 ISSN: 2473-4901
v9 n5915

Finding Factors Influencing IT Industry Job Satisfaction Through Topic Modeling

Muge Kosar
mkosar1@student.gsu.edu
Andrew Young School of Policy Studies

Frank Lee
flee@gsu.edu
J. Mack Robinson College of Business

Georgia State University
Atlanta, Georgia, US

Abstract

This study investigates factors of job satisfaction in the IT industry. Using the Latent Dirichlet Allocation (LDA) machine learning technique, over 5,000 employee reviews from 17 leading digital companies are analyzed. The analysis reveals nine key topics IT employees value: management skills and responsibilities, workplace culture and environment, and job experience and flexibility. The comparison of LDA topics with those of a human rater shows moderate overlap with topics identified by independent researchers in work culture and environment, management, career advancement, work-life balance, and compensation and benefits. The sentiment analysis reveals that most reviews are positive.

Keywords: IT Industry, Job Satisfaction, Latent Dirichlet Allocation, Employee Reviews, Sentiment Analysis

1. INTRODUCTION

The significance of employee satisfaction in organizations cannot be overstated, as it directly impacts a company's success and competitiveness (Oshagbemi, 2003). Given the growing significance of Information and Communication Technologies (ICT) in today's environment, it has become crucial for organizations to understand the factors that contribute to employee satisfaction and align them with their corporate strategy (Holland & Bardoel, 2016). Effective human resource management is critical in motivating employees and reducing turnover (Sainju et al., 2021).

Job satisfaction studies have explored a wide range of factors that impact job satisfaction, including employee motivation, corporate performance, absenteeism, turnover, and the financial performance of companies. Personal attributes and organizational factors such as age, gender, level of education, company size, and industry have also been examined concerning job satisfaction.

This study uses topic modeling algorithms to determine factors in IT industry employee satisfaction and utilizes a multi-step process for text analysis. The first step involves extensive text pre-processing, including removing stop words, lemmatization, extracting nouns, and generating bigrams. The second step employs Latent Dirichlet Allocation (LDA) topic modeling to identify critical topics from a dataset of 5,000 employee reviews from 17 leading digital companies. The third step involves comparison with human rater results to ensure the accuracy and consistency of the analysis. Finally, sentiment analysis is conducted to classify each

 ISSN: 2473-4901
v9 n5915

review's tone, providing insight into employees' positive and negative sentiments. Through this comprehensive process, the study provides valuable insights for human resource management, helping them to improve employee satisfaction and, thus, increase the company's competitiveness.

2. LITERATURE REVIEW

Employee satisfaction is an examination of existing research on how satisfied employees are with their current employment in their field. Some studies identify essential employee satisfaction factors, such as job satisfaction, work-life balance, organizational culture, and employee retention. Additionally, other studies examine the impact of these factors on employee engagement, productivity, and turnover rate. Appendix A, Table 1 summarizes the literature review.

This study examines job satisfaction factors among IT industry employees through the topic modeling algorithm LDA. Text mining has been identified as a practical approach for analyzing employee reviews and identifying job satisfaction factors. Studies such as Jung & Suh (2019) and Luo et al. (2016) have found that text mining can extract information from employee reviews that may be difficult to identify through other methods.

Trends in IT Industry

Since the pandemic in the year 2020, there has been a significant rise in layoffs. Barnett and Li (2023) estimate that there has been an increase in tech layoffs since the pandemic began in 2020. The estimates consist of prominent corporations such as Meta Platforms, the parent company of Facebook, and Amazon, along with smaller enterprises within the United States and abroad (Barnett & Li, 2023). Deagon (2023) reported that this trend is due to companies hiring excessively during the pandemic and laying off workers due to decreased demand for tech products. Furthermore, according to a report by Challenger Gray and Christmas Inc. (2023), the number of job layoffs reported in February 2023 was the most for the month since 2009, indicating the trend of layoffs in the IT business is still ongoing.

Layoffs in the IT sector are expected to affect employee satisfaction and well-being, emphasizing the importance of effective employee support strategies. Businesses may improve employee retention and satisfaction and boost long-term success and competitiveness by doing so.

3. DATA & METHODOLOGY

This study collected over 20,000 United States top digital company reviews from Indeed.com using scraping. The analysis focuses on a subset of more than 5,000 reviews from January 2018 through October 2018 related to 17 selected IT companies. An employee review contains the textual description and a star rating in which 1 means a negative experience, and 5 indicates a positive experience. Appendix B, Figure 1 summarizes all the processes for this paper.

Text Pre-processing

Latent Dirichlet Allocation (LDA), a popular machine learning approach commonly used for topic modeling, is employed to analyze the impact of work-life balance on job satisfaction. After the employee reviews have been collected, the first step is to pre-process the data to prepare it for analysis. The initial analysis stage involves data pre-processing, which includes cleaning the employee reviews by removing HTML tags, web links, punctuation marks, non-alphanumeric characters, special symbols, and white spaces. All the text data is converted to lowercase, and duplicated rows are removed. Tokenization is then performed, where the data is broken down into individual words or phrases. Stop words, which do not hold analytical value (e.g., "a," "and" "the"), are removed. Lemmatization reduces words to their root form, and nouns are extracted. N-grams are generated to capture the relationship between words and the context in which they appear, specifically bigrams which are contiguous sequences of two words. Incorporating N-grams aims to enhance the analysis and gain a more nuanced understanding of the textual data.

Topic Modeling

Topic modeling is a technique for uncovering hidden topics in a large text corpus. The Latent Dirichlet Allocation (LDA) algorithm is one of the most widely used topic modeling methods. It represents documents arising from multiple topics, where a topic is defined as a distribution over a fixed vocabulary of terms by Blei and Lafferty (2009). LDA asserts that probabilistically distributed topics can be represented by words, as described by Blei et al. (2003). The text data must be pre-processed and expressed numerically to apply LDA for topic modeling.

Creating a bag-of-words (BoW) text representation is essential for the LDA model. The bag-of-words model is a statistical framework for representing text data as a numerical vector, where each dimension corresponds to a unique

 ISSN: 2473-4901
v9 n5915

word in the text corpus's vocabulary, and its focus is solely on the frequency of occurrence of each word in the document. (Zhang et al., 2010). The first step in this process is to split pre-processed data into training and test sets with a 0.4 test and 0.6 training split, a commonly used ratio. This split is done to evaluate the performance of the model on unseen data. Next, a dictionary is created to represent the vocabulary of the text corpus using tokenized data. An individual integer ID is given to each distinct word in the corpus. As LDA uses numerical data rather than text data, this is important. Once the dictionary is created, a document-term matrix is generated. Each row indicates a document, and each column shows a word in this numerical representation of the text data. The matrix shows word frequency in the document. This matrix is referred to as a "bag-of-words" (BoW) representation since it only considers the frequency of each expression and ignores the order in which the words appear in the document.

The underlying topics in the corpus can be determined using LDA by numerically expressing the text data using a document-term matrix. The objective is to pinpoint the subjects most relevant to the data and use them to comprehend the text corpus.

The LDA model can be used after creating the BoW representation. The model's hyperparameters must first be optimized before it can be used. This process is important since they control the model's behavior and can significantly affect the quality of the topic model provided. A grid search approach is used to find the alpha and beta values that maximize the coherence score.

When the alpha and beta hyperparameters have been optimized, the LDA model can be used on the training and test set. After using the LDA model on the training and test sets, assessing its performance is crucial to ensure the topics it generates are meaningful and understandable. To evaluate the performance of the LDA model, perplexity and coherence scores, commonly used metrics, are calculated for both the training and test data sets to assess the performance of the LDA model.

Perplexity measures how well the model predicts new data, and lower values indicate better performance (Blei, 2003). Coherence refers to a metric used to evaluate the level of semantic similarity among the most relevant or high-scoring words in a topic. (Röder et al., 2015). It measures how interpretable and meaningful the model's output is. Higher coherence values indicate better performance.

By evaluating both the perplexity and coherence scores, the quality of the topic model produced by the LDA algorithm can be determined. Low perplexity and high coherence scores show that the LDA model has made meaningful topics that accurately represent the text corpus's fundamental themes and recurring patterns.

The final step in the LDA topic process involves assigning topics to all the data and calculating the probability that each text belongs to each topic. This step is essential as it enables us to recognize the underlying themes and patterns existing in the corpus of texts, which can be helpful for later tasks like sentiment analysis.

In conclusion, the alpha and beta hyperparameters of the LDA model are tuned before being applied to the text corpus. The model's performance is assessed, topics are allocated to all data, and the probability of the topic that each text belongs to each subject is calculated.

Comparison with Human Rater

To compare the topics generated by the LDA model with those of human raters, four independent researchers who understand natural language processing and textual analysis were requested to examine employee reviews gathered and identify the topics mentioned. The four researchers belonged to the same university and individually picked 33 reviews at random for this task. The objective is to compare their results to the LDA's to identify similarities or differences.

The Jaccard similarity score measures the similarity between two sets by dividing the intersection size by the union size. This method was also employed by Guo et al. (2017) to compare the dimensions of previous studies and their analysis. This study uses the Jaccard similarity score to compare the topics identified by an LDA model to those identified by independent human researchers. A comparison is made between the topic assignments of the LDA model and those of four independent human researchers (A, B, C, and D). The score ranges from 0 to 1, where 0 suggests no overlap between the two sets and 1 indicates a perfect match.

This approach is helpful for several reasons. It first provides a benchmark against which to compare the performance of the LDA model. We may assess the model's topic assignment accuracy by comparing the topics the model

 ISSN: 2473-4901
v9 n5915

identified with those identified by human raters. Second, this process can give insight into how different people perceive and identify the topics in the same set of texts. The LDA model and human raters can be compared to reveal their strengths and weaknesses. Lastly, the LDA model's topics may be validated using this approach. The topics are significant and relevant to the text corpus if the topics identified by the model and human raters overlap.

Overall, comparing the LDA model's topic assignments with the outcomes of human raters helps assess the model's performance, understand how various individuals perceive topics, and validate the topics the model identified.

Sentiment Analysis
Sentiment analysis analyzes written or spoken material, such as a social media post or review, to assess its emotional tone. (Jung & Suh, 2019) The procedure uses computational and natural language processing to identify and extract text sentiment. Numerous techniques and tools are available for sentiment analysis, with the Natural Language Toolkit (NLTK) being one of the most popular. The NLTK Python library provides plenty of tools and pre-trained models for sentiment analysis. This library allows the text to be analyzed and classified as positive, negative, or neutral. (Yao, 2019) This study uses the VADER (Valence Aware Dictionary for Sentiment Reasoning) lexicon of the NLTK library to generate sentiment analysis for each topic.

VADER measures the text's overall sentiment score by aggregating the sentiment scores of individual words from a dictionary of terms labeled with their corresponding sentiment scores (positive, negative, or neutral). (Elbagir & Yang, 2019) With this technique, it is possible to understand the sentiment expressed in the text in more depth and nuance.
Consequently, by using the VADER lexicon of the NLTK library, we can conduct sentiment analysis on the topics generated by the LDA model and understand the emotional tone and sentiment expressed in the text.

4. RESULTS

After pre-processing, text data only includes nouns and bigrams. According to Figure 2 and Figure 3, the most frequent word is management, and the most frequent bigram is life balance.

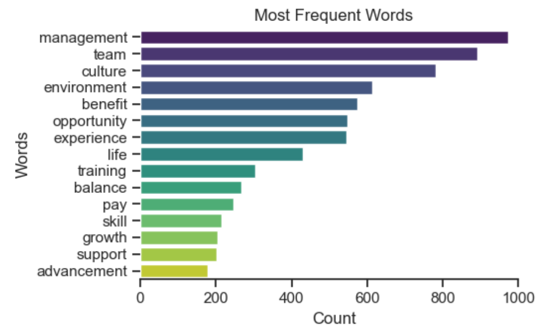


Figure 2 The Most Frequent Words

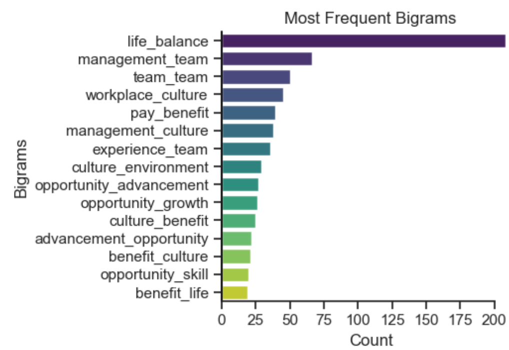


Figure 3 The Most Frequent Bigrams

The final alpha and beta values were found to be 0.31 through the LDA tuning and grid search processes. It is important to emphasize that the entire (100%) corpus, not just a subset (75%), was used for the tuning and grid search process. The whole corpus is used for LDA tuning and grid search to ensure that the resulting topic model is representative of the entire corpus. Suppose only a subset of the corpus was used. In that case, likely, the resulting model might not fully account for the complexity of the text corpus, which could lead to biased or unreliable conclusions.

Alpha and beta are critical parameters in the LDA model that affect the final topic model. As Blei (2012) explains, alpha determines each document's topic distribution, with higher alpha values indicating that each document is more likely to contain diverse topics. In contrast, beta is a parameter that controls the distribution of words within each topic, with higher beta values indicating that each topic is more likely to include multiple words.

Appendix C, Table 2 shows alpha and beta values with the highest coherence score. The topic modeling metric with the highest coherence score is crucial since it shows that the results are more

 ISSN: 2473-4901
v9 n5915

understandable and coherent. High coherence scores make each topic's top words more semantically similar, making them easier for human interpreters to comprehend. Therefore, the alpha and beta values that give the highest coherence score were found to be 0.31.

The LDA model is trained using these values after determining the optimal alpha and beta values. The final model has nine topics. Appendix D, Table 3 shows job satisfaction factors and their associated keywords identified by the LDA model. This information helps us understand the key topics and trends in the text corpus linked to job satisfaction.

By identifying these factors and keywords, we can determine which aspects of employee job satisfaction are most important. This information can assist firms in enhancing job satisfaction by helping them understand the factors that affect it.

Additionally, the LDA model is used to compute the frequency of each topic in the text corpus. The frequency of each topic in the text corpus can give helpful insight into which employees frequently mention topics and, as a result, may be particularly significant or influential for their overall work experience and job satisfaction.

Table 4 displays a detailed ranking of topics from the LDA model based on frequency. Frequencies indicate each topic's importance to study respondents; one such frequency, 23.19% for Management Skills and Responsibilities, topped this list, showing its prominence among employees regarding job satisfaction. By ranking topics this way, we aim to illustrate which factors most contribute to job satisfaction among participants.

Rank	Topic	Frequency	Percentages
1	1	1348 docs	23.19
2	2	923 docs	15.86
3	6	650 docs	11.18
4	9	621 docs	10.68
5	8	576 docs	9.91
6	7	576 docs	9.91
7	3	538 docs	9.25
8	5	344 docs	5.92
9	4	238 docs	4.09

Table 4: Dominant Topics and Their Ranking

Table 5 presents the performance scores for the LDA model, including coherence scores and perplexity scores for both the training and test datasets. Higher coherence scores indicate more coherent topics. Conversely, the perplexity score measures the model's prediction ability, with lower numbers suggesting better ability.

Set	Perplexity Score	Coherence Score
Training	-6.79	0.58
Test	-7.25	0.77

Table 5: Performance Scores

The training dataset's coherence score is 0.58, which indicates that the LDA model's generated topics are moderately coherent. The test dataset's coherence score is more excellent (0.77), demonstrating the LDA model's ability to generalize effectively to new data and provide more coherent topics. This coherence score is crucial since it suggests that the LDA model can accurately capture text corpus patterns and themes even with new data.

The training dataset's perplexity score is -6.79, which shows that the LDA model can predict the text corpus relatively accurately. The test dataset's perplexity score is slightly higher, at -7.25, indicating that the LDA model can still accurately predict it.

According to the scores, the LDA model can produce accurate and coherent topics and generalize effectively to new data. The scores suggest that the LDA model may be able to reveal themes and patterns in large text collections.

To assess the LDA model's accuracy and reliability, topic assignments from the LDA model were compared to the ratings provided by human raters. The comparison explicitly considers whether the LDA model can reliably identify the themes mentioned in the text corpus and whether its conclusions are compatible with human assessors' reports.

The comparison's findings revealed that both the LDA model and human raters identified similar themes, particularly related to work culture, management, work-life balance, and benefits.

Appendix E, Table 6 provides a comprehensive summary of how topic identification by the LDA model and human evaluation is similar. Comparing these results helps us assess the LDA model's accuracy, reliability, and ability to detect text corpus topics.

ISSN: 2473-4901
v9 n5915

According to Jaccard scores in Table 6, The Jaccard similarity scores obtained are as follows:

- LDA vs. Independent Researcher A: 0.42
- LDA vs. Independent Researcher B: 0.6
- LDA vs. Independent Researcher C: 0.42
- LDA vs. Independent Researcher D: 0.5

In this study, the Jaccard similarity scores show moderate agreement between the LDA model and the independent researchers. The LDA model and Independent Researcher B had the highest agreement (0.6), followed by Independent Researcher D (0.5), and the LDA model and Independent Researchers A and C had the lowest agreement (0.42).

In this study's final step, employee reviews were analyzed using sentiment analysis to determine how individuals felt about various job satisfaction-related topics. Figure 4 shows that most reviews were positive, organizational strategy and commitments were the least positive, and compensation and benefits were the most positive.

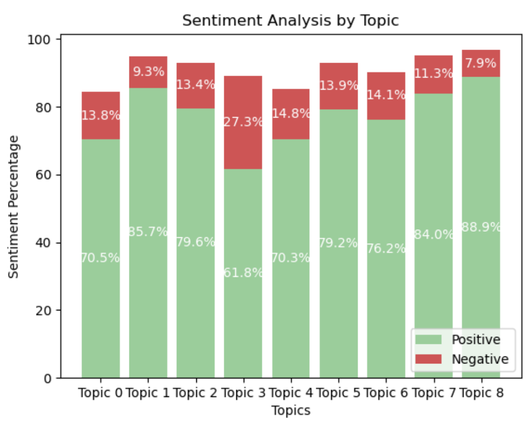


Figure 4 Sentiment Analysis by Topics

Word clouds were created for both positive and negative sentiment words related to all factors to understand better the sentiment conveyed towards specific job satisfaction factors, as seen in Figures 5 and 6. The word clouds highlight the words that refer to job satisfaction factors most frequently in positive and negative ways. According to Figures 4 and 6, benefit, management, and culture are positive words, while complaint, training, and pressure are negative.

Figure 5 Positive Sentiment Words

Figure 6 Negative Sentiment Words

Findings
Based on the analysis conducted in this study, several key findings can be drawn about the factors affecting employee job satisfaction.

This study used the LDA model to find job satisfaction factors in the IT industry and began with text pre-processing as the first step of the analysis. After pre-processing the text data to include only nouns and bigrams, the study of the most frequent words and bigrams identified management as the most frequent word and life balance as the most frequent bigram. This finding supports the importance of management-related factors and work-life balance in employee job satisfaction, as identified in the subsequent analysis using the LDA model and sentiment analysis techniques.

The LDA model identified nine critical topics related to job satisfaction, including management skills and responsibilities, compensation and benefits, work-life balance, and organizational culture and environment. Frequency analysis revealed management skills and responsibilities as the most frequently discussed topic, ranking first overall and underscoring their significance to employee job satisfaction. The coherence and perplexity scores for the LDA model were also evaluated to determine its performance. The results showed that the LDA model could

ISSN: 2473-4901
v9 n5915

generate coherent and accurate topics and generalize well to new data.

The results of the Jaccard similarity score indicate that the LDA model's topic assignments show some similarity to those made by independent researchers. Comparing the LDA model's topic assignments with the results of human raters also showed similarities between the topics identified by the two groups, with the work culture and environment, management, and work-life balance. The results indicate moderate agreement between the LDA model and the human researchers, with some variation observed across different comparisons.

The highest Jaccard similarity score was observed between the LDA model and Independent Researcher B (0.6), suggesting that the model's topic assignments, in this case, were in closer agreement with the human researcher. This could be attributed to the specific nature of the text samples analyzed, the parameters used in the LDA model, or the topic categorization criteria employed by Researcher B, which may have been more consistent with the model's output.

On the other hand, the lowest agreement was observed between the LDA model and Independent Researchers A and C (0.42). The differences between the LDA model and these researchers could be due to various factors, such as differences in the researchers' domain expertise, subjective interpretation of the text, or the inherent limitations of the LDA model in capturing subtle distinctions between topics. This finding suggests that there may be specific topic assignments or nuances in the text data that the LDA model struggled to capture accurately.

The average Jaccard similarity score between the LDA model and Independent Researcher D (0.5) further supports the notion that the model's performance exhibits some similarity to human annotators, but there is still room for improvement. The discrepancies observed could result from the LDA model's assumptions, parameter settings, or the quality of the input data.

Finally, sentiment analysis was used to determine the overall sentiment toward job satisfaction factors. The study showed that most employee reviews were positive, with compensation and benefits receiving the highest positive sentiment. Organizational strategy and commitments received the highest percentage of negative sentiment. The word clouds generated for positive and negative words related to job satisfaction factors highlighted the most frequently used positive and negative words related to all elements.

Discussion

In the IT industry, employee satisfaction has become increasingly important due to the impacts of the COVID-19 pandemic on layoffs. As companies compete to attract potential employees with valuable offers, there is a need for additional research on employee satisfaction and feasible solutions.

Several studies examined how vital employee satisfaction is in the IT industry, and the results show that certain factors are crucial. Ganga's (2022) research found that a positive work environment and an excellent work-life balance were the two most important determinants of job satisfaction across industries. This study also emphasizes the importance of the workplace in the IT sector, suggesting that organizations should prioritize a comfortable and productive workplace.

Moreover, the research by Sainju et al. (2021) demonstrates that management is also a significant factor in the IT sector's job satisfaction. Employees in this industry value effective management practices and need a supportive, empowering work environment to thrive. The study results highlight the significance of companies implementing management strategies prioritizing employee satisfaction.

Moro et al. (2021) identified work exhaustion as the primary cause of job dissatisfaction in the IT industry. However, this current study found that job experience and flexibility are the most dissatisfying factors. These results suggest that companies must create a flexible work environment that allows employees to enhance their job experience.

Finally, Jung and Suh's (2019) study found that project planning is a new factor for job satisfaction. Similarly, this study highlights the importance of project planning in job satisfaction. These findings suggest that companies must prioritize project planning as a critical factor in creating a conducive work environment for their employees.

One of the unique characteristics of this study is its in-depth investigation of factors not addressed prominently in previous research. This research focuses on IT industry specifics, unlike previous studies that focused on generic issues like workplace environments and management. We

ISSN: 2473-4901
v9 n5915

highlight the often under-emphasized significance of job experience and flexibility as key determinants of job satisfaction. The findings show that project planning is crucial to IT job satisfaction. Using powerful NLP techniques and an in-depth dataset, this research reveals what IT employees genuinely value, enabling focused human resource practices to improve job satisfaction.

The Jaccard similarity scores obtained in this study demonstrate that the LDA model's topic assignments show some similarity to those made by human researchers. However, the varying levels of agreement across different comparisons suggest that the model's performance could be further optimized to better align with human assessments. Future research could focus on refining the LDA model's parameters, exploring alternative topic modeling approaches, or incorporating additional domain knowledge to improve the model's performance and achieve better agreement with human annotators.

Overall, this research offers valuable insight into the factors contributing to job satisfaction in the IT industry. By acknowledging the significance of work environment, management practices, work flexibility, and project planning, businesses can create a better workplace for employees and improve their job satisfaction.

Challenges

The results of this study highlight the importance of utilizing human rater and LDA topic modeling in analyzing employee reviews for IT companies. Both methodologies observed work culture and environment, management, and work-life balance and benefits, which were similar. The result indicates that using multiple methods can provide a more comprehensive understanding of the topics in the data.

However, the human rater exercise also revealed challenges in analyzing text data. One of the challenges was the variability in the number of topics identified by each independent researcher. Although the goal was to obtain nine topics from each researcher, some identified only 7-8 topics. The various approaches and findings from each group highlight the subjectivity of human coding and the possibility of inconsistency in the coding process. Therefore, it is essential to establish clear coding guidelines and procedures for ensuring analysis consistency and accuracy. In addition, the complexity of the data requires expertise in both natural language processing and the industry being analyzed. Interdisciplinary collaboration and expertise in conducting text analysis research are needed.

Practical Implications and Implementation for IT Companies

Practicality is paramount in our research findings since theoretical understandings do not lead to real-world changes alone. This analysis diagnoses and informs IT firms on job satisfaction determinants. IT firms seeking actual benefits from our study should consider how it may be used.

By understanding and addressing key elements such as management effectiveness, work-life balance, and organizational culture, companies can improve employee satisfaction rates and enhance retention rates - ultimately decreasing turnover costs while creating a cohesive and productive work environment. Satisfied workers can boost innovation, productivity, and brand image but may also create unique challenges.

Translating research into practice can present unique challenges, with companies facing initial resistance when trying to alter long-standing practices or introduce new ones. Allocating appropriate time and finances can be demanding; furthermore, after implementation, a constant feedback process is needed to guarantee that new strategies fulfill employee demands and adapt to changing dynamics. Despite such difficulties, however, IT companies could stand to gain immensely by making this transition systematically and with dedication; their potential benefits could be immense.

6. CONCLUSIONS

This paper explores IT job satisfaction factors by analyzing five thousand employee reviews from 17 prominent digital companies using LDA. Data pre-processing includes stop word removal, lemmatization, noun extraction, and bigram generation for nuanced text data analysis. The study identifies nine key topics IT employees value, with management being the most frequent word and life balance being the most frequent bigram. Comparison with the topics identified by independent researchers reveals moderate overlap, particularly in work culture and environment, management, and work-life balance and benefits. The Jaccard similarity scores obtained in the study highlight that the LDA model's topic assignments show some similarity to those made by independent researchers.

ISSN: 2473-4901
v9 n5915

The sentiment analysis is also conducted to classify each review's tone, finding that most reviews are positive. Notably, compensation and benefits have the highest percentage of positive sentiment, while organizational strategy and commitments have the highest rate of negative sentiment.

Finally, the analysis visualizes positive and negative sentiment words related to all factors, with benefit, management, and culture as positive words, while complaint, training, and pressure are negative.

However, it is essential to note that this study only analyzed employee reviews from 17 leading digital companies. Thus, the results may not represent the broader IT industry, as smaller or less prominent companies were not included.

In summary, this study provides valuable insights into what IT employees value in their work environment and highlights the importance of factors to improve employee satisfaction and retention. Organizations can use these findings to devise more effective support strategies that ensure success for IT employees. Future studies may expand by studying diverse companies or including more variables for a comprehensive understanding of job satisfaction within IT firms. Still, nonetheless, this work serves as a basis for companies looking to enhance employee well-being.

7. REFERENCES

Barnett, A., & Li, M. (2023, January 3). Tech layoffs are happening faster than at any time during the pandemic*. The Wall Street Journal.* Retrieved March 6, 2023, from https://www.wsj.com/articles/tech-layoffs-are-happening-faster-than-at-any-time-during-the-pandemic-11672705089

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In A. Srivastava, & M. Sahami (Eds.), *Text Mining: Classification, Clustering and Applications* (pp. 71-93). Cambridge: Chapman and Hall/CRC.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Challenger Gray & Christmas Inc. (2023). Job cuts hit 77,770 in February 2023; highest ytd since 2009. *Challenger Gray & Christmas Inc.* Retrieved March 13, 2023, from https://omscgcinc.wpenginepowered.com/wp-content/uploads/2023/03/The-Challenger-Report-February23-1.pdf

Dai, W., & Wu, N. (2017). Profiling essential professional skills of chief data officers. In *Proceedings of Twenty-Third Americas Conference on Information Systems.*

Deagon, B. (2023). Tech industry layoffs show no signs of abating as businesses undo overhiring. *Investor's Business Daily.* Retrieved March 10, 2023, from https://www.investors.com/news/technology/tech-layoffs-show-no-signs-of-slowing/

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, *101*(3), 621–640. https://doi.org/10.1016/j.jfineco.2011.03.021

Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, *122*.

Ganga, A. (2022). Employees satisfaction in different industries: an exploratory review of the literature. *International Journal of Economics and Management Systems*, *7*. http://www.iaras.org/iaras/journals/ijems

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, *59*, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Holland, P., & Bardoel, A. (2016). The impact of technology on work in the twenty-first century: exploring the smart and dark side. *International Journal of Human Resource Management*, *27*(21), 2579–2581. https://doi.org/10.1080/09585192.2016.1238126

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, *38*(3), 635-872.

Jung, Y., & Suh, Y. (2019). Mining the voice of employees: a text mining approach to identifying and analyzing job satisfaction factors from online employee reviews.

ISSN: 2473-4901
v9 n5915

Decision Support Systems, *123*. https://doi.org/10.1016/j.dss.2019.113074

Kalra, V., & Aggarwal, R. (2017). Importance of text data pre-processing & implementation in rapidminer. In *Proceedings of the First International Conference on Information Technology and Knowledge Management*, *14,* 71–75. https://doi.org/10.15439/2018KM46

Luo, N., Zhou, Y., & Shon, J. J. (2016). Employee satisfaction and corporate performance: mining employee reviews on Glassdoor.com. In *Proceedings of the 37th International Conference on Information Systems.*

Moro, S., Ramos, R. F., & Rita, P. (2021). What drives job satisfaction in IT companies? *International Journal of Productivity and Performance Management, 70*(2), 391–407. https://doi.org/10.1108/IJPPM-03-2019-0124

Oshagbemi, T. (2003). Personal correlates of job satisfaction: empirical evidence from UK universities. *International Journal of Social Economics, 30*(11–12), 1210–1232. https://doi.org/10.1108/03068290310500634

Rast, S., & Tourani, A. (2012). Evaluation of employees' job satisfaction and role of gender difference: an empirical study at airline industry in Iran. *International Journal of Business and Social Science, 3*(7), 91–100.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408. https://doi.org/10.1145/2684822.2685324

Sainju, B., Hartwell, C., & Edwards, J. (2021). Job satisfaction and employee turnover determinants in Fortune 50 companies: insights from employee reviews from Indeed.com. *Decision Support Systems, 148*. https://doi.org/10.1016/j.dss.2021.113582

Yang, S., & Zhang, H. (2018). Text mining of Twitter data using a latent dirichlet allocation topic model and sentiment analysis. *International Journal of Computer and Information Engineering, 12*(7).

Yao, J. (2019). Automated sentiment analysis of text data with NLTK. *Journal of Physics: Conference Series, 1187*(5). https://doi.org/10.1088/1742-6596/1187/5/052020

Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics, 1*(1–4), 43–52. https://doi.org/10.1007/s13042-010-0001-0

ISSN: 2473-4901
v9 n5915

APPENDIX A
Table 1: Literature Review

Source	**Purpose**	**Data & Method**	**Findings**
Dai and Wu (2017)	The study examines current chief data officers' professional skills and education (CDOs) by analyzing their LinkedIn resumes using topic modeling techniques.	621 members of the CDO group on LinkedIn Topic Modelling: NMF and Latent Dirichlet Allocation (LDA)	This study finds that CDOs typically have diverse skills and educational backgrounds, including expertise in business strategy, data governance, and data architecture. Additionally, more specific and coherent skills are captured by NMF than LDA.
Edmans (2011)	This study investigates the link between long-run stock returns and employee satisfaction.	Great Place to Work Institute created 57 question survey. ("100 Best Companies to Work for in America." is the data source.) Regression analysis	According to this study, shareholder returns and employee satisfaction are positively associated.
Ganga (2022)	This study analyzes the employee review literature in different industries and finds the reasons for job satisfaction or dissatisfaction.	Twelve research papers from 2010 to 2020 across various industries and many employee review websites. Topic Modelling: LDA	This study finds that work environment and work-life balance play an essential role in most industries.
Guo et al. (2017)	This study examines how tourists express their satisfaction in reviews from TripAdvisor.	Two hundred sixty-six thousand five hundred forty-four hotel reviews from the TripAdvisor website. Topic Modelling: LDA	This study concludes that online reviews could provide valuable insights into tourist satisfaction and that LDA is an effective tool for analyzing these reviews.

 ISSN: 2473-4901
v9 n5915

Jung and Suh (2019)	This study analyzes online employee reviews to identify factors contributing to job satisfaction.	204,659 online employee reviews from 4,347 firms in 10 industries on jobplanet.co.kr (35,063 reviews about 844 firms in the IT industry in South Korea used for topic modeling) Topic Modelling: LDA	This study finds five new job satisfaction factors that had not been considered in the literature: Project, Software development, Inter-firm relationship, Marketing, and Overseas business. This study concludes that employee reviews provide insight into specific aspects of the job and the organization that affects job satisfaction, such as communication, management, and work-life balance.
Kalra and Aggarwal (2017)	This study highlights the importance of text data pre-processing and how it can be implemented using the RapidMiner tool.	The Times of India news site contains the 44th President Obama's letter. Various Text Mining Algorithms	This study concludes that data preparation is crucial in machine learning (ML), and pre-processing text data is necessary for cleaning, transforming, and formatting it for use in ML algorithms.
Luo et al. (2016)	This study analyzes employee satisfaction and its relation to corporate performance from anonymous employee reviews on Glassdoor.com.	Two hundred fifty-seven thousand four hundred fifty-four reviews from 425 companies representing 21 industry sectors on Glassdoor.com from 2008 to 2014. Text mining, descriptive data analysis, and regression analysis.	This study shows the top 5 industries that received the most reviews: Technology, Retailing, Financials, Telecommunications, and Health Care. According to findings, innovation plays a significant role in the technology industry, while quality is the driving force in the retailing and financial sectors. Overall, employee satisfaction and corporate performance have a significant correlation. The study identifies safety, communication, and integrity as negatively correlated with performance.
Moro et al. (2021)	This study examines the factors that contribute to job satisfaction among employees in IT companies	Fifteen thousand reviews from the top 15 US technology companies from Glassdoor.com. Support vector machine (SVM).	This study finds that communication, management, and work-life balance are essential drivers of job satisfaction. This study also finds positive attitudes of coworkers, contributing to job satisfaction. However, the main reason for job dissatisfaction is work exhaustion.

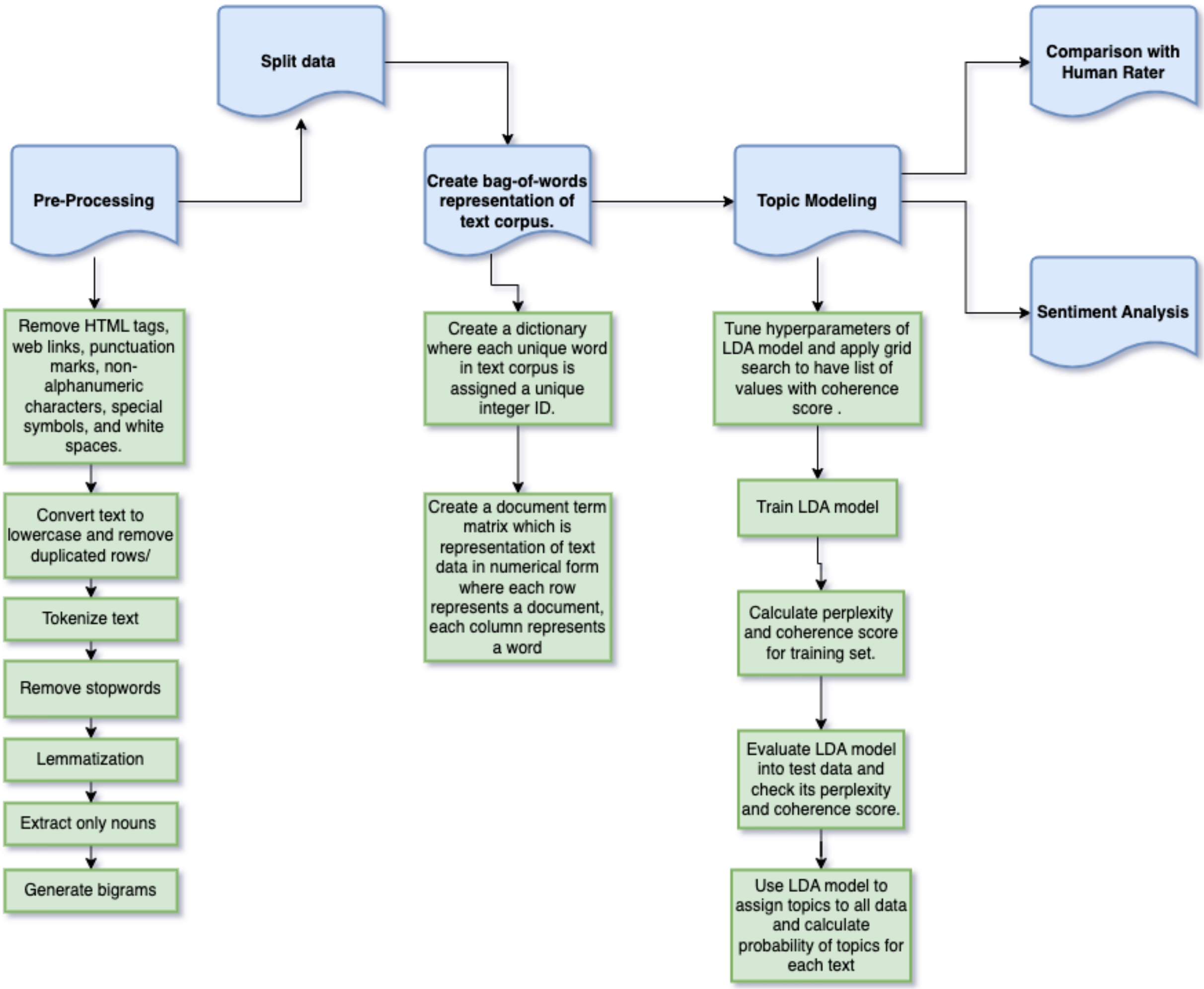
ISSN: 2473-4901
v9 n5915

Oshagbemi (2003)	This study examines the research on the associations between job satisfaction and factors including age, gender, rank, and length of service.	Questionnaire - university teachers in the UK, 23 institutions. Job Descriptive Index, Regression analysis	This study identifies two significant factors that determine an individual's level of job satisfaction in higher education: their academic rank and length of service. It also shows an individual's gender, age, and length of service at their current university do not directly affect their overall job satisfaction. Still, gender and academic rank are statistically significant predictors, as well as age and length of service in higher education. However, a significant positive association exists between job satisfaction and academic rank, while length of service is negatively related.
Rast and Tourani (2012)	This study assesses the degree of job satisfaction among employees and investigates how gender impacts their job satisfaction.	Survey and questionnaires - employees from 3 private airlines in Iran. Descriptive analysis, Independent-sample t-test	This study identifies several key factors that contribute to job satisfaction, including supervision, relationships with coworkers, current salary, work type, and career advancement opportunities. According to the findings, there is no notable disparity in job satisfaction levels between male and female employees, and, in general, employee satisfaction with their jobs is moderate.
Sainju et al. (2021)	This study examines hidden aspects of employee satisfaction.	Six hundred eighty-two thousand one hundred seventy-six employee reviews of Fortune 50 companies from Indeed.com. Structural Topic Modeling (STM).	This study finds that employees in the retail industry prioritize Pay & Benefits and Length of Breaks when it comes to job satisfaction. In contrast, employees in the technology sector place greater emphasis on achieving a healthy Work-Life Balance. This study also suggests that management is a crucial job satisfaction factor.
Yang and Zhang (2018)	This study analyzes audience reviews of Thor movie on the released day from tweets.	One hundred eighty-five thousand one hundred eighty-five retrieved tweets from Twitter. Topic Modeling: LDA Sentiment analysis	This study finds that LDA is an effective tool for identifying the topics discussed on Twitter related to a particular event or topic, and sentiment analysis can be used to determine the overall sentiment of tweets related to that topic.

ISSN: 2473-4901
v9 n5915

APPENDIX B

Figure 1: Proposed Methodology



 ISSN: 2473-4901
v9 n5915

APPENDIX C
Table 2: LDA Tuning Results

Validation Set	Topics	Alpha	Beta	Coherence
75% Corpus	9	0.31	0.31	0.525523
100% Corpus	9	0.31	0.31	0.51706855
75% Corpus	10	0.31	0.31	0.51509178
75% Corpus	10	0.31	0.01	0.51375766

APPENDIX D
Table 3: Job Satisfaction Factors

No	Factors	Keywords
1	Management skills and responsibilities	management, skill, meeting, value, security
2	Workplace culture and environment	culture, environment, support, workplace, compensation
3	Job experience and flexibility	experience, home, role, money, direction
4	Organizational strategy and commitments	politic, mission, strategy, budget, commitment
5	Project planning	project, expectation, stress, goal, user
6	Teamwork and collaboration	team, family, quality, feedback, ability
7	Work-life balance and well-being	life, training, balance, schedule, hr
8	Career advancement and leadership development	opportunity, advancement, leadership, contract, promotion
9	Compensation and benefits	benefit, pay, growth, atmosphere, location

 ISSN: 2473-4901
v9 n5915

APPENDIX E

Table 6: A Comparison of Topics Between LDA and Human Evaluation (Researcher)

Factors	**LDA**	**A**	**B**	**C**	**D**
Management skills and responsibilities	✔	✔	✔	✔	✔
Workplace culture and environment	✔	✔	✔	✔	✔
Job experience and flexibility	✔	✔	✔	×	×
Organizational strategy and commitments	✔	×	×	×	×
Project planning	✔	×	×	×	×
Teamwork and collaboration	✔	×	×	×	✔
Work-life balance and well-being	✔	×	✔	✔	✔
Career advancement and leadership	✔	✔	✔	✔	✔
Compensation and benefits	✔	✔	✔	✔	✔
Coworkers	×	✔	×	×	×
Vacation	×	✔	×	×	×
Diversity	×	✔	×	✔	×
People	×	×	✔	×	×
Education	×	×	×	✔	×
Challenges	×	×	×	✔	×
Networking	×	×	×	×	✔

Notes: ✔ = included × = not included. The Jaccard coefficient is 0.42 between LDA analysis and researcher A. The Jaccard coefficient is 0.6 between LDA analysis and researcher B. The Jaccard coefficient is 0.42 between LDA analysis and researcher C. The Jaccard coefficient is 0.5 between LDA analysis and researcher D.