Volume 7 Issue 1 (2024) Pages 186 - 197
YUME : Journal of Management
ISSN : 2614-851X (Online)

Sentiment Analysis of The Most Viewed YouTube Video: Exploring Gender Bias in The Discussion of Women Workers in Indonesia

Rouna Meisda Paoki ✉[1], Jimmy Moedjahedy[2]
[1] Akademi Sekretari Manajemen Indonesia Klabat, Indonesia
[2] Universitas Klabat, Airmadidi, Indonesia

Abstrak

Isu-isu kekerasan terhadap perempuan dan kelompok rentan, serta penerapan pengarusutamaan gender dan pemberdayaan perempuan, terus menjadi masalah bagi bisnis dan organisasi. Tujuan utama dari penelitian ini adalah untuk menyelidiki sentimen yang diekspresikan oleh netizen atau individu di Indonesia terkait dengan Bias Gender di negara ini, dengan menggunakan lima film yang paling sering dilihat di platform YouTube. Metodologi yang digunakan dalam penelitian ini adalah dengan melakukan analisis sentimen terhadap komentar-komentar yang terkait dengan video-video tersebut, seperti yang telah disebutkan sebelumnya. Komentar-komentar tersebut dianalisis menggunakan Google API, Pandas for Python, dan Natural Language Tool Kit (NLTK). Temuan dari penelitian ini menunjukkan bahwa distribusi balasan dapat dikategorikan sebagai berikut: 53,25% dari tanggapan diklasifikasikan sebagai netral, 31,76% diklasifikasikan sebagai positif, dan 14,99% diklasifikasikan sebagai negatif. Salah satu pendekatan yang memungkinkan untuk penelitian ini adalah dengan menggunakan analisis sentimen dengan menyertakan dan membandingkan beberapa teknik pembelajaran mesin.

Kata Kunci: *Bias Gender; Analisis Sentimen; Pekerja Wanita; Gender Inequality Index*

Abstract

Issues of violence against women and vulnerable groups, as well as the adoption of gender mainstreaming and women's empowerment, continue to be problems for businesses and organizations. The main objective of this study is to investigate the sentiments expressed by netizens or individuals in Indonesia with regard to Gender Bias in the country, utilizing the five most frequently seen films on the YouTube platform. The methodology employed in this study involves conducting sentiment analysis on the comments associated with the videos, as mentioned earlier. The comments are analyzed using Google API, Pandas for Python, and the Natural Language Tool Kit (NLTK). The findings of this study indicate that the distribution of replies can be categorized as follows: 53.25% of the responses were classified as neutral, 31.76% were classified as positive, and 14.99% were classified as negative. One possible approach for this research involves the utilization of sentiment analysis by including and contrasting multiple machine-learning techniques.

Keywords: *Gender Bias; Sentiment Analysis; Woman Worker; Gender Inequality Index*

✉ Corresponding author :
Email Address : rouna.paoki@unklab.ac.id

INTRODUCTION

The data pertaining to women's employment in professional roles in Indonesia has exhibited a consistent upward trend over the past six years. Based on data provided by the Central Bureau of Statistics (BPS) in Indonesia, there is a nearly equivalent representation of both genders in professional occupations on an annual basis. According to the available data spanning from 2016 to 2022, the average proportion of men in professional roles is 51.96%, whereas the corresponding figure for women stands at 48.03%. According to the release data of the research, as mentioned earlier, there was an observed growth in the proportion of women serving in parliament, rising from around 17% to 21.89% by the year 2023 (Badan Pusat Statistik, 2023).

Women in professional roles and women serving in parliament represent some of the variables that impact the Gender Inequality Index (GII) in Indonesia. The GII serves as a metric for assessing the extent of gender-based disparities in various dimensions of human development. These disparities, particularly in the areas of reproductive health, empowerment, and the labor market, hinder the attainment of optimal human development achievements. Optimal human development is contingent upon minimizing the disparity in success among individuals, particularly regarding gender. To ascertain the presence of GII necessitates the utilization of a comprehensive metric that accurately captures the multifaceted nature of inequality while also offering insights into potential remedies. A lower GII score indicates a reduced level of disparity between genders. Figure 1 illustrates a declining trend in the GII score in Indonesia throughout the years.

Typically, employed women experience three distinct social contexts: the familial sphere, the community at large, and the professional setting. Informal social connections have a significant role in shaping power dynamics within the family. In contrast, the professional environment is defined by a structured framework of formal vertical and horizontal relationships, mainly focused on work-related tasks (Ortiz Rodríguez & Pillai, 2019). However, women frequently encounter inequitable treatment in the workplace, sometimes referred to as gender discrimination. Numerous research has provided evidence supporting the existence of gender disparities inside the workplace.

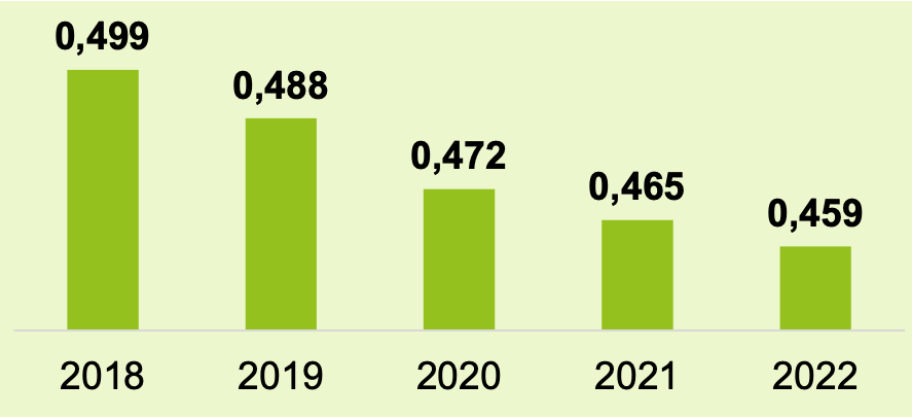

Figure 1. Indonesia GII score 2018-2022

Gender differences in workplaces are commonly complex, driven by a combination of institutional systems, cultural norms, and individual prejudices. Research undertaken by reputable organizations like Catalyst has highlighted the enduring presence of gender-based discrimination in professional environments globally (Catalyst, 2023). These types of

discrimination are evident in differential compensation, restricted opportunities for leadership positions, biased assessments of performance, and insufficient backing for efforts promoting work-life balance among women. Moreover, the challenges experienced by women in professional sectors are further intensified by intersectional characteristics such as race, ethnicity, and socio-economic background, which magnify the impediments to their advancement (Quinlan & VanderBrug, 2016).

To tackle these intricate difficulties, it is necessary to implement comprehensive methods that involve changes in legislation, regulations within organizations, and transformations in public views regarding gender roles. Enacting legislative measures to provide equal compensation for equivalent work, offering parental leave, and enforcing anti-discrimination legislation are essential actions in mitigating gender inequalities within the labor market (Huang, 2019). Furthermore, organizations have a crucial role in cultivating inclusive work environments through promoting diversity in leadership, providing mentorship programs, and implementing flexible work arrangements that cater to caregiving obligations. Research conducted by Dixon-Fyle et al. (2020) has demonstrated that companies that place a high value on diversity and inclusion tend to have better financial results and more innovation. This underscores the compelling argument for promoting gender equality in professional environments.

Moreover, the participation of males as advocates for gender equality in professional environments is essential. Encouraging men to actively participate in discussions about gender prejudices, urging them to question and reject established gender conventions, and advocating for allyship programs can play a big role in breaking down obstacles and cultivating a more inclusive work environment. At a broader societal level, it is crucial to implement educational programs that challenge gender stereotypes and promote equitable opportunities for all genders, starting from a young age. These programs are necessary to address the societal changes that are needed (Loayza & Trumbic, 2021). To mitigate these complex issues, it is essential to tackle them at different levels, including legislative, organizational, and societal. This approach will enable the creation of conducive settings that empower women to excel and make meaningful contributions in their professional capacities, ultimately reducing gender imbalance within Indonesia's workforce.

LITERATURE REVIEW

In their study, Njuki et al. (2022) conducted a comprehensive evaluation of 198 papers that satisfied specific inclusion criteria, including publication in English, origination from low- or middle-income countries, and undergoing peer review. The findings of the analysis revealed that there exist disparities in resource accessibility within the food system between women and men. These disparities encompass various aspects, including but not limited to essential services, knowledge and information, technology deployment, time allocation, and market access.

A study on the subject of female workforce participation in Mexico, specifically focusing on gender equality, was carried out by Ortiz Rodríguez & Pillai (2019). The data utilized in this study comprises the national statistics data of Mexico. The data utilized in this study comprises the national statistics data of Mexico. The sample for this study consisted of 82,846 women who were married or cohabiting and resided in both rural and urban regions of Mexico. The age range of the participants was 15 years and above. Research indicates that there is a tendency for women residing in urban regions to exhibit a higher

level of support for gender equality compared to their counterparts residing in rural areas. Younger females exhibit a greater propensity for endorsing gender equality compared to their older counterparts. The solution that has been identified is enhancing the educational attainment of women and effectively implementing labor regulations, which have the potential to mitigate gender disparities and enhance the well-being of women in both working and non-working roles.

In addition, a research investigation was carried out in Indonesia by Siscawati et al. (2020). Organizations and enterprises continue to face challenges pertaining to the implementation of gender mainstreaming and women's empowerment, as well as addressing issues of violence against women and vulnerable populations. Specifically, contemporary challenges such as intolerance, extremism, and terrorism possess a gendered dimension, undermining endeavors to achieve gender equality, empower women, advance human development, and foster national progress holistically.

In contrast, in the research undertaken by Kurzman et al. (2019), the evidence presented indicates the existence of a scholarly discourse surrounding the conceptualization and understanding of gender equality. The discourse around the establishment of universal definitions of gender equality vs. subjective definitions arose throughout the latter part of the 18th century, coinciding with the initial articulations of women's rights in Western Europe. To see how well women's experiences and views match up with the globally agreed-upon meaning used by the global gender equality index, a comprehensive four-stage testing process was undertaken in conjunction with the gender equality index. The findings of the study indicated that there was no consistent correlation between the equality index and women's perceptions of control over their lives or overall life satisfaction in comparison to men's perceptions.

Based on exposed data and ongoing discussions, the researchers undertook a sentiment analysis of comments posted on YouTube videos pertaining to the prevalent issue of gender bias. Sentiment analysis techniques are frequently employed to assess the reactions to an occurrence, discerning whether they are favorable, negative, or neutral. There are multiple research that serve as references, As shown in Table 1.

Table 1. Research That Employs Sentiment Analysis on YouTube

No.	Author	Title	Result
1	(Pichad, 2023)	Analysing Sentiments for YouTube Comments Using Machine Learning	Texts with sentiments that vary are hard for people to label when it comes to text sentiment classification. A new labeling technique is used in this study to separate texts into those with pure sentiment orientations and those with mixed sentiment orientations.

2	(Thelwall, 2018)	Analysing Sentiments for YouTube Comments Using Machine Learning	Detecting male sentiment poses a greater challenge due to its relatively lower level of explicitness. There is a lack of empirical data to support the notion that gender-specific lexical sentiment analysis can effectively address this problem.
3	(Singh & Tiwari, 2021)	YouTube Comments Sentiment Analysis	The utilization of uni-grams, bi-grams, and tri-gram features exhibited strong performance and provided substantial support to the classifiers, resulting in the attainment of the highest accuracy ratings.
4	(Novendri et al., 2020)	Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes	Based on the findings of sentiment analysis, it is evident that a significant proportion of comments express positive sentiments, suggesting a high level of audience desire for the picture.
5	(Tanesab et al., 2017)	Sentiment Analysis Model Based On YouTube Comment Using Support Vector Machine	Sentiment analysis can be employed to assess the extent to which the performance of an object is influenced by the feedback received from users on YouTube. The classification of weighted data using the Support Vector Machine (SVM) approach has led to the determination that the True Positive rate is 91.1%.

METHOD

The methodology employed in this study involves the utilization of sentiment analysis. It is worth noting that various other studies have also employed sentiment analysis, not limited to the analysis of YouTube data. These studies are documented in Table 2 below. The focus of this study pertains to a film that explores the subject of gender bias, specifically examining female workers who have garnered the most number of views. According to the information presented in Table 3, a total of five videos were examined,

accumulating a combined count of 2,634,254 video views and 6,868 comments. The research employs the Python programming language, executed within the Jupyter Notebook environment using Anaconda. Research procedures are depicted in Figure 2.

Table 2. Research That Employs Sentiment Analysis on General

No.	Author	Title	Result
1	(D'Aniello et al., 2022)	KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis	A novel sentiment analysis model, referred to as ABSA (aspect-based Sentiment Analysis), has been developed. The underlying premise of the model is based on the recognition that sentiment, affect, emotion, and opinion are distinct concepts, and it is fundamentally flawed to employ the same metrics and techniques for their measurement.
2	(Denecke & Reichenpfader, 2023)	Sentiment analysis of clinical narratives: A scoping review	The research conducted by the authors demonstrates that the utilization of sentiment analysis leads to enhanced predictive capabilities in relation to clinical outcomes, specifically death rates and the likelihood of readmission.
3	(Cristescu et al., 2022)	Using Market News Sentiment Analysis for Stock Market Prediction	The findings demonstrate that the incorporation of sentiment ratings as an exogenous variable in linear autoregression enhances the coefficient R's value. Therefore, it can be inferred that the incorporation of sentiment analysis into market regression analysis leads to improved regression outcomes in terms of goodness of fit.
4	(Novák et al., 2021)	Sentiment Analysis in Agriculture	In the field of agriculture, sentiment analysis is predominantly carried out using machine learning techniques, with the naive Bayes algorithm being the prevailing method of choice. The utilization of lexicon-based techniques is

			notably limited, whereas hybrid approaches have yet to be employed.
5	(Rahab et al., 2021)	SANA: Sentiment analysis on newspaper comments in Algeria	The researchers utilized two distinct text collections for their study: SANA, a proprietary dataset generated by the researchers themselves, and OCA, a publically accessible dataset. The findings of the study indicated that Support Vector Machine, Naive Bayes, and K-Nearest Neighbour exhibited promising performance; nonetheless, additional development is necessary.

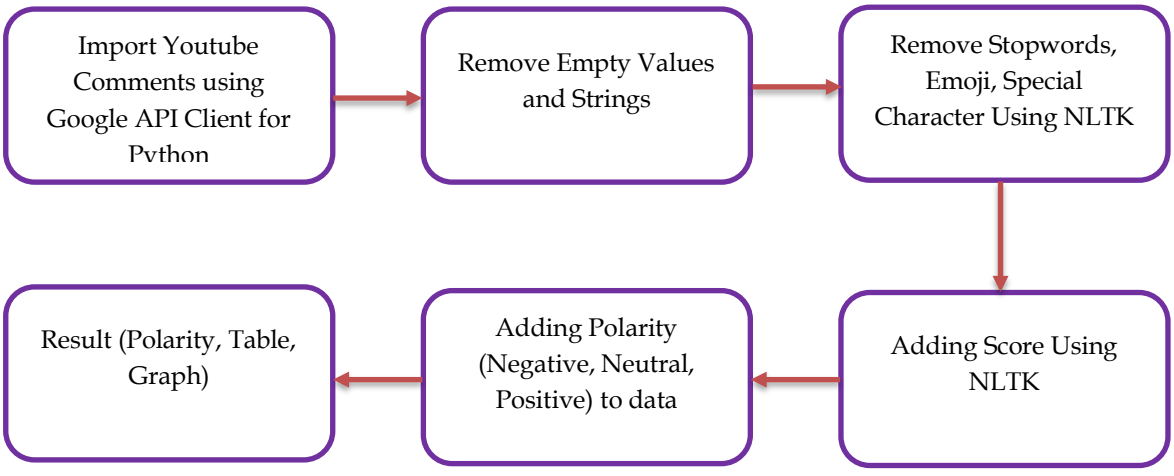

Figure 2. Research Process

This research utilizes three specific tools and libraries, which are:

1. Google API Client for Python: Access to a wide variety of Google products, including Google Maps, Google Drive, and YouTube, can be gained using Google APIs. Google makes available client libraries that, in addition to making it simpler to build code that interacts with these application programming interfaces (APIs), can also make the code you write more reliable (Google, n.d.).

2. Pandas for Python: When dealing with tabular data, such as that which is kept in databases or spreadsheets, pandas is the tool that should be used to address problems. The data will be explored, cleaned, and processed with the assistance of pandas (Pandas, n.d.).
3. Natural Language Toolkit (NLTK) Vader: The process entails the utilization of a predetermined set of rules and heuristics to ascertain the sentiment expressed in a given textual content. The criteria for these principles generally rely on the lexical and syntactic characteristics of the text, encompassing factors such as the inclusion of positive or negative vocabulary and expressions.

Table 3. Most Viewed YouTube Video About Woman Worker

Youtube ID	Total Views	Video Title (Translated)	Total Comments
93FiM3tWT0g	976.906	From Women for Women	901
_7juLPcjp5g	330.407	Why are there so few female leaders?	2.096
BSr2DECCYxs	219.216	We need feminism because women are still victims of sexism	2.512
cKHIzKNheXY	94.550	Smart Women: How Najwa Shihab Addresses the Gender Gap	43
ctjfkk7DyGA	1.013.175	Why Women Should Vote	1.317

RESULT AND DISCUSSION

After the extraction of comments from the five videos, the obtained comment results undergo further processing through the following steps:

1. The process of eliminating empty values and strings from a dataset.
2. The process of eliminating stopwords, emojis, and special characters is facilitated through the utilisation of the Natural Language Toolkit (NLTK).
3. Assigning a numerical rating to each comment using the Natural Language Toolkit (NLTK). The second column in table 3 displays the relevant information.
4. One approach to enhance comments is by incorporating polarity through the inclusion of negative, positive, and neutral values. The fifth column in table 3 exhibits the relevant information.
5. Obtain the outcomes in the format of ultimate polarity, accompanied by graphical representations as depicted in Figures 3, 4, and 5.

Table 4. Fragment of Sentiment Analysis Using YouTube Comment

	comments	scores	compound	polarity
CARISSA RIFAYA SALSABILA	bang perempuan yang ngomong itu kak fergi jnx ya karna dari baju suara dan bentuk rambut aja udah sama ...	{'neg': 0.0, 'neu': 0.933, 'pos': 0.067, 'compound': 0.3612}	0.3612	positive
MAT	ya semoga ga cuma mau setara dalam hal yang senang nya aja semoga mau setara di bagian susahnya juga	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neutral
...	…	**...**	**...**	**...**
Yandi Anjimy	btw ini pertanyaan ku dri tahun tdk pernah terjawab ^ pertanyaannya simple saja mana sih …	{'neg': 0.0, 'neu': 0.963, 'pos': 0.037, 'compound': 0.3612}	0.3612	neutral

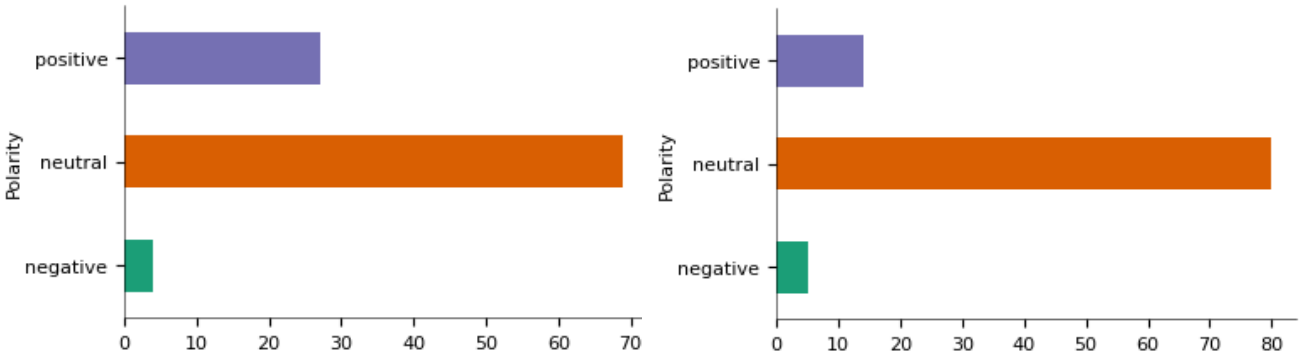

Figure 3. Polarity For Video 1 And 2

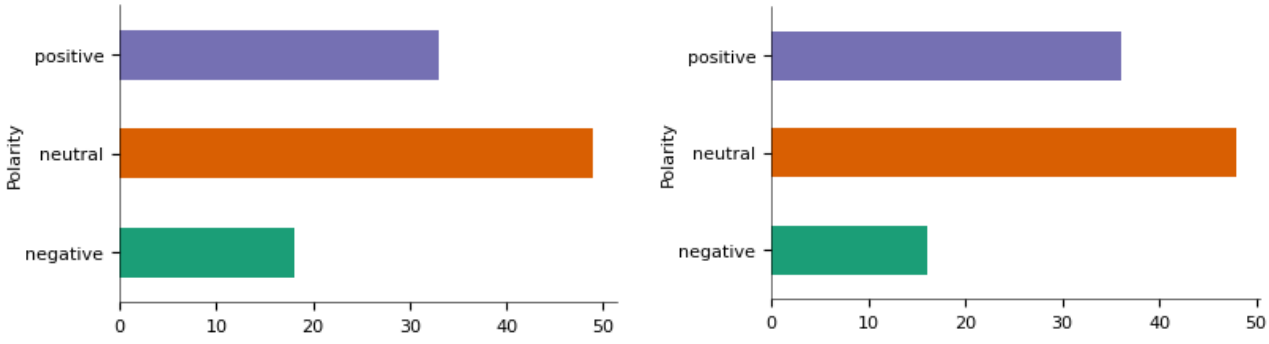

Figure 4. Polarity For Video 3 And 4

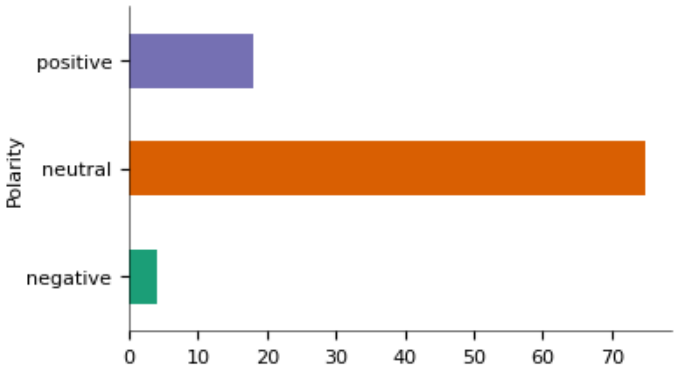

Figure 5. Polarity For Video 5

Table 5. Comments that have polarity

Youtube ID	Video Title (Translated)	Positive	Neutral	Negative
93FiM3tWT0g	From Women for Women	34	**49**	19
_7juLPcjp5g	Why are there so few female leaders?	38	**49**	18
BSr2DECCYxs	We need feminism because women are still victims of sexism	33	**50**	18
cKHIzKNheXY	Smart Women: How Najwa Shihab Addresses the Gender Gap	37	**47**	16
ctjfkk7DyGA	Why Women Should Vote	19	**75**	5

According to the findings presented in Table 5, the sentiment analysis outcomes indicate that the attitudes expressed by internet users and the general public towards women in the workforce are characterized as neutral. This finding aligns with the research conducted by Lee et al. (2022) which demonstrates a positive correlation between the inclusion of women in a team and their overall collective performance. Similar findings were considered as well by Izzaty Shahirah Nor Sham et al. (2021), the inclusion of women in the workforce has the potential to contribute to the advancement of gender equality and the empowerment of women inside organizations, and women are encouraged to assume a fair and equitable share of work responsibilities (Wani & Ahmad, 2021). This can be achieved by enhancing the capabilities of working women, while also prioritizing the establishment of a more favorable equilibrium between their professional and personal lives. The objective is to enhance the capabilities of women by promoting their potential, while concurrently striving to establish a more favorable equilibrium between their

professional and personal lives. The polarity results produced in this study will be used in the examination of accuracy comparison utilizing machine learning in further investigations. Furthermore, it is important to conduct compound analysis on every comment in order to ascertain the sentiment rate. The sentiment analysis findings of this study are presented in the Conclusion section.

CONCLUSION

According to the findings that were presented in the section titled "Results and Discussion," it is possible to see that out of the total of 507 comments that were analyzed to determine their polarity, 161 comments were categorized as positive, 270 comments were defined as neutral, and 76 comments were labelled as negative. According to the data, it is possible to draw the conclusion that the general attitude of internet users or the general public towards women who are employed in office settings is generally neutral. This sentiment accounts for 53.25% of all of the responses. Nevertheless, a sizeable proportion, around 31.76% of those surveyed, had a favorable reaction. On the other hand, just 14.99% of respondents expressed an unfavorable attitude towards the issue being discussed here. This represents a minority. One of the possible directions that this research could go is in the direction of using sentiment analysis by combining and contrasting a number of different machine-learning strategies.

References :

Badan Pusat Statistik. (2023). Berita Resmi Statistik 1 Agustus 2023. https://www.bps.go.id/website/materi_ind/materiBrsInd-20230801155718_rev.pdf

Catalyst. (2023, December). Catalyst Transforms Workplaces.

Cristescu, M. P., Nerisanu, R. A., Mara, D. A., & Oprea, S. V. (2022). Using Market News Sentiment Analysis for Stock Market Prediction. Mathematics, 10(22). https://doi.org/10.3390/math10224255

D'Aniello, G., Gaeta, M., & La Rocca, I. (2022). KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artificial Intelligence Review, 55(7). https://doi.org/10.1007/s10462-021-10134-9

Denecke, K., & Reichenpfader, D. (2023). Sentiment analysis of clinical narratives: A scoping review. Journal of Biomedical Informatics, 140. https://doi.org/10.1016/j.jbi.2023.104336

Dixon-Fyle, S., Hunt, V. (DBE), Dolan, K., & Prince, S. (2020). Diversity wins: How inclusion matters. McKinsey, May.

Huang, L. (2019). Modern Families: An Interview with UN Women on Progress of the World's Women 2019–2020: Families in a Changing World. Chicago Policy Review (Online).

Izzaty Shahirah Nor Sham, N., Salleh, R., & Syahirah Syed Sheikh, S. (2021). Women Empowerment and Work-Life Balance of Women Engineers in the Malaysian Energy Sector: A Conceptual Framework. SHS Web of Conferences, 124. https://doi.org/10.1051/shsconf/202112408009

Kurzman, C., Dong, W., Gorman, B., Hwang, K., Ryberg, R., & Zaidi, B. (2019). Women's Assessments of Gender Equality. Socius, 5. https://doi.org/10.1177/2378023119872387

Lee, T., Good, L., Lipton, B., & Cooper, R. (2022). Women, work and industrial relations in Australia in 2021. Journal of Industrial Relations, 64(3). https://doi.org/10.1177/00221856221099624

Loayza, N., & Trumbic, T. (2021). Women, Business and the Law 2021: Women´s economic empowerment is critical to resilient recovery efforts. World Bank Blog.

Njuki, J., Eissler, S., Malapit, H., Meinzen-Dick, R., Bryan, E., & Quisumbing, A. (2022). A review of evidence on gender equality, women's empowerment, and food systems. Global Food Security, 33. https://doi.org/10.1016/j.gfs.2022.100622

Novák, J., Benda, P., Šilerová, E., Vaněk, J., & Kánská, E. (2021). Sentiment Analysis in Agriculture. Agris On-Line Papers in Economics and Informatics, 13(1). https://doi.org/10.7160/aol.2021.130109

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes. Bulletin of Computer Science and Electrical Engineering, 1(1). https://doi.org/10.25008/bcsee.v1i1.5

Ortiz Rodríguez, J., & Pillai, V. K. (2019). Advancing support for gender equality among women in Mexico: Significance of labor force participation. International Social Work, 62(1). https://doi.org/10.1177/0020872817717323

Pichad, S. (2023). Analysing Sentiments for YouTube Comments using Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 11(5). https://doi.org/10.22214/ijraset.2023.51973

Quinlan, J., & VanderBrug, J. (2016). Gender Lens Investing: Uncovering Opportunities for Growth, Returns, and Impact. In Gender Lens Investing.

Rahab, H., Zitouni, A., & Djoudi, M. (2021). SANA: Sentiment analysis on newspapers comments in Algeria. Journal of King Saud University - Computer and Information Sciences, 33(7). https://doi.org/10.1016/j.jksuci.2019.04.012

Singh, R., & Tiwari, A. (2021). YOUTUBE COMMENTS SENTIMENT ANALYSIS. International Journal of Scientific Research in Engineering and Management (IJSREM.

Siscawati, M., Adelina, S., Eveline, R., & Anggriani, S. (2020). Gender Equality and Women Empowerment in The National Development of Indonesia. Journal of Strategic and Global Studies. https://doi.org/10.7454/jsgs.v2i2.1021

Tanesab, F. I., Sembiring, I., & Purnomo, H. D. (2017). Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine. International Journal of Computer Science and Software Engineering (IJCSSE), 6(8).

Thelwall, M. (2018). Gender bias in sentiment analysis. Online Information Review, 42(1). https://doi.org/10.1108/OIR-05-2017-0139

Wani, A., & Ahmad, M. (2021). WOMEN IN THE WORK PLACE – AN EMPIRICAL ANALYSIS OF THE CHALLENGE OF WORK LOAD. Advancing Women in Leadership Journal, 40(1). https://doi.org/10.21423/awlj-v40.a377