Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 23, No. 2, March 2024, pp. 389∼404
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v23i2.3883

Unsafe Conditions Identification Using Social Networks in Power Plant Safety Reports

Annisaul Mubarokah[1], Rita Ambarwati[2], Dedy [3]
[1]Universitas Muhammadiyah Sidoarjo, Sidoarjo Indonesia
[2]PLN Nusantara Power, Surbaya, Indonesia
[3]Samarkand Institute of Economics and Service, Samarkand, Uzbekistan

Article Info

Article history:

Received January 19, 2024
Revised February 20, 2024
Accepted March 20, 2024

Keywords:

Occupational Health and Safety
Social Network Analysis
Unsafe Condition

ABSTRACT

Power plants in Indonesia grapple with significant challenges in managing occupational health and safety. Power generation companies urgently need to reduce workplace accidents yearly and need an application to report every potential workplace hazard. The huge reporting data in applications such as IZAT requires thorough analysis to determine the pattern and distribution. This research aimed to facilitate the company in hazard mitigation by identifying reported unsafe conditions and building a semantic association network to understand the nature of unsafe conditions between Paiton and Indramayu generating units. The research method used was social network analysis, which is carried out by preprocessing the data using programming to remove noise and then converting the data into a readable format. Then, semantic relationships between words were analyzed, and the data was visualized using the ForceAtlas2 algorithm. The findings revealed a different focus between the two units, where 6.597 reports from the Paiton generating unit mainly highlighted team response and accident-prone workplace conditions. In contrast, 5.840 reports from the Indramayu unit emphasized specific conditions, locations, and equipment that pose accident risks.

Corresponding Author:

Rita Ambarwati, +628113400992
Faculty of Business, Law and Social Science,
Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia,
Email: ritaambarwati@umsida.ac.id

How to Cite:
A. Mubarokah, R. Ambarwati, D. Dedy, and M. Alimova, ”Unsafe Conditions Identification Using Social Networks in Power Plant Safety Reports”, *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 23, No. 2, pp. 389-404, Mar, 2024.

1. INTRODUCTION

Recently, accidents in Indonesia have received national attention, along with the high number of work accidents that have occurred in recent years. Based on data from internal companies, PT PLN Nusantara Power has a high risk in the field of Occupational Safety and Health (K3), where the history of work accidents at PT PLN Nusantara Power from 2015 to 2020 with losses of more than 13 billion in the last incident of work accidents and fires in the workplace. From the problem of work accidents that occur, the OHS Development Division at the Head Office (PLN Nusantara Power) developed a strategy in an effort to create Zero Accident for all Generating Units, which creates a safe, healthy and comfortable work environment for all employees, in order to support the company in achieving optimal performance. Based on accident statistics at PT PLN Nusantara Power from 2016 to 2023, the number of work accidents increased in 2018 and began to decline in 2021. There was 1 serious injury accident in 2016, and no serious injury accidents were recorded in subsequent years. Injury is a term that encompasses a wide range of physical harm that can affect individuals in various ways [1]. Meanwhile, there were fatality accidents with 1 case in 2016, 1 case in 2017, 4 cases in 2018, 1 case in 2019, 2 cases in 2020, and no fatality accidents in 2021 to 2023. In the electrical industry, fatality refers to the occurrence of death resulting from electrical accidents [2]. Based on this explanation, it can be interpreted that PLN Nusantara Power can almost meet the Zero Accident target at work. PT PLN needs to maintain and continue to develop to achieve zero accidents in the company.

PT PLN Nusantara Power has a website and mobile-based application to facilitate the OHS patrol system through the use of smartphones [3]. IZAT (Zero Accident Assistant Application) is an application that optimizes the entire OHS business process by creating a management application to plan, implement, control, and evaluate the OHS process. In addition, this application facilitates the implementation of OHS activities by scheduling patrols, collecting findings reports, and carrying out follow-up actions. The IZAT application has a status-finding feature; the reporter can report hazardous conditions into four types: positive (non-hazardous conditions), near miss, unsafe action, and unsafe condition. In Heinrich's theory (1931), one of the factors causing work accidents is unsafe conditions. An unsafe condition in the work environment can potentially increase the risk of work accidents [4]. Unsafe conditions can be defined as situations that pose a danger to oneself, others, or local facilities and infrastructure. Ninety-eight percent of occupational accidents and diseases are preventable, and establishing a safety culture is the key to their elimination; safety attitudes influence safety risk perception and hazard identification [5, 6].

Previous research provides solutions for preventive measures of accident occurrence using various methods. Research conducted by Pramod Kumar analyzed retrospective accident reports by classifying accidents based on system activity and human error categories. Researchers applied fuzzy mathematics to estimate the likelihood of human error as a contributing factor to the occurrence of accidents [7]. The research conducted by XU Na identified risk factors for workplace accidents using Text Mining (TM) technology. The approach explains the risk factors obtained and describes the important causes contributing to workplace accidents in China [8]. Furthermore, research conducted by Bilal to determine preventive measures for work accidents by predicting construction incidents using Latent Class Clustering Analysis (LCCA) and Artificial Neural Networks (ANNs). Researchers provide practical preventive measures to companies to avoid incidents according to prediction results [9]. The research conducted by Botao Zhong proposes a learning-based approach to analyzing text with accident reports. Researchers use Latent Dirichlet Allocation (LDA) based network analysis methods to visualize factors contributing to work accidents [10]. The novelty of this research is to provide solutions to prevent work accidents by using Social Network Analysis to determine the distribution of unsafe action reports by utilizing IZAT application report data in the Paiton generating unit and Indramayu generating unit. This research aims to make it easier for companies to decide on accident prevention measures in the workplace by categorizing hazard report data and visualizing it. This follows the research conducted by Seyedeh and Adel, which provides categories based on rational data to create a portrait of OHS hazard prevention in cyclical activities in the mining sector [11]. Prevention of harm is important because Occupational Health and Safety (OHS) is a thought and effort to ensure wholeness and perfection both spiritually and physically, which aims to maintain the comfort and safety of labor to achieve work power, physical endurance, and a high level of health [12].

The limited literature on using social network analysis (SNA) to analyze unsafe conditions is a gap in this study. This research further fills the lack of literature on using SNA in analyzing unsafe conditions. This study uses social network analysis by utilizing programming applications such as Python, Wordij, and application Gephi to process large amounts of data regarding reports of unsafe conditions in the workplace. The results of this categorization and comparison can be a reference for companies to facilitate the prevention and handling of work accidents. Categorization in reporting aims to classify words that describe unsafe conditions that occur. This classification process can make it easier for employees to classify unsafe conditions in each potential hazard report. Employees' ease can increase their awareness of unsafe conditions and thus improve their reporting mechanisms. Thus, companies can develop strategies to reduce the risks associated with these unsafe conditions. Companies can identify risks to workers and reduce them in ways such as replacing unreliable equipment and improving uncomfortable work environments. This can help reduce workplace accidents and optimize occupational safety and health [13, 14]. This study compares the unsafe conditions between the Paiton and Indramayu generating units because both units have similarly high productivity and have active reporting activities.

In addition, the Paiton and Indramayu generating units combine 3 previous small units. The results of visualization data will be analyzed based on statistical data and graphical forms. Data visualization can facilitate the use of data, such as by illustrating the relationship between types of hazardous materials and the risk of occupational accidents using software tools [15]. In the case of hazard identification, categorization and visualization of data can assist in identifying relationships between factors that may influence occupational accidents and optimize data analysis [16]. This research has two objectives, namely to categorize the results of unsafe condition findings in the IZAT application in the Paiton and Indramayu units and to determine the comparison of the visualization results of unsafe condition findings in the IZAT application that occurred between the Paiton unit and the Indramayu unit.

2. RESEARCH METHOD

This study employs a qualitative descriptive research methodology, a subset of qualitative research focused on delineating the characteristics of a specific phenomenon [17]. The primary objective of qualitative descriptive research is to facilitate a thorough and nuanced comprehension of the phenomenon in question, grounded in its real-world manifestations. Such research is instrumental in interpreting and depicting the existing data and circumstances. It excels at furnishing detailed descriptions and analyses of the phenomena under investigation, encompassing both scientific and humanistic perspectives. This approach is adept at handling data in various non-quantitative formats, including textual, visual, and other forms of non-numerical information, thereby offering a holistic view of the subject matter. The researcher's data source is secondary data from the IZAT application, obtained from interviews and company document data. The research subjects are all IZAT users who report while at work. The object of research is data containing comments on the condition column from IZAT application users who fall into the category of unsafe conditions at the Paiton generating unit and the Indramayu generating unit. The research location is the IZAT application.

The availability of quite a lot of data in the IZAT application certainly has a pattern of distribution of reports that resembles each transaction so that it can be utilized and used as consideration through the understanding of data mining science so that it can be utilized and taken into consideration through the understanding of data mining [18]. The problem in this study is the communication pattern of reporting through the IZAT application to the reporter's response to risk findings with the unsafe condition category and the relationship between actors from the graph analysis formed in the network structure. The method that can be used is Social Network Analysis (SNA), which is part of the Social Computing technique to extract information on unstructured data and has a large volume. Social Network Analysis (SNA) is a subset of social networking methods that leverages graph theory to study human relationships and extract information from large datasets. SNA uses nodes and edges connected in social networks to visualize social [19]. SNA deals with the analysis of information embedded in social networks. Such information can be categorized into two types: the first is structure-based information, which represents the topological structure of the network, and the second is content-based, which represents features related to entities and their relationships [20]. Social Network Analysis (SNA) is a statistical and analytical technique used to analyze the structure, interactions, and relationships between individuals, groups, or organizations in a social network. SNA uses interaction data to build graphs that describe the relationships between entities in a social network. Using statistical methods and programming algorithms, SNA can understand the structure of social networks and identify relationships between different entities [21, 13].

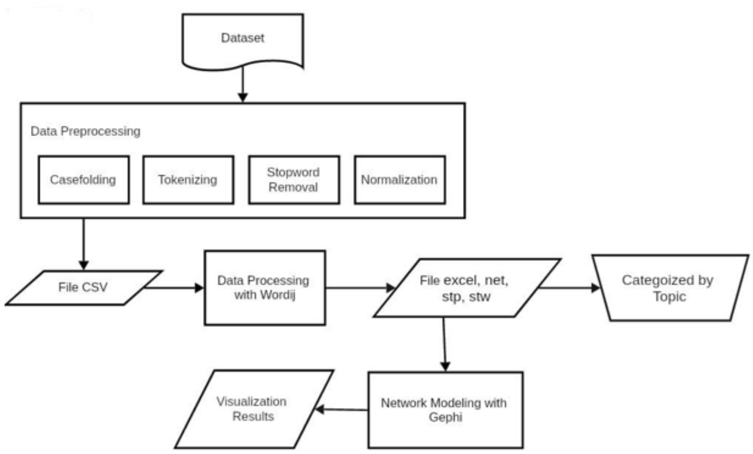

Figure 1. Research Flowchart

The concept of this research is illustrated in Figure 1 for designing social network analysis. The formulation of social network analysis uses points, symbols, and user interactions represented by lines. This requires an analysis that can offer a new perspective on how people connect with individuals or communities [22]. The initial stage in the research process is a literature review of the phenomena that occur in the form of historical data through the Zero Accident Assistant Application (IZAT). The data was then extracted (scrapping) in Excel, covering the coverage of three contents: unsafe condition findings, title of findings, and generating units. Data pre-processing is required before entering the analysis and visualization stage. The purpose of pre-processing in this study is to ensure that the data collected in text format only consists of terms that are relevant to the analysis. This process involves four main stages: case folding, tokenizing, normalizing, and filtering. In implementation, it uses the Python language, a high-level and multipurpose programming language. Many programming paradigms, such as procedural programming languages and object-oriented programming, are supported with Python [23]. After that, the data will be processed using the Wordij application to change its format so that computer systems can read it. The processed data will be used for word categorization and visualization, which will be further analyzed for the information contained in the data. Gephi 0.9.2 software was used to identify and visualize the structure of the formed property network. Gephi is an interactive tool used to visualize, examine, or assess various types of simple and complex networks and dynamic and hierarchical graphs. This study presents the results of research and discussion, which is the result of a comparison of network property calculations from each generating unit and images that show words that often appear in the subject of findings in each generating unit.

2.1. Data Source

Data collection comes from secondary data obtained from company data, which contains reporting findings based on the problem's topic, which is the research object. The data obtained comes from the IZAT application, which contains findings on unsafe conditions. Based on the findings and reporting on the IZAT application, it was found that reporting on unsafe conditions was the most reported hazard category at PT PLN Nusantara Power. The data shows the work environment's condition in the Paiton and Indramayu Generating Unit, which generates units that utilize coal energy. The data obtained is the result of reporting from 2020 to 2023.

2.2. Data Preprocessing

Data processing is necessary to remove noise in the data, such as abbreviations and informal words that are difficult for computers to understand [24]. Process preprocessing has several stages; the following are the stages of preprocessing [25]. The data preprocessing process consists of four main stages. Firstly, case folding involves converting all uppercase letters in the dataset into lowercase. This step ensures consistency in text format throughout the dataset. Next, tokenization breaks down sentences into individual words, effectively removing noise and simplifying further analysis. Stopword removal follows, where common but insignificant words are filtered out to focus on relevant content, thus refining the dataset. Finally, normalization standardizes the document structure to prevent duplication, streamlining data preparation and processing for more efficient analysis. These stages enhance data quality and usability, facilitating meaningful insights in subsequent analytical tasks.

2.3. Data Processing

The data in CSV format that has been generated in the data preprocessing process is then processed using wordij then the results are in the form of data formats such as excel, net, stp and stw. Wordij is a program that moves windows through text to count word pairs based on proximity. The results of wordij will be classified based on predetermined categories following the research problem.

2.4. Categorization by Topic

Data categorization was conducted to define the data mining needs in filling potential hazard reporting at IZAT. Researchers provide appropriate categories based on data processing results to create a table of unsafe condition categories in reporting activities in the company. Data categorization is made by entering each word from Wordij's Excel format and processing results into a predetermined topic category. Researchers identified five topic categories: place, response, cause, tools, and condition. The five topics are selected based on the company's work environment. The categories cover the needs that can be used for hazard identification. Data categorization makes hazard identification easier by looking at how many words appear from each topic.

2.5. Network Modeling and Visualization

In this process of creating a network model, Gephi software visualizes the network model. Data visualization is the process of displaying data that has gone through processing into a visual form to provide a clearer and easier-to-understand picture. Data visualization in this research will produce nodes and edges representing actors and their relationships. To see the keywords or topics most often reported on the IZAT application. The final step involves the application of Gephi to create a social network derived from Bigram phrases. This is done to analyze the value of the network structure and visualize the relationship between the words obtained. In Gephi, the ForceAtlas2 algorithm is used for setup, with varying colors and sizes assigned to each node and edge to enhance the interpretability and visual clarity of the network analysis. ForceAtlas2 is an algorithm for forcibly orienting nodes to show the relationship between each node [26].

3. RESULT AND ANALYSIS

This research contains data comparing the findings of the Paiton Generating Unit and Indramayu Generating Unit in the IZAT application regarding unsafe condition reporting from 2020 to 2023. Word analysis must be done so researchers can understand the text more deeply to identify topics better and increase rigor and flexibility [27].

3.1. Data Source

Data regarding reports on the IZAT application owned by PT PLN Nusantara Power is obtained directly through company archives. The data collected includes unsafe condition reports on the Paiton and Indramayu generating units. Data on unsafe conditions obtained from the IZAT application results in the number of data:

Table 1. Number of Unsafe Condition Data

Content	Generating Unit	Year	Amount of data
Unsafe Condition	Paiton	2020-2023	6.597
	Indramayu		5.840

Table 1 shows that the amount of data obtained from the IZAT application with the keywords to be studied focuses on unsafe conditions. This study collected data from 2020 to 2023 with a total of 12.437 findings reported from the Paiton Generating Unit and Indramayu Generating Unit. There are 6597 findings from the Paiton Generating Unit reporting and 5840 findings from the Indramayu Generating Unit reporting.

3.2. Data Preprocessing

The data that has been obtained will go through the preprocessing stage. This process is used to clean the data from noise and is ready for use in the next process. The preprocessing process has several stages, and the following are the preprocessing stages.

1. Case Folding
   Case folding is used to make all words the same font size because font sizes are not always used consistently in reports. This process can also remove punctuation and redundant spaces, preventing uppercase and lowercase letters from being detected as having different meanings. Figure 2 is the code snippet used to implement the case folding stage. Table 2 displays the results obtained from the case folding process.

```
# ------ Case Folding --------
# make all sentence lowercase
df_olah1['Text'] = df_olah1[' title'].str.lower()
```

Figure 2. Code Snippet for Case Folding

Table 2. The Result of The Case Folding Process

Example of Report Text	Case Folding
SU #1 Welding Receptacle open & chaotic wiring	su #1 welding receptacle open & chaotic wiring
DST A panel box opened, and box door detached	dst a panel box opened and box door detached

2. Tokenizing The tokenizing process is done to separate a series of words in a sentence, paragraph, or page into tokens or single-word chunks. This process facilitates the calculation of the presence of the word in the document and the frequency of occurrence of the word. At the same time, tokenizing also removes characters other than letters, such as punctuation marks. Figure 3 is the code snippet used to implement the tokenizing stage. Table 3 displays the results obtained from the tokenizing process.

```
# ------ Tokenizing ---------

def remove_tweet_special(text):
    text = str(text)
    # remove tab, new line, ans back slice
    text = text.replace('\\t'," ").replace('\\n'," ").replace('\\u'," ").replace('\\',"")
    # remove non ASCII (emoticon, chinese word, .etc)
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove mention, link, hashtag
    text = ' '.join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\/\/\S+)"," ", text).split())
    # remove incomplete URL
    return text.replace("http://", " ").replace("https://", " ")

df_olah1['Text'] = df_olah1['Text'].apply(remove_tweet_special)

#remove number
def remove_number(text):
    return  re.sub(r"\d+", "", text)
```

Figure 3. Code Snippet for Tokenizing

Table 3. The Result of The Tokenizing Process

Example of Report Text	Tokenizing
SU #1 Welding Receptacle open & chaotic wiring	welding receptacle open chaotic wiring
DST A panel box opened, and box door detached	panel box opened, and box door detached

3. Stopword Removal This stopword removal process is used to remove words that have no effect on the sentiment process. Words from report data are compared with words contained in the stopword database, and the result of this process is to eliminate words detected in the database. Stopword removal removes frequently occurring words that are not important in the document. This is done without changing the document's meaning to be processed further. Figure 4 is the code snippet used to implement the stopword removal stage. Table 4 displays the results obtained from the stopword removal process.

```
# get stopword English
list_stopwords = stopwords.words('english')


# ---------------------------- manualy add stopword  -----------------------------------
# append additional stopword
list_stopwords.extend(['and', 'or', 'am'])

# ----------------------- add stopword from txt file -----------------------------------
# read txt stopword using pandas
txt_stopword = pd.read_csv("C:/Users/USER/Kamus/stopwords-en.txt", names= ["stopwords"], header = None)

# convert stopword string to list & append additional stopword
list_stopwords.extend(txt_stopword["stopwords"])

# ---------------------------------------------------------------------------------------

# convert list to dictionary
list_stopwords = set(list_stopwords)


#remove stopword pada list token
def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]

df_olah1['Text_tokens_WSW'] = df_olah1['Text_tokens'].apply(stopwords_removal)


print(df_olah1['Text_tokens_WSW'].head())
```

Figure 4. Code Snippet for Stopword Removal

Table 4. The Result of The Stopword Removal Process

Example of Report Text	Stopword Removal
SU #1 Welding Receptacle open & chaotic wiring	”welding” ”receptacle” ”open” ”chaotic” ”wiring”
DST A panel box opened, and box door detached	”panel” ”box” ”opened” ”door” ”detached”

4. Normalization Normalization is used as a process of converting data into a standard or 'normal' form to facilitate data processing and speed up data reading operations. The main purpose of data normalization is to reduce and even eliminate data redundancy or repetition without changing the data's meaning. Figure 5 is the code snippet used to implement the normalization stage. Table 5 displays the results obtained from the normalization process.

```
# Indonesian slang | English can be added
path = "C:/Users/USER/ARTIKEL/Tools_Pendukung/Kamus/Data_slang.csv"
normalizad_word = pd.read_csv(path)

normalizad_word_dict = {}

for index, row in normalizad_word.iterrows():
    if row[0] not in normalizad_word_dict:
        normalizad_word_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalizad_word_dict[term] if term in normalizad_word_dict else term for term in document]

df_olah1['tweet_normalized']= df_olah1['Text_tokens_WSW'].apply(normalized_term)

df_olah1['tweet_normalized'].head(10)
```

Figure 5. Code Snippet for Normalization

Table 5. The Result of The Normalization Process

Example of Report Text	Normalization
SU #1 Welding Receptacle open & chaotic wiring	"weld" "receptacle" "open" "chaotic" "wire"
DST A panel box opened, and box door detached	"Panel" "box" "open" "door" "detach"

3.3. Data Processing

The data obtained from the preprocessing stage is then processed so that a graph is obtained from the JSON file. Graph representation in this study uses Wordij software. Wordij is a system based on a connectedness or network model representing textual information. The fundamental unit of analysis is word pairs, or two-word phrases, rather than individual terms. Wordij is a text analysis program that treats words as interconnected nodes and edges for network and other statistical analyses. The first stage is to replace the text content from the preprocessing results so that all text does not have a sign, $/\,'$, and provide a distance between data with enter. Replace the $','$ sign into space. Replace the "[' and [' signs with blanks. Replace the ']" & '] signs into $\backslash n$. After that, save the file and process it with Wordij. Data in CSV format is then converted into data formats such as Excel, net, stp and stw. The Wordij application is displayed in Figure 6, and the results of the Wordij process are shown in Table 6.

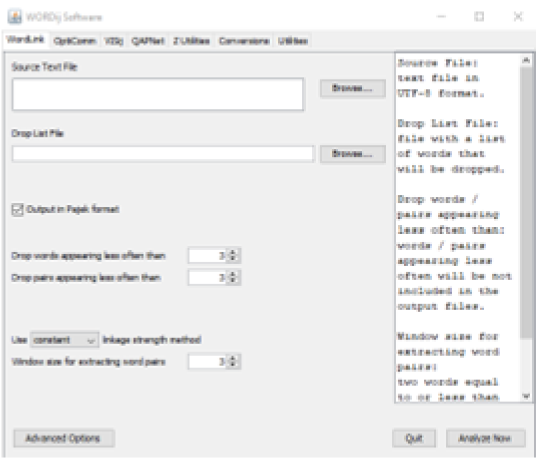

Figure 6. Wordij Display

Table 5 displays the total number of words, the number of unique words, and the average number of all words per unique word as a result of data processing using the Wordij tool. The data will form a visualization that connects the words in the dataset, and each word is called a node or word phrase. The number of words is the number of words that appear in the data retrieval from the IZAT application. Unique words are filtered words that come from many words to avoid repetition of words that appear [28]. Table 2 shows that 11.555 words in total, including 594 unique words that appeared, and an average of 19,453 words appeared by the Paiton generating unit. In the meantime, the Indramayu generating unit produced 14.014 words in total, with 738 words appearing in the unit and 18,989 words appearing on average. The number of words generated by the IZAT application regarding unsafe conditions is higher in the Indramayu generating unit than in the Paiton generating unit.

3.4. Categorized by Topic

The results processed using wordij produce words with the frequency of the number of words. Word frequency is obtained from the total number of words occurring in both generating units. Where the number of words that appear is divided by all the total words. The words were selected according to similar criteria and then grouped into their respective categories by analyzing their interrelationships and logic. The researcher identified five category topics: place, response, cause, tools, and condition. The five categories represent the common characteristics of the reporting sentences from both units.

Table 6. Identification of Thematic Word Categories

Category 1: Place		**Category 2: Response**		**Category 3: Cause**		**Category 4: Tools**		**Category 5: Condition**	
Unit	1,26%	Patrol	13,15%	Water	1,34%	Lamp	2,45%	Damaged	2,61%
Room	0,98%	Patrole	2,63%	Rubbish	0,76%	Door	1,34%	Broken	1,27%
Floor	0,98%	Check	1,34%	Ash	0,46%	Lighting	0,86%	Used	1,01%
Appear	0,92%	Independent	1,23%	Stak	0,42%	Hydrant	0,85%	Leak	0,88%
CCR	0,73%	Routine	0,57%	Oil	0,28%	Cable	0,83%	Leave	0,6%
Street	0,68%	Finding	0,55%	Waste	0,18%	Box	0,77%	Dirty	0,44%
Boiler	0,67%	Cek	0,52%	Grass	0,09%	Pipe	0,61%	Close	0,32%
Place	0,64%	Care	0,23%	Gas	0,08%	Panel	0,6%	Access	0,3%
Roof	0,62%	Inspection	0,12%	Mud	0,04%	AC	0,57%	Empty	0,23%
Stairs	0,49%	Repair	0,1%	Mask	0,04%	Hose	0,52%	Open	0,2%

Table 6 shows the authors' top 10 words for each category. The category results display how often words are mentioned when reporting unsafe conditions at both generating units. It is clear that Category 2 ”response,” Category 4 ”equipment,” and Category 5 ”condition” are the three categories of most concern for reporting at Paiton and Indramayu Generating Units with a particular focus on ”patrol” (13.15%), ”damaged” (2.61), ” lamp” (2.45%), and ”patrole” (2.63%). The relationship between Category 1, ”Place,” and Category 3, ”Cause,” is almost the same in each generating unit, indicating that reporting related to these categories has the same level of findings in reporting in both generating units.

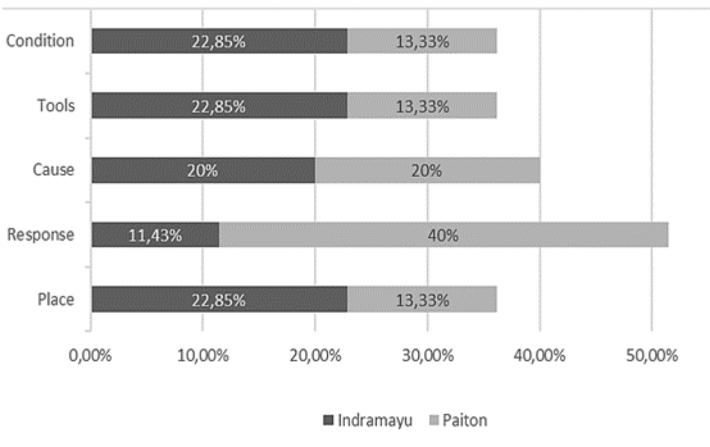

Figure 7. Category Distribution on Both Generating Units

Figure 7 shows the relative differences of the five categories across the two generating units. The X-axis of the figure represents the percentage of high-frequency words in the five categories in each generating unit. A higher percentage indicates more attention paid to that category. The Y-axis represents the five categories. In general, the relationship of the five categories is quite different between the two generating units. The percentages of ”Response” and ”Cause” are higher, while the percentages of ”Condition,” ”Tools,” and ”Place” are relatively low. Category 2 response appears more prominent in the Paiton Generating Unit (40%) than in the Indramayu Generating Unit (11.43%), indicating that the responses obtained from the reports are noticed in the Paiton Generating Unit [29]. Category 3, ”Cause,” was comparable between Paiton and Indramayu Generating Units, indicating that both generating units listed the causes of the unsafe condition. Category 1 ”Place,” Category 4 ”Tool,” and Category 5 ”Condition” did not differ much between Indramayu Generating Unit (22.85%) and Paiton Generating Unit (13.33%), which may be related to the reporting of findings received often mentioning tools, places, and conditions where it describes unsafe conditions.

The categorization of safety hazards can assist in determining the safety measures required for solving each type of unsafe condition, ensuring that employees are protected from potential hazards. This category contributes to the general objective of creating a safer work environment and reducing occupational hazards. Employers can better prioritize risks and allocate resources by categorizing hazards correctly, focusing on the most important dangers first. Employers can communicate risks to their hiring managers by categorizing unsafe conditions, allowing them to understand potential risks and take appropriate precautions. This category aims to protect workers’ well-being and reduce workplace risks. Categorization enables employers to effectively communicate to their employees the risks associated with specific unsafe conditions, raising awareness and promoting a safety culture in the workplace [30]. Such information can improve the capacity to predict, avoid, and respond to threats and their impacts [31]. Companies can use this classification to assess the efficacy of their existing health and safety programs, identify gaps, and continuously improve their strategies to protect employee health.

3.5. Network Modeling and Visualization

To investigate the structural characteristics of the network for each generating unit, the preliminary results obtained from Wordij necessitate further processing utilizing Gephi. This critical phase will engender edges and facilitate sophisticated data visualization. In this investigation, the application of data visualization is aimed at creating nodes and edges, which symbolically represent actors and their interactions, respectively. Nodes are configured as geometric shapes that embody either words or actors, playing a pivotal role in the structural representation of the network. Conversely, edges are utilized to depict the connections between these entities, offering a visual articulation of the relational dynamics within the network. Each node is intertwined or connected, called an edge, which is presented in the form of a graph in the form of a node and represents a two-way interaction [32]. Gephi was used for network visualization in this study to make the semantic relationships of thematic words more intuitive. In visualizing the relationship of bigram phrases, researchers used the ForceAtlas2 algorithm. ForceAtlas2 is an algorithm for forcibly orienting nodes to show the relationship between each node [26]. Compared to other layout algorithms, ForceAtlas2 has better measurable quality [27]. ForceAtlas2 is a force-directed layout: it simulates a physical system to spatialize a network. Nodes repulse each other like charged particles, while edges attract their nodes like springs. These forces create a movement that converges to a balanced state. The use of ForceAtlas2 keeps nodes that are not very related to other nodes away, making it easier to understand the relationship between nodes. This final configuration is expected to help the interpretation of the data. The following is a display of the use of ForceAtlas2 for visualization.

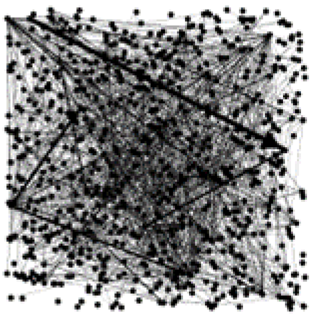

Figure 8. Before using ForceAtlas2

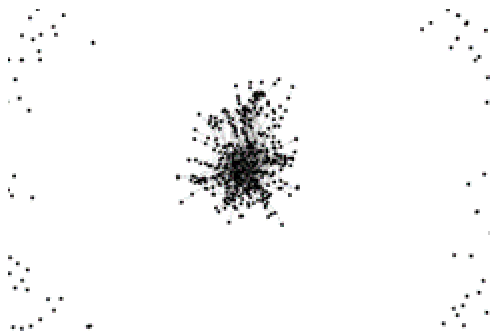

Figure 9. After using ForceAtlas2

Using betweenness centrality scores indicated by font size and modularity conveyed by color differences will result in a clearer and more engaging visualization. Choose a font size corresponding to the nodes' betweenness centrality score, then adjust the color according to modularity. Node and edge coloring or font size can be adjusted as needed. The visualization results for bigram phrase relationships are in Figure 10 and Figure 11.

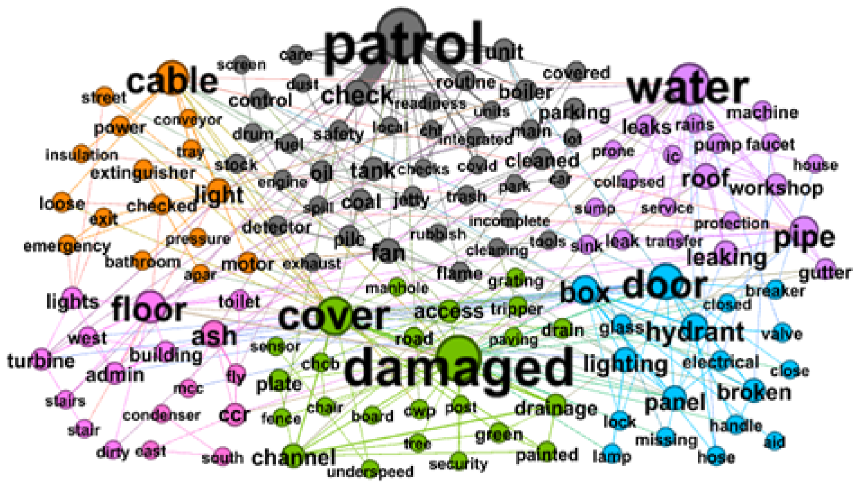

Figure 10. Visualization of The Paiton Generating Unit

Figure 10 shows the visualization of the unsafe condition network in the Paiton generating unit from 2020 to 2023. The visualization of unsafe conditions displayed produces various colors that distinguish each existing network node and edge. Figure 10 produces visualization results with six colors: gray, green, blue, orange, purple, and pink. The gray network of nodes and edges is more directed towards the response and location of the unsafe condition. The green nodes and edges discuss more about the tools and conditions where unsafe conditions are reported. The purple network of nodes and edges is more directed towards the conditions and causes of unsafe conditions. The blue nodes and edges are more about the tools and conditions where unsafe conditions occur. The orange network of nodes and edges is more directed towards unsafe tools and conditions. Meanwhile, the pink nodes and edges are more directed towards the tool and the location of the unsafe condition. Based on this analysis, the results of reporting on the Paiton unit more often mention the responses and conditions found in the workplace related to unsafe conditions. The core nodes in the Paiton generating unit network can be seen from how large the scale is in the image, such as the words "patrol," "damaged," "door," and "water." This shows that unsafe conditions that occur in the Paiton generating unit are due to a lack of attention related to the routine inspection system, and unsafe conditions often occur in several facilities and infrastructure of the Paiton unit.

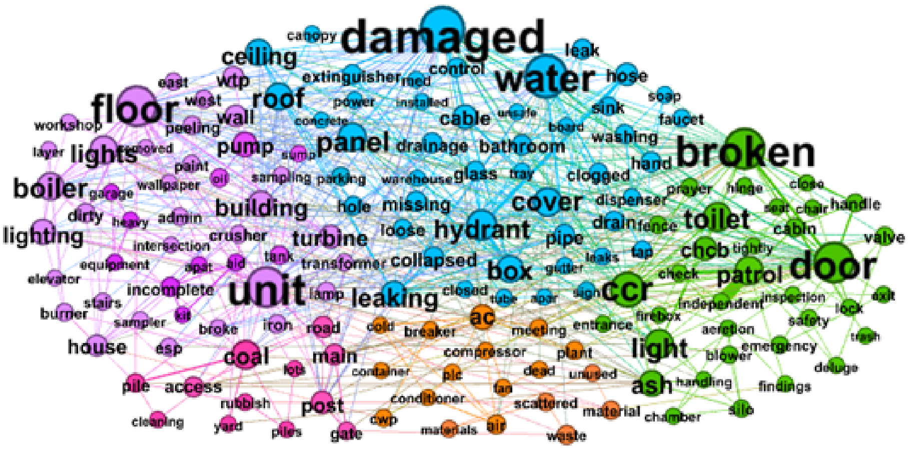

Figure 11. Visualization of The Indramayu Generating Unit

Figure 11 shows the network visualization of unsafe conditions in the Indramayu generating unit from 2020 to 2023. Unsafe condition visualization in Figure 11 produces visualization results with five colors: purple, blue, orange, pink, and green. The purple network of nodes and edges is more directed toward the tools and locations where unsafe conditions occur. The blue nodes and edges discuss more about the cause and condition where unsafe conditions are reported. The orange nodes and edges network is more directed towards the tools of the unsafe condition. Pink-colored nodes and edges are more about the cause and place where unsafe conditions occur. The network of green nodes and edges is more directed towards the conditions, tools, and places of unsafe conditions. Based on this analysis, the reporting results more often discuss the conditions, tools, and places found in the workplace that can cause unsafe conditions. The core nodes in the Indramayu generating unit network can be seen from the magnitude of the scale in the image, such as the words "damaged," "door," "broken," "unit," and "floor." This shows that the unsafe condition in the Indramayu generating unit is due to a lack of attention to the maintenance of equipment conditions and the need for routine checks in each room in the unit.

Table 7. Network Properties of Paiton and Indramayu Generating Units

Network Properties	Value	
	Paiton	Indramayu
Nodes	142	155
Edges	305	624
Average Degree	1,455	2,388
Average Weighted Degree	10,222	15,674
Network Diameter	11	10
Modularity	0,659	0,549
Average Path Length	3,998	3,52

Through data processing, generating units can be compared by creating a network with property values. A network node property describes the number of keyword interactions in Table 8. Networks with more tied nodes lead to far faster penetration [29]. There are 142 nodes in the Paiton generating unit and 155 nodes in the Indramayu generating unit in 2020-2023. The larger the number of nodes, the more words that match the keywords are used. The edge property network is a network or line that connects nodes and illustrates how much discussion between nodes [33]. There are 305 edges in the Paiton generating unit and 624 edges in the Indramayu generating unit. The data distribution is improved with more edges because there are more connections or interactions between nodes. Additionally, the average network property is the number of links connecting nodes. The average degree describes the ratio of the total number of network ties to the sum of all ties [29]. The Paiton generating unit network is 1,455, while the Indramayu generating unit is 2,388. The greater the average degree, the faster and easier the spread of information. In a network graph, the average number of link weights connecting nodes to other nodes is described by the network property known as the average weighted degree. The average weighted degree produced by the Paiton generating unit network is 10,222, whereas the Indramayu generating unit yields an average of 15,674. A higher average weighted degree number indicates a better average speed of information dissemination [33].

The network diameter property is the maximum or longest distance in a network. Network diameter represents the distance traveled in the network [28]. The Paiton generating unit network shows a network diameter of 11, while the Indramayu generating unit is 10. Information spreads more quickly and easily with a smaller or shorter diameter. Therefore, a small diameter reduces the time required for the information step. The network property modularity describes how strong a group is in the network. The Paiton generating unit network shows a modularity of 0.659, while the Indramayu generating unit shows a modularity of 0.549. The greater the modularity in a network, the better, meaning that the groups formed in the network have a solid relationship [33]. Average path length is the average geodetic distance or the average path taken by each node to another node. The network properties of the Paiton generating unit show an average path length of 3,998, while the Indramayu generating unit shows an average path length of 3,52. The researcher argues that the thematic tendency discussed by the reporter reflects the OHS performance in each unit. For example, the Paiton unit needs to communicate with all parties involved in the OHS patrol to give more supervision to equipment or machines that often experience damage by monitoring the maintenance schedule so that the work safety and security system operate properly. Not much different at the Indramayu unit, the OHS system also needs more supervision of equipment and locations where unsafe conditions often occur to reduce or prevent the occurrence of unsafe conditions. There needs to be a comprehensive application of identifying unsafe conditions related to the work environment and equipment used at work.

4. CONCLUSION

This study found differences in the core issues in the two power generation units. Paiton unit workers' reports were more directed towards OHS responses related to patrols and routine checks. The statistical analysis of thematic word classification results shows that the reporting on the Paiton unit leads to a response of 40%, in contrast to the Indramayu unit, which is 11.43%. This finding shows that the service level of the patrol team has a big influence on unsafe conditions that occur in the Paiton unit. However, reporting on the Indramayu unit is more directed towards the conditions, tools, and places that cause unsafe conditions, with a concern of 22.85%. The results indicate differences in the reporting of unsafe conditions in the two steam power generation units (PLTU). The reporting on the Indramayu generating unit is more focused on the situations, locations, and equipment that may result in hazardous conditions while reporting on the Paiton generating unit is more focused on the response and circumstances that lead to unsafe conditions. It implements that workplace accidents' most frequently reported causes, such as unsafe conditions, can be avoided. The research indicates that power generation units should improve the performance of the Occupational Health and Safety (OHS) team and strengthen worker involvement in providing more detailed and factual reports about workplace problems.

An overview of unsafe condition reporting is anticipated from the SNA method analysis of unsafe conditions and the category and visualization results to assist each generating unit in better meeting the company's objectives and receiving positive input from the community. This research has limitations related to the data analyzed because it only uses unsafe condition reports. Unsafe conditions are only one of the hazards that can cause work accidents. A suggestion for future research is to use the Social Network Analysis (SNA) method on similar big data such as traffic accident reports or employee performance reports. Future researchers can also use other causes of work accidents, such as unsafe acts or near misses. Then, related research can be carried out in other fields, such as construction, health, business, and so on.

5. ACKNOWLEDGEMENTS

We thank our colleagues from Universitas Muhammadiyah Sidoarjo, who provided insight and expertise that greatly assisted the research.

6. DECLARATIONS

AUTHOR CONTIBUTION

The first author is me, Annisaul Mubarokah, as the author for each data collection and data extraction on the dataset. The second author is Dr. Rita Ambarwati Sukmono, S.E., M.MT., as analytical research on every aspect of existing data and also conducting research in research methodology. The third and fourth authors are Dedy and Mashhura Toironovna Alimova, as helpers, correct the results of each result of the object and also check the results.

FUNDING STATEMENT

Statements regarding financing in conducting research are the personal property of authors 1, 2, and 3.

COMPETING INTEREST

The authors declare that there are no competing interests in the publication of this research.

REFERENCES

[1] T. Horino, "Acute Kidney Injury: Definition and Epidemiology," in *Acute Kidney Injury and Regenerative Medicine*. Singapore: Springer Singapore, 2020, pp. 3–20, https://doi.org/10.1007/978-981-15-1108-0_1.

[2] B. Brenner and D. Majano, "Expanding Workplace Electrical Safety to Nonelectrical Occupations: Strategies for Reducing Fatal Electrical Injuries," *IEEE Industry Applications Magazine*, vol. 27, no. 6, pp. 76–81, nov 2021, https://doi.org/10.1109/MIAS.2021.3065319.

[3] D. G. Fitri, H. Tolle, and H. M. Az-zahra, "Evaluasi dan Perbaikan Antarmuka Aplikasi Mobile IZAT (Aplikasi Zero Accident Assistant) Bidang K3 PT PJB UP Gresik Menggunakan Metode Human-Centered Design," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 10, pp. 9661–9670, 2019, https://doi.org/https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6528/3134.

[4] E. Ratman, S. R. Karimuna, and S. K. Saptaputra, "Gambaran Tindakan Tidak Aman (Unsafe Action) Dan Kondisi Tidak Aman (Unsafe Condition) Pada Pekerja Proyek Kantor Perakilan Bank Indonesia (KPwBI) Di Kota Kendari Tahun 2019," *Jurnal Kesehatan dan Keselamatan Kerja Universitas Halu Oleo*, vol. 1, no. 1, pp. 28–35, jun 2020, https://doi.org/10.37887/jk3-uho.v1i1.12615.

[5] Z. F. Olcay, S. Temur, and A. E. Sakalli, "A research on the knowledge level and safety culture of students taking occupational health and safety course," *Cypriot Journal of Educational Sciences*, vol. 16, no. 1, pp. 187–200, feb 2021, https://doi.org/10.18844/cjes.v16i1.5519.

[6] D. Kashmiri, F. Taherpour, M. Namian, and E. Ghiasvand, "Role of Safety Attitude: Impact on Hazard Recognition and Safety Risk Perception," in *Construction Research Congress 2020*, vol. 007, no. 1994. Reston: American Society of Civil Engineers, nov 2020, pp. 583–590, https://doi.org/10.1061/9780784482872.063.

[7] P. Kumar, S. Gupta, and Y. R. Gunda, "Estimation of human error rate in underground coal mines through retrospective analysis of mining accident reports and some error reduction strategies," *Safety Science*, vol. 123, p. 104555, mar 2020, https://doi.org/10.1016/j.ssci.2019.104555.

[8] N. XU, L. MA, Q. Liu, L. WANG, and Y. Deng, "An improved text mining approach to extract safety risk factors from construction accident reports," *Safety Science*, vol. 138, p. 105216, jun 2021, https://doi.org/10.1016/j.ssci.2021.105216.

[9] B. U. Ayhan and O. B. Tokdemir, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001762.

[10] B. Zhong, X. Pan, P. E. Love, L. Ding, and W. Fang, "Deep learning and network analysis: Classifying and visualizing accident narratives in construction," *Automation in Construction*, vol. 113, p. 103089, may 2020, https://doi.org/10.1016/j.autcon.2020.103089.

[11] S. A. Baghaei Naeini and A. Badri, "Identification and categorization of hazards in the mining industry: A systematic review of the literature," *International Review of Applied Sciences and Engineering*, vol. 15, no. 1, pp. 1–19, jan 2024, https://doi.org/10.1556/1848.2023.00621.

[12] A. Hasibuan, B. Purba, R. Makbul, A. Mansida, L. O. M. Y. Amsah, F. Rachim, V. Dharmawan, E. Bachtiar, P. Sumantrie, S. Simbolon, Erdawaty, Muadzah, and R. Herno Della, "Penerapan Sistem Manajemen Keselamatan dan Kesehatan Kerja," in *Teknik Keselamatan dan Kesehatan Kerja*, 1st ed., J. Simarmata, Ed. Indonesia: Yayasan Kita Menulis, 2020, ch. 12, p. 204, https://doi.org/{https://repositori.uin-alauddin.ac.id/19792/1/2020_Book%20Chapter_Teknik%20Keselamatan%20dan%20Kesehatan%20Kerja.pdf}.

[13] N. Nurrokhman, H. Dwi Purnomo, and K. Dwi Hartomo, "Utilization of Social Network Analysis (SNA) in Knowledge Sharing in College," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 4, no. 2, pp. 259–271, aug 2020, https://doi.org/10.29407/intensif.v4i2.14460.

[14] D. Kurniawan, A. Iriani, and D. Manongga, "Pemanfaatan Social Network Analysis (SNA) Untuk Menganalisis Kolaborasi Karyawan Pada Pt. Arum Mandiri Group," *Jurnal Transformatika*, vol. 17, no. 2, p. 149, jan 2020, https://doi.org/10.26623/transformatika.v17i2.1646.

[15] P. G. S. Anggraini and I. Yuadi, "Tren Publikasi Climate Change (Perubahan Iklim) Tahun 2020-2023 Pada Scopus," *JIPI (Jurnal Ilmu Perpustakaan dan Informasi)*, vol. 8, no. 2, p. 213, nov 2023, https://doi.org/10.30829/jipi.v8i2.15917.

[16] N. Noeraida, "Perkembangan publikasi internasional bidang pemantauan radiasi tahun 2011-2019 melalui basisdata Scopus," *Berkala Ilmu Perpustakaan dan Informasi*, vol. 16, no. 1, pp. 68–82, jun 2020, https://doi.org/10.22146/bip.v16i1.296.

[17] C. T. Reis, S. G. Paiva, and P. Sousa, "The patient safety culture: a systematic review by characteristics of Hospital Survey on Patient Safety Culture dimensions," *International Journal for Quality in Health Care*, vol. 30, no. 9, pp. 660–677, nov 2018, https://doi.org/10.1093/intqhc/mzy080.

[18] F. N. Hasan, A. S. Aziz, and Y. Nofendri, "Utilization of Data Mining on MSMEs using FP-Growth Algorithm for Menu Recommendations," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 2, pp. 261–270, mar 2023, https://doi.org/10.30812/matrik.v22i2.2166.

[19] N. Prabowo, "Social Network Analysis for User Interaction Analysis on Social Media Regarding E-Commerce Business," *IJIIS: International Journal of Informatics and Information Systems*, vol. 4, no. 2, pp. 95–102, sep 2021, https://doi.org/10.47738/ijiis.v4i2.106.

[20] A. Kumari, R. K. Behera, K. S. Sahoo, A. Nayyar, A. Kumar Luhach, and S. Prakash Sahoo, "Supervised link prediction using structuredbased feature extraction in social network," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 13, jun 2022, https://doi.org/10.1002/cpe.5839.

[21] M. K. Bratawisnu and A. Alamsyah, "Social Network Analysis Untuk Analisa Interaksi User Di Media Sosial Mengenai Bisnis E-Commerce," *Almana : Jurnal Manajemen dan Bisnis*, vol. 2, no. 2, pp. 107–115, aug 2018, https://doi.org/http://journal.unla.ac.id/index.php/almana/article/view/143/107.

[22] A. J. Al-Bayati, "Impact of Construction Safety Culture and Construction Safety Climate on Safety Behavior and Safety Motivation," *Safety*, vol. 7, no. 2, p. 41, may 2021, https://doi.org/10.3390/safety7020041.

[23] M. Mambwe, E. M. Mwanaumo, W. D. Thwala, and C. O. Aigbavboa, "Evaluating Occupational Health and Safety Management Strategy Success Factors for Small-Scale Contractors in Zambia," *Sustainability*, vol. 13, no. 9, p. 4696, apr 2021, https://doi.org/10.3390/su13094696.

[24] E. J. Tetzlaff, K. A. Goggins, A. L. Pegoraro, S. C. Dorman, V. Pakalnis, and T. R. Eger, "Safety Culture: A Retrospective Analysis of Occupational Health and Safety Mining Reports," *Safety and Health at Work*, vol. 12, no. 2, pp. 201–208, jun 2021, https://doi.org/10.1016/j.shaw.2020.12.001.

[25] D. I. Af'idah, D. Dairoh, S. F. Handayani, R. W. Pratiwi, and S. I. Sari, "Sentimen Ulasan Destinasi Wisata Pulau Bali Menggunakan Bidirectional Long Short Term Memory," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 607–618, jul 2022, https://doi.org/10.30812/matrik.v21i3.1402.

[26] D. C. U. Lieharyani and R. Ambarwati, "Visualisasi Data Tweet di Sektor Pendidikan Tinggi Pada Saat Masa Pandemi," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 116–123, jun 2022, https://doi.org/10.47065/bits.v4i1.1551.

[27] Z. Hou, F. Cui, Y. Meng, T. Lian, and C. Yu, "Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis," *Tourism Management*, vol. 74, no. January, pp. 276–289, oct 2019, https://doi.org/10.1016/j.tourman.2019.03.009.

[28] A. N. Navisha, R. Ambarwati, and M. Hariasih, "Twitter Social Network Interaction As Customer Engagement In Competition For E-Commerce E-Health Performance In Indonesia," *Jurnal Manajerial*, vol. 10, no. 02, p. 303, jun 2023, https://doi.org/10.30587/jurnalmanajerial.v10i02.5279.

[29] E. Muller and R. Peres, "The effect of social networks structure on innovation performance: A review and directions for research," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 3–19, mar 2019, https://doi.org/10.1016/j.ijresmar.2018.05.003.

[30] L. Yuliana and D. Ardhyaksa, "Analysis Of Unsafe Action And Unsafe Condition Based On Occupational Health And Safety Reporting Programs," *Journal of Global Research in Public Health*, vol. 4, no. 2, pp. 78–86, 2019, https://doi.org/10.30994/jgrph.v4i2.40.

[31] B. Fakhruddin, J. Kirsch-Wood, D. Niyogi, L. Guoqing, V. Murray, and N. Frolova, "Harnessing risk-informed data for disaster and climate resilience," *Progress in Disaster Science*, vol. 16, no. September, p. 100254, dec 2022, https://doi.org/10.1016/j.pdisas.2022.100254.

[32] S. Azad and S. Devi, "Tracking the spread of COVID-19 in India via social networks in the early phase of the pandemic," *Journal of Travel Medicine*, vol. 27, no. 8, pp. 1–9, dec 2020.

[33] A. S. Rachma, R. Ambarwati, and M. Yani, "Comparison of Twitter Users' Perception of Content Marketing Effectiveness and Service Quality in Two Online Transportation," *Almana : Jurnal Manajemen dan Bisnis*, vol. 7, no. 1, pp. 134–146, apr 2023, https://doi.org/10.36555/almana.v7i1.2132.

[This page intentionally left blank.]