Kontigensi: Jurnal Ilmiah Manajemen
Vol 11 , No. 2 December 2023, pp. 409-417
ISSN 2088-4877

Implementation of Sentiment Analysis in HR Management and Development Planning Using Topic Modelling on Employee Review

Furqon Sandiva Utomo Putra, Ira Puspitasari
Universitas Airlangga, Surabaya, Indonesia
E-mail: furqonsand@gmail.com

DOI: https://doi.org/10.56457/jimk.v11i2.358

Received: May 2023	Accepted: November 2023	Published: December 2023

ABSTRACT

Employee satisfaction is pivotal in influencing workplace motivation, performance, and well-being, with dissatisfaction potentially leading to high turnover and organizational disruption. Traditional methods of gauging employee satisfaction often face challenges due to hesitancy in candid expression. This study addresses this gap by leveraging sentiment analysis and topic modelling to analyze 9,000 anonymous employee reviews from Accenture on Glassdoor. This approach enables the extraction of nuanced insights into employee perceptions, categorizing them into prevalent themes. Our analysis utilized advanced data analytics tools to dissect both positive and negative sentiments, revealing key areas of employee satisfaction and dissatisfaction. The findings were then subjected to opportunity mapping, pinpointing specific domains for organizational improvement. This study not only sheds light on the actual sentiments of employees but also provides a strategic framework for organizations to prioritize areas for enhancement in HR management and development planning. The methodology and insights from this research offer a novel approach for organizations to harness employee feedback for constructive development.

Keywords: Employee Satisfaction, Data Analytics, Opportunity Mapping

INTRODUCTION

Employee satisfaction is a job indicator that reflects employees' feelings and views on various aspects and experiences of their work (Astuti & Iverizkinawati, 2018). Modern organizations consider this a crucial indicator in the workplace (Lee, 2019), because it is one of the key aspects influencing an individual's motivation, performance, and mental well-being (Saidah, 2022). Research has found that employee satisfaction significantly impacts service quality and customer satisfaction, thereby enhancing company profits (Alnaser et al., 2018). Thus, employees who are satisfied with their jobs tend to be more effective and productive (Okpo et al., 2021).

On the other hand, low employee satisfaction is a strong predictor of employee turnover (Stamolampros et al., 2019). This becomes a problem because companies incur significant costs in replacing and training their departing employees. Additionally, employee turnover can disrupt organizational effectiveness and performance as the workload is shifted to remaining employees, reducing productivity and morale (Verma & Kesari, 2020). On the other hand, high satisfaction increases employee commitment to the organization (Mahmood Aziz et al., 2021). Therefore, companies and organizations must enhance employee satisfaction to maintain productivity, performance, and retention rate. This can be achieved by planning human resource (HR) management and development strategies that aim to uphold employee satisfaction.

To do so, managers need to understand the aspects that make employees happy or unhappy with their respective companies. This information can be used to develop improvement plans for enhancing employee satisfaction. However, obtaining this information can be difficult as it needs to be factual, accurate, and unbiased. This becomes a challenge because not all employees openly express their opinions due to fear of consequences. Furthermore, acquiring data through surveys and questionnaires also possesses weaknesses, in that factors such as lengthy questionnaires, cognitive burden, and question density can impose fatigue and respondent burden (Abay et al., 2021), leading to fewer, incomplete, and lower-quality responses (Marshall et al., 2019).

To solve this problem, companies can utilize third-party platforms, such as career websites, where employees can leave anonymous reviews about their workplaces. However, for large companies with hundreds or thousands of employees, an overwhelming amount of textual data must be interpreted. In addition, management needs to identify which aspects within their organization need prioritization for improvement to enhance employee satisfaction. To address this issue, this study implements data analytics approaches using sentiment analysis, topic modelling, and opportunity mapping.

Sentiment analysis is a data analytics approach belonging to the family of data mining and text mining, which is the process of exploring and analyzing large datasets to discover meaningful patterns, rules, and correlations (De et al., 2022). Sentiment analysis enables researchers to classify text based on the emotions expressed in the sentences by categorizing individual words as positive, negative, or neutral in context (Tonkin, 2016). On the other hand, topic modelling is an unsupervised machine learning approach used to identify words and discover patterns of association within representative clusters (Gupta et al., 2022). Using this method, researchers can derive keywords from textual datasets. The combination of sentiment analysis and topic modelling in the research context provides information on the aspects expressed by employees in their satisfaction reviews and whether these aspects tend to be evaluated negatively or positively.

Previous studies have employed various measurement methods for employee satisfaction, such as OLS regression models, correlation analysis, Principal Component Analysis, and Confirmatory Factor Analysis (Abdelwahed, et al., 2022). This study, however, utilizes sentiment analysis with topic modelling to identify the aspects of employee satisfaction reviews and measure the level of satisfaction for each aspect through sentiment analysis. Previous studies have also examined factors influencing employee satisfaction through sentiment analysis and regression (Chittiprolu et al., 2021). However, this study expands its scope by incorporating opportunity mapping.

Opportunity mapping is conducted to help management identify which aspects, among those extracted through text mining, need to be prioritized for improvement. Although previous research has combined opportunity mapping with text mining, these studies focused on product planning, whereas this study extends the context to management and HR development planning (Irawan, et al., 2020). This is achieved by viewing the HR management and development process as a product, with employees as customers who need to be served by management. The data utilized in this study comprise reviews left by employees and former employees of Accenture, which are publicly available on the Glassdoor website.

LITERATURE REVIEW

1. EMPLOYEE SATISFACTION

Employee job satisfaction refers to employees' feelings regarding their work efforts. These feelings can be positive (satisfied) or negative (disappointed) (Pawoko & others, 2019; Setia et al., 2022). It reflects how employees perceive the tasks and responsibilities assigned by the company. The level of employee satisfaction influences how employees view their organizations (Eyo et al., 2022; Nabahani & Riyanto, 2020). Satisfaction can motivate employees to work harder, and dissatisfaction can be a reason for leaving the job.

Employee satisfaction elements can be divided into four categories. The first is job satisfaction, which includes work conditions, working hours, and company reputation. The second is employee relationships, which involve relationships among co-workers. The third is compensation, benefits, and organizational culture, which includes salary, bonuses, promotions, development opportunities, and a supportive culture. Fourth is employee loyalty to the company, as satisfied employees tend to be loyal to the organization. On the other hand, employees' decisions to leave the company can be influenced by their job satisfaction (Abdullah, et al., 2013). Managers need to monitor the factors that affect job satisfaction.

2. TEXT MINING

Text mining is an approach used to transform text to derive meaning and find solutions to problems. Patterns and trends are discovered in unstructured data in order to obtain high-quality information. This is done to predict whether a line, sentence, paragraph, or document belongs to a particular category based on similar patterns. Unstructured text information is transformed into vectors or numbers, which can then be processed using statistical methods, algorithms, and data mining techniques. Typically, this is accomplished by applying Statistical Natural Language Processing (SNLP) methods.

The general steps involved in text mining begin with preprocessing. This is a method for converting raw data into a more suitable form for analysis (Gupta et al., 2022). Then, the analysis is carried out using various techniques such as information extraction, clustering, and summarization. First, information extraction involves extracting specific information, structures, and attribute relationships from text data. This can be achieved by creating pattern matching rules to identify specific relations or by using a supervised machine learning approach to train a model with specific patterns. Second, clustering is used to classify documents into relatively homogeneous groups. These groups or clusters contain a set of objects that have certain similarities or proximities (Februariyanti & Santoso, 2017). This method is employed when the grouping of the dataset is unknown (Yudiarta, et al., 2018). Lastly, summarization is a technique for condensing long text data. The goal is to create a clear and concise summary that captures the essence of the text.

3. Sentiment Analysis and Topic Modelling

Sentiment analysis is a text mining method that utilizes natural language processing (NLP) and machine learning algorithms. In this method, a program is trained to detect whether a message has a positive, negative, or neutral context. This is accomplished by breaking down the message into various topics, each assigned a sentiment value (Hauthal, et al., 2020).

Topic modelling is a text mining method used to extract topics (clusters of words) from large textual datasets to discover hidden semantic structures within the data (Likhitha et al., 2019). It is an unsupervised approach for uncovering hidden structures in text documents (Vayansky & Kumar, 2020). This means that topic modelling does not rely on tags, training data, or predefined taxonomies (Günther & Quandt, 2018). Instead, modelling is based on the frequency and co-occurrence of words in one or more documents, clustering frequently co-occurring words together (Salah et al., 2018). Topic modelling is used to gain insights into the specific themes possessed by a group of documents (Abdelrazek, et al., 2022). This analysis can help organizations with large amounts of data gain easier access by categorizing their data into specific groups (Vayansky & Kumar, 2020).

4. Opportunity Mapping

Opportunity mapping is a concept within the Opportunity-Driven Innovation (ODI) framework introduced by Anthony W. Ulwick. It is a method for estimating market opportunity and determining which efforts should be prioritized in product development (Ulwick, 2016). This method provides decision-makers with insights into what needs to be done to satisfy customers.

Opportunity mapping has two axes: importance and satisfaction. Importance represents the value customers attribute to the importance of a product quality aspect, while satisfaction indicates how satisfied they are with the current quality of that aspect. Mapping is performed with the X-axis representing importance and the Y-axis representing satisfaction. A line is drawn from the origin (0,0) to point (10,10) where X represents the average importance and Y represents the average satisfaction. Then, a line is drawn from the maximum point (10) to the average importance and average satisfaction. Topics mapped above the satisfaction line are considered overserved, topics within the area between the importance and satisfaction lines are considered served right, and topics mapped below the underserved line are considered underserved.

Topics that fall into the underserved area indicate aspects that are less satisfying relative to their importance. Over-served topics mean that customers are already highly satisfied with those aspects relative to their importance. Served right topics indicate a proportional balance between satisfaction and importance for those aspects.

METHOD

To test the model, sample data of employee reviews from Accenture is used. A dataset of 9000 publicly available employee review of the company from the Glassdoor website has been scrapped (Raut, 2023). Each of the data from the website consists of two reviews (pros and cons working in the company), resulting in 18000 distinct review data. The data analysis model used in this study is illustrated in Figure 1.

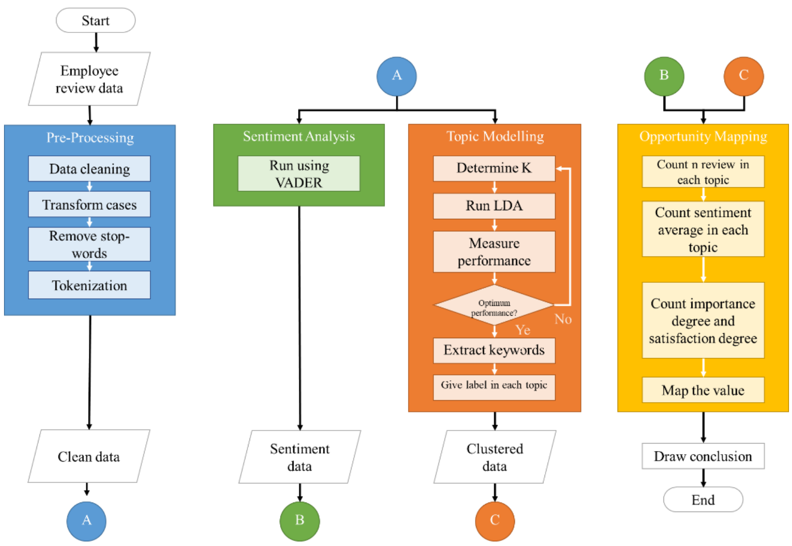

Figure 1. Data Analysis Model

After the data are obtained, pre-processing is performed to prepare the data for processing. This process involves data cleaning, transform cases, stopword removal, and tokenization. Data cleaning is the process to identify error in the data (Ridzuan, et al., 2019). After all errors have been corrected or eliminated, transform cases are performed to convert all text data into lowercase letters. Then, stopwords are removed, which are words that do not have semantic importance, such as conjunctions, that are not useful for analysis (Rania & Lobiyal, 2018). Finally, tokenization is performed. This is the process by which sentences are broken down into smaller words for analysis (Lamurias & Couto, 2019).

In this study, the VADER and lexicons were used for sentiment analysis. It is a baseline-based approach to sentiment analysis, where the lexicons have predefined lists of words associated with negative and positive contexts. VADER is a lexicon created based on words and terms commonly used in social media (Kandasamy, et al., 2020). The input for this process is clean data, and the output is the polarity score for each text data.

For topic modelling, the Latent Dirichlet Allocation (LDA) algorithm was employed in this research. This method clusters textual data into k predetermined topics based on Probabilistic Latent Semantic Analysis (PLSI) (Hofmann, 1999). The Jensen–Shannon divergence (JSD) value was computed to determine the optimal number of topics (k). JSD is a measure of dissimilarity between probability distributions and is commonly used in topic modelling to evaluate model performance (Osán, et al., 2022). Lower JSD values indicate that the topic model captures the underlying themes of a corpus (a collection of words in documents) better and produces more coherent and meaningful topics. Hence, topic modelling is iterated with different values of k, and the k with the lowest JSD value is selected.

The results from sentiment analysis and topic modelling are then utilized for opportunity mapping. In the context of this study, Importance is measured based on the number of reviews in a topic, while Satisfaction is measured through the average sentiment polarity of the reviews within a topic (Irawan, et al., 2020). These calculations are then normalized within the interval (0,10) using the formula:

$$\boldsymbol{importance_i = 10 \times \frac{CS_i - CS_{min}}{CS_{max} - CS_{min}}} \quad \textbf{(1)}$$

$$\boldsymbol{satisfaction_i = 10 \times \frac{SS_i - SS_{min}}{SS_{max} - SS_{min}}} \quad \textbf{(2)}$$

Where:

- $Importance_i$: importance score for topic i
- CS_i: number of reviews in topic i
- CS_{min}: number of the fewest review among all topics
- CS_{max}: number of the most review among all topics
- $Satisfaction_i$: satisfaction score in topic i
- SS_i: average review sentiment polarity in topic i
- SS_{min}: the average review sentiment polarity that has the least value among all topics
- SS_{max}: the average review sentiment polarity that has the most value among all topics

RESULT and DICUSSION

The sentiment analysis step returned the polarity score of each of the rows of text data. In the topic modelling, it is found that the optimal k used is 4. The JSD score for each iteration of k can be seen in Figure 2. Keywords are then extracted from each topic and used to label each of these topics. The keywords can be seen in Table 1.

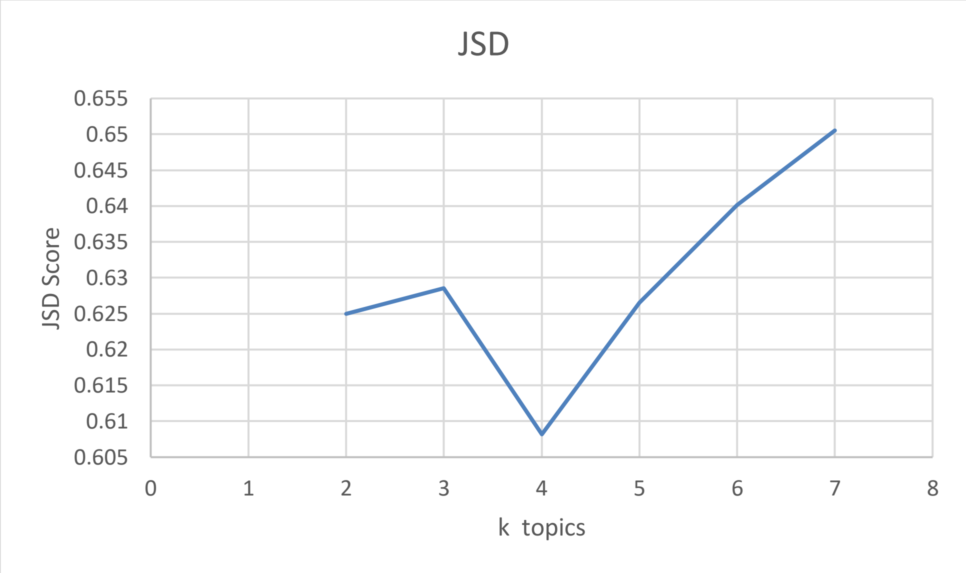

Figure 2. JSD score for different k

Table 1. Keywords for each topics

Topic 0		Topic 1		Topic 2		Topic 3	
0.1543	work	0.0378	con	0.0704	NUM	0.056	project
0.1518	good	0.0336	manag	0.0462	work	0.0528	good
0.0394	balanc	0.0164	compani	0.0364	salari	0.051	learn
0.0372	cultur	0.0153	accentur	0.0316	less	0.038	opportun
0.0367	life	0.0145	project	0.0298	hike	0.0323	technolog
0.0352	compani	0.0143	employe	0.024	pay	0.0229	new
0.0349	environ	0.012	peopl	0.0206	time	0.0207	get
0.0259	great	0.0116	polit	0.0202	hour	0.0196	work
0.0188	place	0.0102	NUM	0.0171	project	0.017	lot
0.0159	employe	0.0101	none	0.0164	manag	0.0161	skill

Looking at the keywords and sample comments in each topic, it is concluded that topic 0 is labelled as work-life balance, topic 1 as aspects of company management, topic 2 as incentives, and topic 3 as learning opportunities in companies. Then, this information is used for opportunity mapping. Calculation using Formula 1 and 2 results in Table 2. From there, an opportunity landscape map is plotted as shown in Figure 3. From the figure, it can be concluded that topics 0, 1, and 2 are served right, while topic 3 is overserved.

Table 2. Opportunity mapping calculation result

Topic no	Number of Comments	Polarity Avg	Importance Degree	Satisfaction Degree
0	7179	0.43	10.00	10.00
1	3763	0.05	1.69	0.00
2	3989	0.06	2.24	0.40
3	3069	0.28	0.00	6.15
Max	7179	0.43		
Min	3069	0.04		

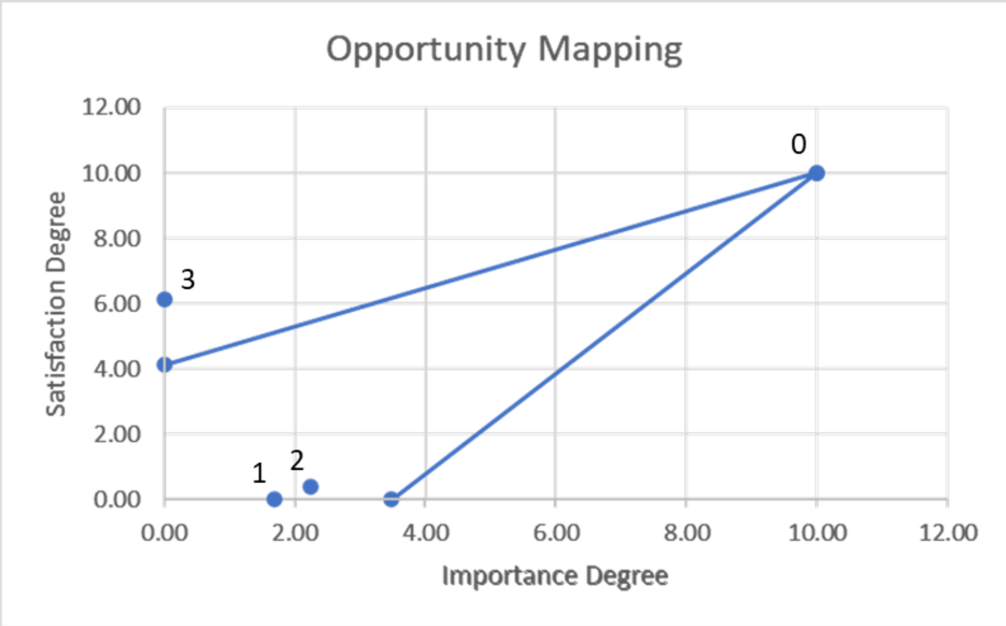

Figure 3. Opportunity landscape map

The fact that no topics fall into the under-served category indicates that employees, on average, feel satisfied with the company's quality of work. Moreover, topic 3 being over-served suggests that the company's investments in those areas have paid off, and employees receive highly satisfactory training and development. This is a positive finding as it demonstrates that the company invests in the future of its employees and assists in developing their skills, which can enhance employee satisfaction and retention.

However, the "served right" position for topics 0, 1, and 2 highlights that, although these aspects of the company are adequately addressed, there is still room for improvement. Organize training for managers to enhance their leadership skills and learn effective communication strategies. Research has shown that leadership training can significantly improve employees' perceptions of leaders in terms of alignment, communication, and integrity (Awad, et al., 2004). Studies have also indicated a positive relationship between communication quality within an organization and employee satisfaction (Madalina & Catalin, 2016).

Regarding incentive complaints, the company needs well-designed and implemented incentive schemes, as they can have a significant positive impact on performance (Brown & McIntosh, 2003). Research has demonstrated that financial incentives and rewards can motivate employees to have a positive attitude toward the workplace, leading to job satisfaction and increased productivity. Therefore, employers should pay attention to these aspects (Ekpudu & Ojeifo, 2014).

CONCLUSION

Based on the research findings, it is evident that information regarding employee satisfaction can be obtained through reviews on career websites using sentiment analysis and topic modelling. Sentiment analysis provides insights into the positive or negative perceptions of employees towards the company based on their reviews, while topic modelling reveals the aspects they comment on. In the case of Accenture, the most dominant topics identified were work-life balance, management, incentives, and learning opportunities.

Furthermore, the opportunity mapping approach can be utilized to measure and map the aspects that need improvement to enhance employee satisfaction by identifying what is already satisfactory, lacking, and highly satisfying for employees. In the case of Accenture, based on the results of opportunity mapping, the aspects of work-life balance, management, and incentives were found to be quite satisfactory, while learning opportunities were highly satisfying.

Subsequent research can use other categorical variables (e.g., department, gender, and age) to map the results of sentiment analysis and topic modelling, to see the impact of these aspects on employee satisfaction.

REFERENCES

Abay, K. A., Berhane, G., Hoddinott, J. F., & Tafere, K. (2021). *Assessing response fatigue in phone surveys: Experimental evidence on dietary diversity in Ethiopia* (Vol. 2017). Intl Food Policy Res Inst.

Alnaser, F., Ghani, M., & Rahi, S. (2018). Service quality in Islamic banks: The role of PAKSERV model, customer satisfaction and customer loyalty. *Accounting*, *4*(2), 63–72.

Astuti, R., & Iverizkinawati, I. (2018). Pengaruh Kepemimpinan dan Lingkungan Kerja terhadap Kepuasan Kerja Karyawan pada PT. Sarana Agro Nusantara Medan. *Jurnal Ilman: Jurnal Ilmu Manajemen*, *6*(1), 26–41.

Chittiprolu, V., Singh, S., Bellamkonda, R. S., & Vanka, S. (2021). A text mining analysis of online reviews of Indian hotel employees. *Anatolia*, *32*(2), 232–245.

De, S., Dey, S., Bhatia, S., & Bhattacharyya, S. (2022). An introduction to data mining in social networks. In *Advanced Data Mining Tools and Methods for Social Computing* (pp. 1–25). Elsevier.

Eyo, E. I., Sunarsi, D., Herling, H., Nurjaya, N., & Affandi, A. (2022). A Literatur Review in Learning Organization. *International Journal of Educational Administration, Management, and Leadership*, 11–16.

Februariyanti, H., & Santoso, D. B. (2017). *Hierarchical agglomerative clustering untuk pengelompokan skripsi mahasiswa*.

Günther, E., & Quandt, T. (2018). Word counts and topic models: Automated text analysis methods for digital journalism research. In *Rethinking research methods in an age of digital journalism* (pp. 75–88). Routledge.

Gupta, R. K., Agarwalla, R., Naik, B. H., Evuri, J. R., Thapa, A., & Singh, T. D. (2022). Prediction of research trends using LDA based topic modeling. *Global Transitions Proceedings*, *3*(1), 298–304.

Lee, Y. (2019). Workplace health and its impact on human capital: seven key performance indicators of workplace health. *Indoor Environment and Health*, *43*.

Likhitha, S., Harish, B. S., & Kumar, H. M. K. (2019). A detailed survey on topic modeling for document and short text data. *International Journal of Computer Applications*, *178*(39), 1–9.

Mahmood Aziz, H., Jabbar Othman, B., Gardi, B., Ali Ahmed, S., Sabir, B. Y., Burhan Ismael, N., Abdalla Hamza, P., Sorguli, S., Ali, B. J., & Anwar, G. (2021). Employee commitment: The relationship between employee commitment and job satisfaction. *Aziz, HM, Othman, BJ, Gardi, B., Ahmed, SA, Sabir, BY, Ismael, NB, Hamza, PA, Sorguli, S., Ali, BJ, Anwar, G.(2021). Employee Commitment: The Relationship between Employee Commitment And Job Satisfaction. Journal of Humanities and Education Development*, *3*(3), 54–66.

Marshall, N. T., Schroeder, J. H., & Yohn, T. L. (2019). An incomplete audit at the earnings announcement: Implications for financial reporting quality and the market's response to earnings. *Contemporary Accounting Research*, *36*(4), 2035–2068.

Nabahani, P. R., & Riyanto, S. (2020). Job satisfaction and work motivation in enhancing generation Z�s organizational commitment. *Journal of Social Science*, *1*(5), 234–240.

Okpo, E.-I. E., Nwoyi, O. E., Ogar, G. W., Arrey, M.-E. V., & Nkamare, S. E. (2021). The Impact of Job Satisfaction on Employees Productivity. *IIARD International Journal of Economics and Business Managemen*, *7*(1), 55–62.

Pawoko, G., & others. (2019). The role of work-life balance, achievement motivation on organizational commitment through satisfaction athlete's in Indonesia. *Academy of Strategic Management Journal*, *18*(5), 1–11.

Saidah, M. (2022). Pengaruh Pendidikan, Kompetensi, Disiplin dan Kesejahteraan Karyawan Tehadap Kinerja Karyawan Supermarket Sakinah Surabaya. *MALIA: Jurnal Ekonomi Islam*, *13*(2), 261–274.

Salah, A., Ailem, M., & Nadif, M. (2018). Word co-occurrence regularized non-negative matrix tri-factorization for text data co-clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).

Setia, B. I., Yuniarsih, T., Gaffar, M. F., Suryadi, E., Affandi, A., & Rohmawati, T. (2022). Job satisfaction as a mediator in improving employee performance through talent and knowledge management. *Journal of Eastern European and Central Asian Research (JEECAR)*, *9*(5), 749–762.

Stamolampros, P., Korfiatis, N., Chalvatzis, K., & Buhalis, D. (2019). Job satisfaction and employee turnover determinants in high contact services: Insights from Employees� Online reviews. *Tourism Management*, *75*, 130–147.

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582.

Verma, B. K., & Kesari, B. (2020). Does the morale impact on employee turnover intention? An empirical investigation in the Indian steel industry. *Global Business Review*, *21*(6), 1466–1488.