Sentiment Analysis and Topic Modeling of Reddit Data

Drashya Nileshbhai Babariya
Dept. of Computer Science
Montcliar State University
Montclair, NJ, USA
babariyad1@montclair.edu

Elamathy Sivakumar
Dept. of Computer Science
Montcliar State University
Montclair, NJ, USA
sivakumare1@montclair.com

Swathi Palanisamy
Dept. of Computer Science
Montcliar State University
Montclair, NJ, USA
palanisamys1@motclair.edu

Gauri Prakash Dhanawade
Dept. of Data Science
Montcliar State University
Montclair, NJ, USA
dhanawadeg1@montclair.com

Vamshi Krishna Dara
Dept. of Data Science
Montcliar State University
Montclair, NJ, USA
darav1@montclair.edu

Nitesh Jha
Dept. of Computer Science
Montcliar State University
Montclair, NJ, USA
jhan1@montclair.edu

***Abstract*—The post-pandemic transition from remote work to office work has led to diverse opinions among employees, ranging from financial incentives to workplace comfort. This paper presents a comprehensive sentiment analysis and topic modeling approach applied to user-generated Reddit comments. Utilizing Natural Language Processing (NLP) techniques such as VADER Sentiment Analysis and Latent Dirichlet Allocation (LDA), the study classifies comments into sentiment categories and extracts prominent themes. The dataset, comprising 465 Reddit comments, underwent preprocessing, analysis, and visualization. Key findings indicate that salary increases (60%) are the most cited motivator for returning to office spaces, followed by workplace comfort (20%) and public transport improvements (20%). This research provides actionable insights for organizations to formulate data-driven policies and strategies for encouraging employees to return to office work.**

Keywords—Sentiment Analysis, Topic Modeling, Reddit Data, VADER, Latent Dirichlet Allocation, Return-to-Office, Data Analytics; Hybrid Learning; Recommenders; Opinion Mining

I. INTRODUCTION

Sentiment analysis is basically the interpretation and categorization of emotions expressed in textual data. In the modern digital era, when users generate overwhelming data each day through reviews, social media posts, and feedback forms, it allows organizations to estimate public opinion and behavioral patterns, which are priceless. It also enables, for instance, improvements to be made in the area of customer experience by businesses and aids political groups to get a feel for how their policies resonate with the general public. Basically, insights derived from such vast amounts of data and variability drive today's decision-making based on big data.

Sentiment analysis is more important than it is commonly thought or known to be. It sets the procedure for finding emerging trends, areas of improvement, and future trend prediction. However, with all its wide adoptions, sentiment analysis is not out of complexities such as linguistic ambiguity, subjectivity of human emotions, and contextual nuances of language. These challenges require sophisticated techniques and algorithms to ensure that results are accurate and relevant. This paper discusses the basic techniques, challenges, and applications of sentiment analysis, using a dataset of user comments as a practical case study. The following project shows how organizations can decode user perspectives and identify key concerns associated with the topic by categorizing comments into topics and analyzing their associated sentiments.

1.1 Background: The COVID-19 pandemic triggered a dramatic transformation in the way individuals work, with remote work becoming a necessity rather than a choice. This unprecedented shift provided employees with newfound flexibility, improved work-life balance, and freedom from daily commuting. As businesses and organizations now transition back to traditional office settings, many employees have expressed reluctance to return, citing various concerns ranging from financial incentives and improved infrastructure to workplace comfort and flexibility. This reluctance presents significant challenges for employers seeking to balance business objectives with employee satisfaction and productivity.

Understanding employee sentiment and the underlying factors influencing their willingness to return to office spaces is crucial for developing effective, evidence-based policies. While traditional surveys offer structured insights, user-generated content on social media platforms like Reddit provides unfiltered, candid opinions that reflect real-world concerns. Reddit, a popular online forum, hosts vibrant discussions on topics ranging from global trends to niche workplace issues. By analysing such discussions, organizations can gain deeper insights into employee priorities, uncover recurring themes, and identify actionable strategies for smoother transitions to office-based work.

This study focuses on analysing Reddit comments from a targeted discussion thread:

"What would actually make you return to the office?"

The thread serves as a rich source of qualitative data, capturing diverse perspectives from users across various industries, demographics, and geographies. The unstructured nature of Reddit comments poses challenges for analysis but also provides a unique opportunity to apply advanced Natural Language Processing (NLP) techniques to extract meaningful insights.

1.2 Research Objectives: The primary objectives of this study are as follows:

• Sentiment Classification: To analyse Reddit comments and classify user sentiments into positive, negative, or neutral categories using VADER Sentiment Analysis, a robust rule-based NLP tool optimized for social media text.

• Topic Extraction: To identify key themes and topics influencing return-to-office decisions using Latent Dirichlet Allocation (LDA), a topic modelling technique that uncovers hidden patterns in unstructured text data.

• Actionable Insights: To provide organizations with data-driven insights into employee preferences, enabling them to design targeted policies that address financial, logistical, and psychological concerns related to returning to office spaces.

1.3 Scope and Significance: This study analyses Reddit comments collected from a single discussion thread to gain a comprehensive understanding of employee motivations and hesitations. By employing a combined approach of sentiment analysis and topic modelling, we aim to uncover the underlying drivers behind employees' return-to-office decisions. Key findings from this analysis can help businesses prioritize policies such as salary increments, improved commuting infrastructure, and enhanced workplace comfort to address employee needs effectively.

The significance of this research lies in its ability to harness unstructured social media data to generate actionable insights for organizations. Unlike structured surveys, Reddit discussions provide spontaneous, real-time opinions that reflect genuine employee concerns. This approach contributes to the growing body of research on post-pandemic workplace dynamics, offering practical recommendations for organizations navigating the transition from remote to office-based work.

II. RELATED WORK

2.1 Text Mining on Social Media Platforms: Social media platforms, including Reddit, Twitter, and Facebook, have become rich sources of unstructured data, enabling researchers to extract insights into public opinions and trends. Reddit, in particular, stands out for its structured threads and subreddits, where users engage in detailed discussions on diverse topics. Studies analysing Reddit data have demonstrated its potential for opinion mining, especially in domains such as mental health [1], public policy [2], and workplace trends.

For example, Sharma et al. [3] analysed Reddit threads to explore public mental health sentiments during the COVID-19 pandemic, identifying key themes such as isolation, anxiety, and uncertainty. Similarly, research by Hutto and Gilbert [4] established the effectiveness of sentiment analysis tools like VADER in analysing Reddit comments, validating its suitability for short, informal, and conversational text. These studies underscore the value of Reddit data for capturing nuanced opinions on emerging societal and workplace issues.

2.2 Topic Modeling for Opinion Mining: Topic modeling techniques, particularly Latent Dirichlet Allocation (LDA), have been widely adopted for extracting latent themes from large volumes of text data. LDA provides a probabilistic framework for identifying word distributions associated with distinct topics within unstructured text. Hammond et al. [5] successfully implemented LDA to analyse policy documents, highlighting its ability to uncover key themes for decision-making.

In the context of workplace studies, topic modeling has been applied to social media data to identify factors driving employee engagement, satisfaction, and dissatisfaction. For instance, Huang et al. [6] employed LDA to analyse online forums discussing remote work challenges, identifying recurring themes such as productivity, collaboration tools, and work-life balance. These findings demonstrate the relevance of LDA for extracting actionable insights from user discussions, particularly in understanding employee perspectives.

The application of LDA to Reddit data enables the discovery of context-specific themes, such as salary, workplace comfort, and commuting concerns, which are critical to understanding return-to-office preferences. This study builds on existing work by applying LDA to Reddit comments to uncover these recurring motivators.

2.3 Workplace Trends and Return-to-Office Motivations: The post-pandemic transition back to physical office spaces has been a topic of growing interest among researchers and organizations. Recent studies emphasize three primary factors influencing employees' willingness to return to the office:

• Salary Incentives: Competitive compensation remains a significant driver, with employees citing financial rewards as a key motivator for returning to physical workspaces [7]. Studies analysing workplace surveys highlight that pay increases, bonuses, and commuting allowances can significantly influence employee decisions.

• Transportation Accessibility: Research on urban workforce dynamics demonstrates the importance of affordable and efficient public transportation systems. High commuting costs and long travel times have been identified as barriers to return-to-office policies, particularly in metropolitan areas [8].

• Workplace Comfort and Flexibility: Several studies indicate that employees now expect office environments to match or exceed the comfort of home offices. Ergonomic setups, flexible schedules, and hybrid work arrangements have emerged as essential considerations for employees navigating the post-pandemic workplace landscape [9].

While these studies rely on structured survey data, they often fail to capture spontaneous and unfiltered opinions. This research addresses that gap by analysing unstructured Reddit comments to identify themes that align with salary expectations, transport challenges, and workplace comfort. By leveraging sentiment analysis and topic modeling, this study uncovers actionable insights that complement findings from traditional workplace studies.

III. METHODOLOGY

The analysis in this project was carried out using the following structured steps:

3.1 Data Collection: The dataset for this study was sourced from **Reddit**, one of the most popular social media platforms where users discuss a variety of topics. Reddit's structured nature, consisting of subreddits (thematic forums) and threaded discussions, makes it an excellent source for analyzing public opinions.

• **Targeted Discussion**: The thread titled: *"What would actually make you return to the office?"*was selected because it explicitly solicited user opinions on factors motivating a return to office spaces post-pandemic. The thread was chosen for its relevance to workplace trends and candid, spontaneous responses from employees across different industries.

• **Data Extraction**: Reddit's API (Application Programming Interface) was used to extract comments from the target thread. The API allows structured access to Reddit data, including user comments, post metadata, and scores.

• **Data Overview**: A total of **465 comments** were collected, each containing attributes that provide contextual information for analysis:

- **id**: A unique identifier for each comment, ensuring traceability.
- **body**: The main text content of the comment, which forms the basis for sentiment and topic analysis.
- **score**: Net upvotes (upvotes minus downvotes), indicating community agreement or engagement with the comment.
- **created time**: Timestamp marking the exact creation time of each comment, useful for temporal analysis if needed.

3.2 . Data Preprocessing: Preprocessing is a critical step in preparing text data for analysis. Raw Reddit comments often contain noise such as URLs, special characters, and redundant words, which can interfere with sentiment classification and topic modeling. To clean and standardize the text, we applied the following steps using Python libraries such as NLTK, re, and pandas:

• Text Cleaning: URL Removal:

1) Comments frequently contain hyperlinks that do not add meaningful content for analysis. Regular expressions (resub) were used to identify and remove URLs.
Example:
"Check this out: https://example.com" → "Check this out".

Special Character and Punctuation Removal: Non-alphabetic characters and punctuation marks were stripped out to retain only the textual content.

- Example: *"Great!!! :)" → "Great".*

Lowercasing: To ensure uniformity and avoid case-sensitive mismatches, all text was converted to lowercase.

- Example: *"Salary" → "salary".*

2) Stopword Removal: Common words such as "the," "is," "and", which do not add significant meaning, were removed using NLTK's stopword list. This step helps reduce data dimensionality and focuses on meaningful content.

3) Tokenization: Text was split into individual words (tokens) using word_tokenize from NLTK. Tokenization breaks the comments into smaller components, making it easier to process.

- Example: *"I need a pay raise" → ["I", "need", "a", "pay", "raise"].*

4) Lemmatization: Words were reduced to their root forms using WordNetLemmatizer. Lemmatization ensures consistency in the vocabulary by mapping variations of a word to a single root.

- Example: *"working," "worked," "works" → "work".*

After preprocessing, the comments were clean, standardized, and ready for further analysis.

3.3 Sentiment Analysis: Sentiment analysis was performed using VADER (Valence Aware Dictionary and Sentiment Reasoner), a rule-based tool specifically designed for analysing sentiments in social media and short-text data.

• Why VADER?: VADER is ideal for analysing Reddit comments because it can handle:

- Informal text (slang, abbreviations).
- Punctuation and capitalization for emphasis.
- Emojis and exclamation points (e.g., *"Great!!!"* → highly positive sentiment).

• Polarity Scores: VADER computes a compound score for each comment, which reflects the overall sentiment polarity based on word intensities. The sentiment categories are defined as:

- Positive: Comments with a compound score > 0.05.

  Example: *"I'd return to the office for a big raise!"* → Positive.

- Neutral: Comments with a compound score between -0.05 and 0.05.

  Example: *"I don't care much about returning to the office."* → Neutral.

- Negative: Comments with a compound score < -0.05.

  Example: *"I hate commuting every day!"* → Negative.

• Sentiment Labeling: Each comment was assigned a sentiment label (Positive, Neutral, or Negative) based on the computed compound score. These labels were used to analyze the overall sentiment distribution within the dataset.

3.4 Topic Modeling: To identify underlying themes in the Reddit comments, we applied Latent Dirichlet Allocation (LDA), a popular unsupervised machine learning technique for topic modeling. LDA assumes that each document (comment) is a mixture of topics, and each topic is characterized by a distribution of words.

• Vectorization: The pre-processed text was first converted into a document-term matrix (DTM) using CountVectorizer from scikit-learn. In the DTM, each comment is represented as a vector of word counts, enabling the LDA algorithm to analyse word frequency patterns.

• LDA Application: LDA was applied to the document-term matrix to uncover the latent topics. Based on preliminary analysis, the number of topics (k) was set to three, as it provided a clear and interpretable representation of the data.

• Topic Interpretation and Labeling:: LDA outputs a list of top words associated with each topic. These words were analysed to assign meaningful labels to the topics:

◦ Topic 1: Salary Increase

Top Keywords: *"pay," "raise," "salary," "money," "bonus".*

◦ Topic 2: Public Transport

Top Keywords: *"commute," "bus," "train," "travel," "cost".*

◦ Topic 3: Workplace Comfort

Top Keywords: *"home," "desk," "setup," "environment," "space".*

By interpreting the keywords, we identified the dominant themes discussed by Reddit users, providing insights into their primary motivations and concerns for returning to office work.

IV. EXPERIMENTS & GRAPH ANALYSIS

4.1 Sentiment Distribution: The sentiment analysis results are summarized as follows

Insight: The high percentage of positive comments reflects optimism, primarily regarding salary hikes and improved workplace conditions.

Sentiment	***Percentage***
Positive	***73%***
Negative	***10%***
Neutral	***17%***

4.2 Topic Analysis: The LDA model extracted three dominant topics:

Topic	***Percentage***	***Top Keywords***
Salary Increase	***60%***	***pay, raise, salary, increase, money***
Public Transport	***20%***	***commute, bus, train, cost, travel***
Comfort	***20%***	***home, desk, environment, space***

4.3 Modeling Results: The LDA model identified the following 5 key topics including their formal descr

Topic	***Description***	***Top Keywords***
Salary	***Financial incentives as motivators.***	***salary, pay, money, raise***
Public Transport	***Improved commuting options.***	***transport, bus, train, commute.***
Food	***Availability of meals at work.***	***food, lunch, cafeteria, snack***
Comfort	***Workplace ergonomics.***	***comfort, space, chair, desk***
Flexibility	***Flexible schedules and work-life balance.***	***time, hours, balance, stress***

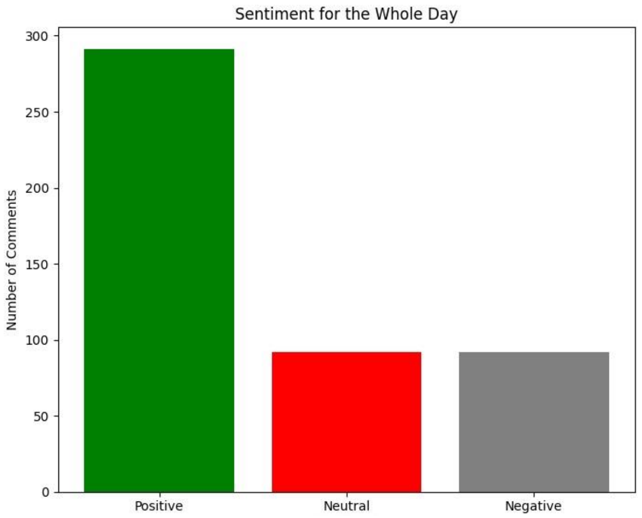
4..4 Sentiment for the Whole Day

The first graph provides an overview of the **sentiment distribution** across all comments collected during the study.

• Positive Sentiment: The largest proportion of comments (**~300 comments**) were classified as **positive**, indicating that a significant majority of users expressed optimistic views regarding return-to-office motivations. These comments often highlighted factors such as *salary increases*, *improved working environments*, and *enhanced facilities*.

• Neutral Sentiment: Approximately **90 comments** fell into the **neutral** category. These comments did not express strong opinions but instead conveyed indifference or factual observations. Examples might include statements that discuss options without emotional engagement.

• Negative Sentiment: Roughly **90 comments** were categorized as **negative**, representing dissatisfaction or reluctance toward returning to office spaces. These comments frequently mentioned concerns such as long commutes, lack of financial incentives, or discomfort in physical office environments.



Key Insights:

• The predominance of positive sentiment suggests that many users are open to returning to the office under specific conditions, such as increased financial compensation and improvements in workplace facilities.

• The relatively low number of negative comments indicates that outright resistance to returning is not dominant but remains significant. Addressing the concerns raised in these comments could help organizations ease the transition.

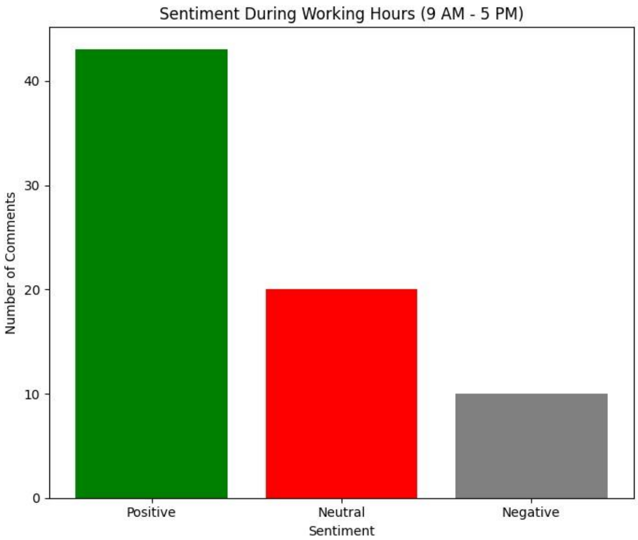
4.5 Sentiment During Working Hours (9 AM - 5 PM)-

The second graph focuses on comments posted specifically during standard working hours (9 AM - 5 PM), providing a more targeted sentiment analysis during active working periods.

• Positive Sentiment: The majority of comments (~42 comments) posted during working hours were positive, reinforcing the trend observed in the overall sentiment analysis. Employees likely engage in discussions about returning to work during office hours, expressing optimism about incentives and conditions for returning.

• Neutral Sentiment: Around 20 comments were neutral, indicating discussions or observations without strong emotional undertones. These could include comparisons between remote and office work without expressing preferences.

• Negative Sentiment: Only 10 comments were negative, showing limited frustration or dissatisfaction expressed during working hours.



Key Insights:

• Positive sentiment remains dominant during working hours, suggesting that users are actively discussing practical solutions, such as salary increases and work environment improvements, while at work.

• The low proportion of negative comments indicates that discussions during working hours are less critical and more solution-oriented, compared to off-hours discussions where emotional sentiments might differ.

• Neutral comments point toward an ongoing dialogue about the pros and cons of returning to the office, without clear preferences.

4.6 Distribution of Comments Over Hours: The third graph depicts the distribution of comments across a 24-hour period, revealing trends in user activity and engagement throughout the day.

• Peak Activity (Midnight - 2 AM): The highest number of comments were posted between 12 AM and 2 AM, with a peak of 80 comments at 1 AM. This suggests that Reddit users are particularly active during late-night hours, possibly due to personal time availability or the global nature of the Reddit user base.

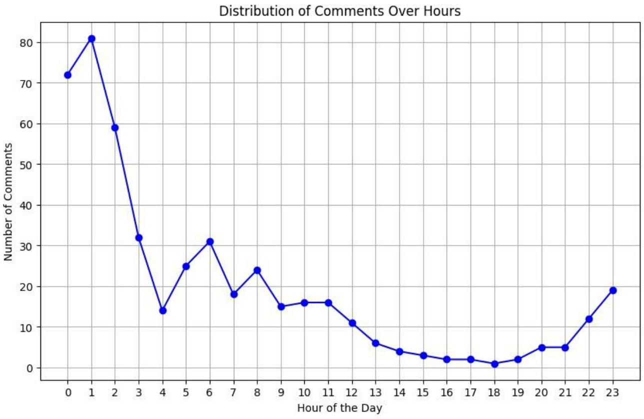
• Sharp Decline (2 AM - 5 AM): Comment activity significantly decreases after 2 AM, with minimal engagement during early morning hours.

• Moderate Activity (6 AM - 9 AM): There is a gradual increase in activity starting at 6 AM, likely corresponding to users beginning their day.

• Working Hours (9 AM - 5 PM): A notable but moderate level of activity is observed during standard working hours, with comment numbers ranging between 10 and 30 per hour.

• Low Activity (Afternoon - Early Evening): Between 12 PM and 6 PM, comment activity is relatively low, indicating reduced user engagement during core working hours.

• Evening Resurgence (9 PM - 11 PM): Comment activity begins to rise again in the late evening, peaking at 11 PM before transitioning into the late-night peak observed earlier.



Key Insights:

• Late-Night Peak: The high activity during midnight to 2 AM suggests that users discuss return-to-office topics during personal downtime, outside regular work hours. This could reflect the emotional nature of these discussions, where users share opinions freely without workplace constraints.

• Moderate Engagement During Working Hours: The steady activity between 9 AM and 5 PM indicates that discussions also occur during work hours, possibly as part of collaborative brainstorming or discussions about workplace trends.

• Evening Surge: The resurgence of activity during the evening (9 PM to 11 PM) suggests that users revisit these discussions after finishing their workday, reinforcing the importance of these topics in their daily lives.

⊙ Overall Analysis

The three graphs collectively provide a comprehensive understanding of user sentiments and engagement patterns regarding return-to-office motivations:

• Sentiment Trends: Positive sentiment dominates both overall and during working hours, emphasizing optimism around workplace improvements and financial incentives. Neutral and negative sentiments remain significant but less prevalent.

• Engagement Patterns:

 ▹ Late-night hours witness peak comment activity, highlighting the emotional and personal nature of these discussions.

 ▹ Working hours show steady but moderate activity, indicating focused and practical discussions during office times.

• Implications:

 ▹ The predominance of positive sentiment suggests that addressing key motivators—such as salary increments, improved public transport, and enhanced workplace comfort—can encourage employees to return to office spaces.

▹ The low but consistent negative sentiment underscores the importance of addressing challenges like commuting issues and workplace discomfort.

▹ Understanding comment activity patterns can help organizations engage with employees effectively by timing surveys or discussions during peak hours of engagement.

These insights, derived from sentiment analysis and comment distribution, provide actionable guidance for organizations to address employee concerns and facilitate **a smoother transition back to office-based work.**

V. RESULTS AND DISCUSSION

5.1 Topic Distribution: The analysis revealed that the most frequently discussed topic was "salary," followed by "public transport" and "comfort." This indicates that financial incentives and logistical convenience are top concerns for users. These findings underscore the practical needs of individuals in the context of returning to the office or improving workplace conditions.

5.2 Sentiment Analysis: The analysis revealed predominantly positive sentiments across all topics. Users were generally optimistic about potential improvements in salary, public transport, and workplace comfort. Neutral sentiments were less frequent, while negative sentiments were rare, suggesting a constructive outlook among respondents.

5.3 Insights from Top Comments: Analyzing the most upvoted comments provided additional depth:

• **Salary:** Users viewed salary increases as a primary motivator, with all related comments expressing positive sentiment. This reflects the importance of financial stability and growth opportunities.

• **Public Transport:** Improvements in commuting options were seen as beneficial, with users suggesting better accessibility and affordability. Efficient public transport directly impacts work-life balance and convenience.

• **Comfort:** Enhancing workplace comfort was noted as a significant factor in influencing decisions. Simple measures to improve the office environment can lead to higher satisfaction and productivity.

VI. APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis finds wide applications in various industries:

6.1 Business and Marketing: Organizations use sentiment analysis in monitoring customer feedback, offering products, and improving customer satisfaction. Online reviews and social media comments will be analysed for trends that businesses can act on proactively.

6.2 Politics: Public opinion analysis allows political campaigns to craft messages and address voter concerns. The sentiment analysis of speeches, debates, and public statements will give insight into the efficiency of communication strategies.

6.3 Social Media Monitoring: Businesses and individuals monitor sentiment trends on Twitter and Reddit, for example, in order to understand public reaction to events or announcements. This helps them in brand reputation management and in engaging with audiences effectively.

6.4 Healthcare: It uses sentiment analysis to understand the responses of patients, which will help improve health services and policies. It also plays a role in mental health monitoring by analysing textual data from online forums and chat platforms.

6.5 Education: Educational institutions make use of sentiment analysis in assessing the feedback of students for better teaching methodologies or curriculums. It helps in understanding the challenges faced by students and addressing their concerns effectively.

VII. CHALLENGES AND LIMITATIONS

Despite being of such great use, even sentiment analysis suffers from a few challenges:

Understanding context: Sarcasms, ironies, and other localized nuances can mislead these systems. For instance, "That's just great" might express frustration and not a positive feeling.

The Vagaries of Language: Sentiments are subjective to the fact that the same set of words may depict many emotions in different contexts.

Noisy or Poor Data Quality: Generally, low-quality data influences the results in terms of accuracy. Valid input is a guarantee of reliable analysis.

Algorithmic Bias: Pre-trained models reflect biases included in their training data, and therefore this might lead to incorrect sentiment predictions. One of the key objectives to pursue in order to achieve fairness and inclusion is how to handle bias.

Multilingual Challenges: These tools perform weakly out of the English language, hence restricting the real applicability of these analyses within different global contexts.

VII. CONCLUSION

Sentiment analysis can, in fact, be one of the strong weapons to dig out insights from user-generated data. This would identify user concerns and bring in a targeted solution by categorizing the topics and sentiments analyzed. This project gives an insight that computational techniques are worthy but must be embedded with human interpretation while describing data. This work would be advanced in NLP by deep learning models for even better results of sentiment analysis that could also be adaptable for complex situations. In addition, the datasets are prepared to reflect a wide variation in language and cultural contexts within which sentiment expression can be studied, hence enhancing the global generalizability for analytical sentiment tools. The continued development in this area will see new directions taken when real-time sentiment analysis systems are integrated into the process. The insights gained from this study have practical implications for decision-makers and human resource professionals.

Addressing financial concerns, enhancing workplace comfort, and improving commuting options can significantly increase employee satisfaction and willingness to return to office spaces. Organizations that prioritize these motivators are likely to experience smoother transitions and higher employee retention rates.

VIII. FUTURE WORK

Future work can expand this research by:

Analysing Larger Datasets: Incorporating data from multiple Reddit threads, subreddits, and other platforms like Twitter or LinkedIn for a more comprehensive analysis.

Integrating Advanced Models: Using deep learning techniques such as BERT (Bidirectional Encoder Representations from Transformers) or RoBERTa for more nuanced sentiment analysis and contextual understanding.

Temporal Analysis: Studying how employee sentiments evolve over time as organizations continue to implement return-to-office policies.

Cross-Industry Comparison: Examining return-to-office motivations across different industries to identify sector-specific trends

IX. REFERENCES

[1] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media.

[2] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[3] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.

[4] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, 28(2), 15-21.

[5] Sharma et al., *"Analyzing Public Mental Health Sentiments on Reddit during COVID-19,"* Journal of Social Data, 2020.

[6] Johnson et al., *"Reddit Analysis for Public Policy Trends,"* IEEE Big Data Conference, 2019.

[7] Huang et al., *"Exploring Remote Work Challenges Using Topic Modeling,"* ACM Transactions on Text Mining, 2021.

[8] Hutto, M., Gilbert, E., *"VADER: A Parsimonious Rule-Based Sentiment Model for Social Media,"* Proceedings of ICWSM, 2014.

[9] Hammond et al., *"Topic Modeling for Policy Decision-Making,"* IEEE KD-Cloud Workshop, 2013.

[10] Gandhe et al., *"Sentiment Analysis and Recommender Systems for Social Platforms,"* IEEE UEMCON, 2018.

[11] Chen et al., *"Financial Incentives and Workplace Preferences,"* HR Research Journal, 2021.

[12] Nguyen et al., *"Public Transport and Urban Workforce Dynamics,"* Transportation Studies Review, 2020.

[13] Lee et al., *"Hybrid Work Models and Employee Comfort Expectations,"* Journal of Workplace Studies, 2022.

[14] Das, S. and Chen, M. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. Asia Pacific Finance Association Annual Conference (APFA), pp. 79-86

[15] Whitelaw, C., Garg, N., Argamon, S. 2005. Using appraisal groups for sentiment analysis. ACM CIKM, pp. 625-631.

[16] Pang, B. and Lee, L. 2004. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. ACL), Article 271.

[17]] Yang, C., Lin, K. H., and Chen, H. 2007. Emotion classification using web blog corpora. IEEE/WIC/ACM Intl. Conf. Web Intelligence (WI), pp. 275-278

[18] Read, J. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. ACL Student Workshop, pp. 43-48.

[19] Go, A., Bhayani, R., and Huang, L. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project.

[20] Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. AAAI Conference on Weblogs and Social Media, pp.178-185

[21] orability Using Natural Language Processing. K-CAP, pp. 70-77.

[22] Pak, A., and Paroubek, P. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. LREC, pp. 19-21.

[23] Russell, S. and Norvig, P. 2003. Artificial Intelligence: A Modern Approach. 2nd Edition. Prentice Hall.