation in the social sciences. *J Am Stat Assoc*. 2016;111(515):988–1003.
67. Roberts ME, Stewart BM, Tingley D. stm: an R package for structural topic models. *J Stat Softw*. 2019;91:1–40.
68. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural topic models for open-ended survey responses. *Am J Polit Sci*. 2014;58(4):1064–82.
69. Saini GK, Jawahar IM. The influence of employer rankings, employment experience, and employee characteristics on employer branding as an employer of choice. *Career Dev Int*. 2019;24(7):636–57. <https://doi.org/10.1108/CDI-11-2018-0290>.
70. Sainju B, Hartwell C, Edwards J. Job satisfaction and employee turnover determinants in Fortune 50 companies: insights from employee reviews from Indeed Com. *Decis Support Syst*. 2021;148:113582.
71. Sanja M, Eftimov L. Calculating the cost for employee turnover in the IT industry in Macedonia by using a web calculator. *J Human Res Manag*. 2016;1(24):23.
72. Santacreu-Vasut E, Teshima K. Foreign employees as channel for technology transfer: Evidence from MNC's subsidiaries in Mexico. *J Dev Econ*. 2016;122:92–112.
73. Schneckenberg D. Web 2.0 and the empowerment of the knowledge worker. *J knowl Manag*. 2009;13(6):509–20.
74. Sekar S, Edakkat Subhakaran S, Chattopadhyay D. Unlocking the voice of employee perspectives: exploring the relevance of online platform reviews on organizational perceptions. *Manag Decis*. 2023;61(11):3408–29.
75. Shan S, Li C, Yao W, Shi J, Ren J. An empirical study on critical factors affecting employee satisfaction. *Syst Res Behav Sci*. 2014;31(3):447–60.
76. Shirmohammadi M, Au WC, Beigi M. Remote work and work-life balance: Lessons learned from the COVID-19 pandemic and suggestions for HRD practitioners. *Hum Resour Dev Int*. 2022;25(2):163–81.
77. Stamolampros P, Korfiatis N, Chalvatzis K, Buhalis D. Job satisfaction and employee turnover determinants in high contact services: insights from employees' online reviews. *Tour Manage*. 2019;75:130–47.

78. Sutherland I, Sim Y, Lee SK, Byun J, Kiatkawsin K. Topic modeling of online accommodation reviews via latent dirichlet allocation. *Sustainability*. 2020;12(5):1821.
79. Symitsi E, Stamolampros P, Daskalakis G, Korfiatis N. The informational value of employee online reviews. *Eur J Oper Res*. 2021;288(2):605–19.
80. Ton Z, Huckman RS. Managing the impact of employee turnover on performance: the role of process conformance. *Organ Sci*. 2008;19(1):56–68. <https://doi.org/10.1287/orsc.1070.0294>.
81. Torre C, Tommasetti A, Maione G. Technology usage, intellectual capital, firm performance and employee satisfaction: the accountants' idea. *TQM J*. 2021;33(3):545–67. <https://doi.org/10.1080/13645579.2016.1185255>.
82. Toscano F, Zappalà S. Social isolation and stress as predictors of productivity perception and remote work satisfaction during the COVID-19 pandemic: The role of concern about the virus in a moderated double mediation. *Sustainability*. 2020;12(23):9804.
83. Van Mol C. Improving web survey efficiency: the impact of an extra reminder and reminder content on web survey response. *Int J Soc Res Methodol*. 2017;20(4):317–27.
84. Wang, C., & Blei, D. M. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2021; pp. 448–456
85. Williams ML, Brower HH, Ford LR, Williams LJ, Carraher SM. A comprehensive model and measure of compensation satisfaction. *J Occup Organ Psychol*. 2008;81(4):639–68.
86. Wu WT, Li YJ, Feng AZ, Li L, Huang T, Xu AD, Lyu J. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res*. 2021;8:1–12.
87. Xia H, Muskat B, Vu HQ, Law R, Li G. Leveraging employee online reviews for improving hotel competitiveness in the great resignation. *Int J Hosp Manag*. 2023;113: 103529.
88. Xiang Z, Schwartz Z, Gerdes JH Jr, Uysal M. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int J Hosp Manag*. 2015;44:120–30.
89. Xu X. Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants. *J Bus Res*. 2020;106:196–210. <https://doi.org/10.1016/j.jbusres.2018.07.022>.
90. Yee RW, Yeung AC, Cheng TE. The impact of employee satisfaction on quality and profitability in high-contact service industries. *J Oper Manag*. 2008;26(5):651–68. <https://doi.org/10.1016/j.jom.2008.01.001>.
91. Yoo M, Lee S, Ha T. Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit. *Inf Process Manage*. 2019;56(4):1565–75.
92. Zhang LE, Zhao S, Kern P, Edwards T, Zhang ZX. The pursuit of indigenous innovation amid the tech cold war: the case of a Chinese high-tech firm. *Int Bus Rev*. 2023;32(6): 102079.
93. Zivnуска SL, Carlson DS, Carlson JR, Harris KJ, Harris RB, Valle M. Information and communication technology incivility aggression in the workplace: implications for work and family. *Inf Process Manage*. 2020;57(3): 102222.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.