[1]Jerin Johnkutty

[2]Dr.Serajul Haque

[3]Jerry Davis T

Exploratory Data Analysis using Machine learning – Behaviour Based Safety

Abstract: - This study takes a gander at the usage of machine learning techniques for exploratory data examination (EDA) in the field of behavior-based safety (BBS), with a particular focus on the examination of safety dimension datasets got from worker reviews drove in industrial settings. We utilize a methodology that integrates an extent of visualization methods, statistical examinations, and parameter evaluations to uncover complex encounters into safety perceptions and ways of behaving. We do this by utilizing the Python programming language and its strong data investigation libraries, including MATplotlib, Seaborn, and Pandas. Our survey means to assist proof based decision-making strategies, proactively distinguish potential for developing safety protocols, and develop a organizational culture that is safety-centric driven by eagerly examining worker feedback and behavioral patterns. This research features the significant role of EDA and machine learning in deciphering complex datasets, advancing substantial improvements in occupational safety, and putting a high priority on workers' well-being in dynamic workplaces through synergistic cooperation between data analytics and domain expertise.

Keywords: Exploratory Data Analysis (EDA), Behaviour Based Safety (BBS), Machine Learning Techniques, Worker Reviews, Occupational Safety, Data Analytics

Introduction

Rapid technological advancement and ideas like lean production in today's culture create brand-new, complex hazards [1]. Consequently, businesses that want to operate successfully in a more sustainable way must take proactive measures to reduce their negative effects on the economy, environment, and society[2, 3]. One factor that can affect an organization's safety performance is the impact on society, the environment, and the economy [4]. The definition of safety performance is the quality of safety-related work. Improving an organization's safety performance can raise its resilience or robustness, lowering the chance of accidents. Poor safety performance, on the other side, might raise the organization's vulnerability and hence the likelihood of accidents [5]. Poor design, gaps in oversight, and unworkable processes are examples of latent circumstances that are hypothesized to cause accidents in organizations. Employee attitudes, beliefs, perceptions, and values (safety culture), the environment that impacts employees (working environment), and routines and procedures are all examples of latent conditions (safety activities) [6, 7]. The work defines safety performance as the complete performance in a safety culture, working environment, and safety activities,Workplace safety culture and environment [8]. Statistics on occupational injuries are critical for determining how well workers are protected from workplace hazards and dangers. Workplace safety and health are critical components of decent work [9, 10]. An occupational accident is defined as an unexpected and unplanned incident, including acts of violence, which occurs during or in connection with work and causes personal harm, sickness, or death to one or more workers [11]. A case of occupational injury is one worker who sustains an occupational injury as a result of a single occupational mishap [12, 13]. An occupational injury can be deadly (as a result of an occupational accident and death occurs within one year of the event) or non-fatal, resulting in missed work time.

Behaviour Based Safety (BBS) is a strategy that uses safety observations to inform management and employees about the overall safety of the workplace [15]. BBS is designed to draw workers' attention to their own and their

[1] Research Scholar, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai

[2]Assistant Professor, Dept. of Mechanical Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai

[3]Assistant Professor, Dept. of Mechanical Engineering, Government Engineering College Thrissur, Kerala

* jerinjohnkutty18@gmail.com ,** serajulhaque@crescent.education

***jerrydavist@gectcr.ac.in

Corresponding Author - Dr.Serajul Haque

(serajulhaque@crescent.education)

colleagues' regular safety behaviour. BBS program aims to increase the employee safety of the organization [16, 17]. Observers (workers trained to perform on-site safety checks)also conduct reviews of other employees with an eye on their behaviour when implementing a BBS [18]. These observers document safe and dangerous conduct as well as safe and unsafe workplace circumstances. The observer then informs the worker of his or her observations and gives comments [19, 20]. Positive feedback is welcome. Discussing strategies for employees to conduct their activities more safely helps workers and observers become more conscious of their actions. BBS programmers are built on a continuous feedback loop in which employees and observers offer comments on how to enhance safety to one another, and safety professionals use the data gathered during the observations to continuously improve the BBS programme. Organizations that adopt a BBS programmer establish the proper list of behaviours to watch depending on their organization's particular habits and hazards [21-23]. Safety professionals often create a checklist style that is simple and quick for field observers to complete and outlines the objective behaviours [24-25]. Each organization must have a unique strategy for installing BBS. Employees are responsible for their own personal safety while working with BBS, according to the firm. For the BBS deployment inside the organization to be effective, all workers must participate in the program [26-27]. Every employee, regardless of their position in the organization's structure, must comprehend and implement the benefits of BBS. Every employee, from the CEO to the front-line workers, must be included in the BBS implementation process, including hourly, salaried, and union personnel. Change is unavoidable since the organization's policies, methods, and/or systems will need to be modified in order to achieve the intended behavioural changes. Any modification requires the participation of the whole crew [23, 29].

Literature Review

Occupational safety and health (OSH) remain critical concerns in various industries worldwide. Understanding the dynamics of safety stressors, social support systems, and safety performance is crucial for ensuring worker well-being and organizational productivity.Sampson, [21] explore the intricate relationship between safety stressors, social support, and safety performance among unionized pipefitters. Their study, grounded in action theory, highlights the nuanced impact of safety obstacles and uncertainty on safety compliance and participation. While safety uncertainty and obstacles were negatively correlated with safety participation, safety compliance was mainly affected by uncertainty. Manager support, especially positive job-related communication, significantly improved safety performance, highlighting the significance of viable communication channels in advancing safety culture. Interventions on the basis of (BBS) are a proactive method for decreasing working environment injuries and accidents. A meta-analysis was done by [22] to evaluate the viability of BBS interventions in various occupational settings. Their outcomes show a statistically significant diminishing in accidents and injuries following the execution of BBS, despite methodological limitations. The authors do, however, issue a warning against exaggerating these discoveries since some examination had subpar procedure. Solid intervention plans that are adjusted to specific workplace prerequisites are supported, just like the utilization of control groups in evaluations to expand their validity.

As the tasks of the construction industry are by their very nature complex, cautious planning and execution are important to maximize efficiency and productivity. [23] use structural equation modeling (SEM) to dissect the critical factors influencing the safety risk tolerance of construction workers. Their review features what risk versatility is a perplexing construct that is impacted by a people perspectives, past experiences, traits of the job, and safety techniques. Curiously, safety management is viewed as a central part influencing risk tolerance, featuring the fundamental job that organizational safety culture plays in picking the mindsets of representatives toward security. The pursuit for expanding construction productivity and efficiency has earned attention, particularly in quickly arising economies like India. As per [24], labor deficiencies, unexpected disturbances, material delays and plan revisions are a piece of the significant obstructions preventing construction efficiency. To expand construction activities and shortening delays, their study underlines the significance of effective project management methodology, brief material procurement, and proactive risk moderation techniques. Credit risk the management is fundamental to safeguarding financial stability and decreasing the risk of loan in the banking business. [28] investigate loan default patterns and credit risk in Ghanaian banks. Banks actually struggle with high credit default rates even with an assortment of credit risk management techniques in light of the CAMPARI model. The authors stress that to decrease credit risks and assurance financial supportability, severe credit assessment practices, the formation of credit reference agencies, and further developed client instruction programs are fundamental.

Exploratory data analysis (EDA) methods help identify patterns and generate hypotheses by providing insightful information about intricate datasets. A thorough introduction to EDA principles and computational tools is given by [29], who also emphasizes how complementary EDA is to confirmatory data analysis (CDA). Through the development of a deeper understanding of data structures and underlying patterns, analytical model refinement and the formulation of strong hypotheses are made possible by EDA.

RESEARCH GAP:

Author	Proposed Methodology	Results	Research Gap
Sampson et al.	Action theory framework, survey method	Identified negative relationship between safety stressors and performance	Lack of exploration into specific safety stressors' impact on performance
Tuncel et al.	Meta-analysis of BBS interventions	Statistically significant reduction in accidents/injuries	Need for methodologically robust studies in BBS interventions
Kabil & Sundararaju	Behavioral safety partnership, survey method	Improved safety behavior outcomes observed	Lack of detailed examination on specific safety behavior interventions
Wang et al.	Structural equation modeling (SEM), questionnaire surveys	External factors had larger impact on risk tolerance	Limited understanding of internal vs. external factors influencing risk
Subramani & Rajiv	Survey method, factor analysis	Identified critical factors affecting construction productivity	Need for empirical studies to evaluate effectiveness of interventions
Ntow-Gyamfi & Boateng	Survey method, analysis of credit risk management tools	Banks using varied risk management tools, default rates remain high	Inefficacy of current risk management tools to mitigate loan defaults
Behrens	Literature review, contrasting EDA with CDA	EDA complements CDA, emphasizes need for EDA integration in data analysis	Limited incorporation of EDA techniques in statistical training

The table summarizes various research approaches and results from multiple studies. The application of action theory by Sampson et al. clarifies the complex relationship between safety stressors and performance, emphasizing the need for more targeted research into the effects of particular stressors. The meta-analysis by Tuncel et al. highlights that in order to validate the efficacy of Behavior-Based Safety (BBS) interventions, methodological rigor is necessary. The structural equation modeling of Wang et al. highlights the limited understanding of internal versus external risk factors and emphasizes the dominance of external factors in influencing risk tolerance. Critical factors influencing construction productivity are identified by Subramani & Rajiv's factor analysis, which calls for empirical research to assess the efficacy of interventions. Ntow-Gyamfi & Boateng's analysis of credit risk management tools reveals the inefficacy of current practices in mitigating loan defaults, indicating a need for improved risk management strategies. Behrens' review accentuates the complementary nature of Exploratory Data Analysis (EDA) alongside Confirmatory Data Analysis (CDA), advocating for the integration of EDA into statistical training and research methodologies to enhance data interpretation and hypothesis development. Overall, these findings collectively stress the imperative for methodological robustness and targeted investigations to address existing research gaps in safety, risk management, productivity enhancement, and data analysis across various domains.

Proposed Methodology:

1. Data Acquisition and Preprocessing:

We commence our methodology by acquiring tabular data from safety dimensions review datasets. The dataset, represented as D, comprises m samples and n features, where m=473 and n=38. Each sample pertains to worker reviews on various safety dimensions. Upon loading the dataset into a Pandas DataFrame, denoted as **X**, we preprocess the data to ensure its readiness for analysis. The preprocessing steps include handling missing values, encoding categorical variables if any, and standardizing numerical features, if necessary.

2. Exploratory Data Analysis (EDA):

Under this phase, we embark on exploring the fundamental characteristics and relationships within the dataset. Our exploratory analysis encompasses:

2.1. Data Exploration:

We employ descriptive statistics to summarize the central tendencies, dispersions, and distributions of the dataset. The summary statistics include the mean (μ), standard deviation (σ), and quartiles of each feature.

2.2. Data Visualization:

Utilizing Python's visualization libraries, including Matplotlib and Seaborn, we create various visualizations such as histograms, scatter plots, and box plots to elucidate the distributional properties and relationships between features.

2.3. Correlation Analysis:

To unveil the interdependencies among safety dimensions, we compute the Pearson correlation coefficient (ρ) between pairs of features. The correlation matrix **C** generated facilitates the identification of strong positive or negative correlations between safety dimensions.

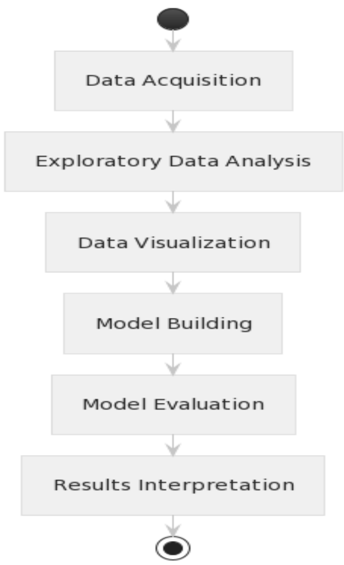


Figure 1: Machine Learning Process (Data analysis & Visualization techniques)

Figure 1 depicts a sequential flow of activities within a machine learning process. It begins with "Data Acquisition," signifying the initial step of obtaining tabular data from relevant sources. The procedure then switches to "Exploratory Data Analysis," in which the obtained data is scrutinized to determine its essential traits and connections. After data exploration, "Data Visualization" techniques are used to produce a variety of visual representations, including scatter plots and histograms, to clarify the relationships and distributional properties of the features. Following an adequate measure of data exploration and visualization, the process continues to "Model Building," where predictive models are fabricated utilizing the information accumulated from the previous stages. In the "Model Evaluation" stage that follows model construction, the models' performance is assessed utilizing measurements like mean squared error and coefficient of determination. In the "Results Interpretation" stage, the results are interpreted, offering critical information about the models' predictive capacity and highlighting important variables impacting the phenomena under study.

3. Modeling:

We move to model building to forecast the Safety Behavior dimension based on other safety dimensions after gaining insights from EDA. The following steps are part of our modeling pipeline:

3.1. Feature Selection:

A subset of safety dimensions (Xselected) that are thought to be significant in predicting safety behavior are chosen. The components of stress recognition, safety awareness, safety commitment, teamwork, and safety compliance are incorporated in this subset.

3.2. Data Splitting:

Utilizing a 60-40 split, we partition the dataset into training (Xtrain) and test (Xtest) sets. While the test set surveys the model's generalization performance, the training set makes model parameter assessment simpler.

3.3. Linear Regression Modeling:

To model the relationship between Safety Behavior and the chosen safety dimensions (Xselected), we use linear regression. The following represents the linear regression model:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

Where xi are the chosen safety dimensions, y^is the expected safety behavior, β0 is the intercept, and the coefficients are βi.

3.4. Model Evaluation:

Measurements like the coefficient of determination (R2) and mean squared error (MSE) on the test set (Xtest) are utilized to assess the model's performance. These measurements measure how well the model predicts results and makes sense of variety in safety behavior.

4. Results Interpretation:

The outcome of our examination shed light on significant factors affecting working environment safety discernments and ways of behaving and offer insightful data about the predictive force of safety aspects on safety behavior.

We understand the intricate connections among safety aspects thorough numerical and statistical investigations, and we construct a predictive model that can direct designated interventions intended to advance a more secure and better workplace.

RESULTS:

Linear Regression Modeling and Performance Evaluation:

The developed linear regression model, which forecasts safety behavior in light of the picked security dimensions, displays favorable outcomes. Utilizing critical measurements on the test dataset, like mean squared error (MSE) and coefficient of determination R^2 , is vital for an exhaustive evaluation of the model's effectiveness. While the MSE estimates the precision of the model's predictions, the R^2 value is a strong indicator that clarifies the percentage of changeability in safety behavior explained by the chose aspects. Carefully laid out performance measurements for the model alongside the coefficients identified by the linear regression model are reflected in the organized showcase given by Table 2.

Table 2: Linear Regression Model Performance and Coefficients

Model Performance Metrics	
R^2	0.76

60% of the data set is taken as training data set and the balance is test data set. From the training data set, we kept the five dimensions (Safety Commitment, Safety compliance, Safety awareness, Teamwork, and Stress Recognition) as independent and one dimension as dependent. We trained a linear regression model to predict one dimension (Safety Behavior) of a 40% data set. The model predicted the output with an accuracy of 76%. (R^2 value 0.769). Important information about the operation and coefficients of the linear regression model used to forecast safety behavior based on chosen dimensions is summarized in Table 2. The model performance metrics show that the selected dimensions account for 76% of the variance in safety behavior, with a coefficient of determination R^2 value 0.76. The coefficients of the linear regression model further delineate the impact of each safety dimension on safety behavior. This comprehensive overview aids in understanding the predictive capacity of the model and underscores the relative importance of individual safety dimensions in shaping workplace safety perceptions and behaviors.

```
r_squared = reg.score(X_test, y_test)
r_squared
```

```
0.7690405426215614
```

Working With the Data Sets

The data we're using is from safety dimensions review data set. We will analyze the data and consider possible options.

1. Import the Pandas libraries in the first step from Numpy Package.

2. Read the relatively large Safety Questionnaire CSV file as a data frame df (variable name). It displays the data sets as rows and columns. There are 473 rows and 38 columns in our CSV file. To return the top 5 rows of the data frame, we used the .head () method.

3. To learn more about the data frame, we used df.describe(). This will return the average, mean, standard deviation, and so on for integer and float-type values in the data frame.

4. In the Next step organised the column names of questions from each dimension into a separate list.

5. We used df.corr () to find correlations between each question and created a heat map for each correlation. Figure 4 depicts this pairwise correlation as a heat map.

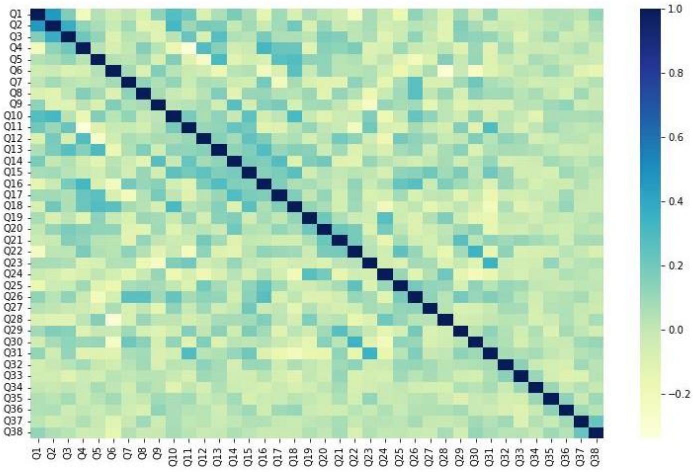
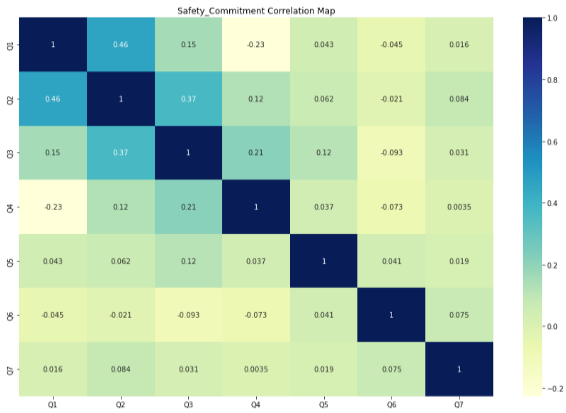
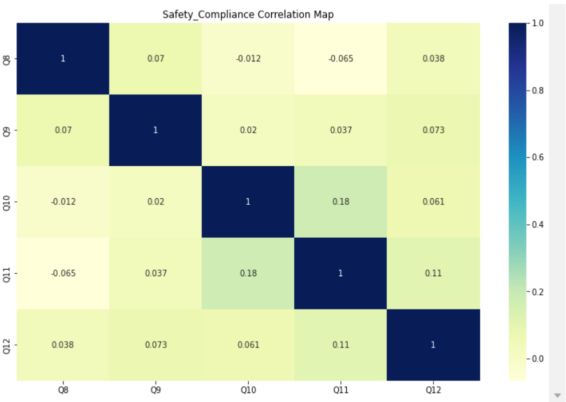


Figure 4. Heat Map of all features

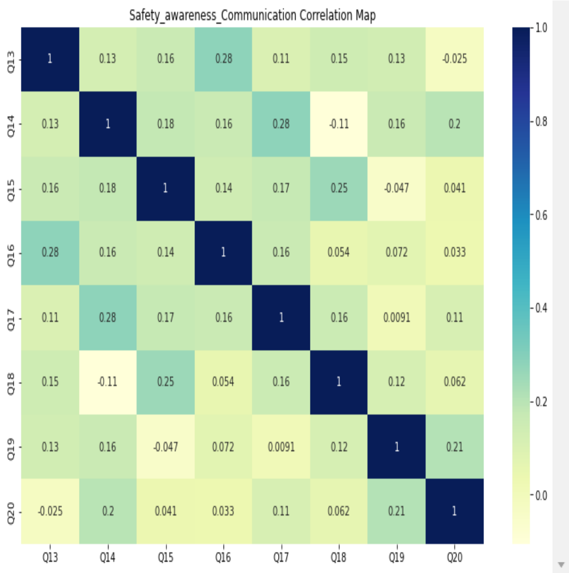
6. Then we obtained the correlation between the questions raised in each safety dimension. The plots of each dimension are shown in Figure 5.



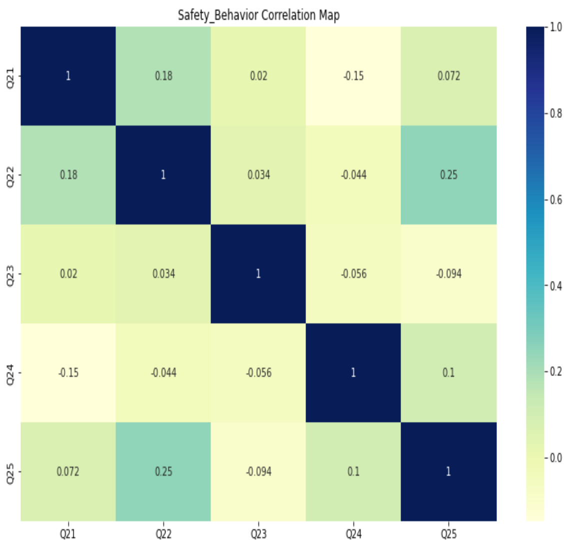
(a)



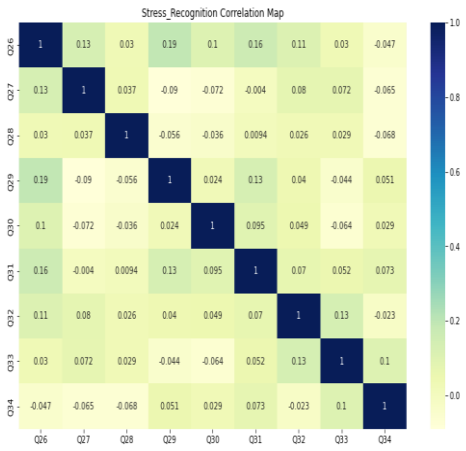
(b)



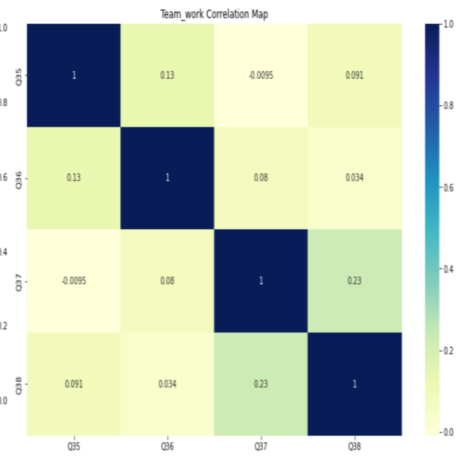
(c)



(d)



(e)



(f)

(a). Safety Commitment (b) Safety Compliance (c) Safety Awareness (d) Safety behaviour (e) Stress recognition (f) Team Work

Figure 5 Heat Map of various dimensions.

7. To conclude on most influenced questions, we have counted each rating with values greater than or equal to 4. Then assigned it as a data frame with a variable name influ. Figure 6 shows the "for loop" used to obtain the above.

```
"""to find columns which have rating more than 4"""
fields = []
for col in df.columns:
#     print(col)
    v = list(df[col])
    count = 0
    for i in v:
        if i >= 4:
            count +=1
    fields.append([col, count])
```

```
influ = pd.DataFrame((fields))#, columns={"max", "feature"})
```

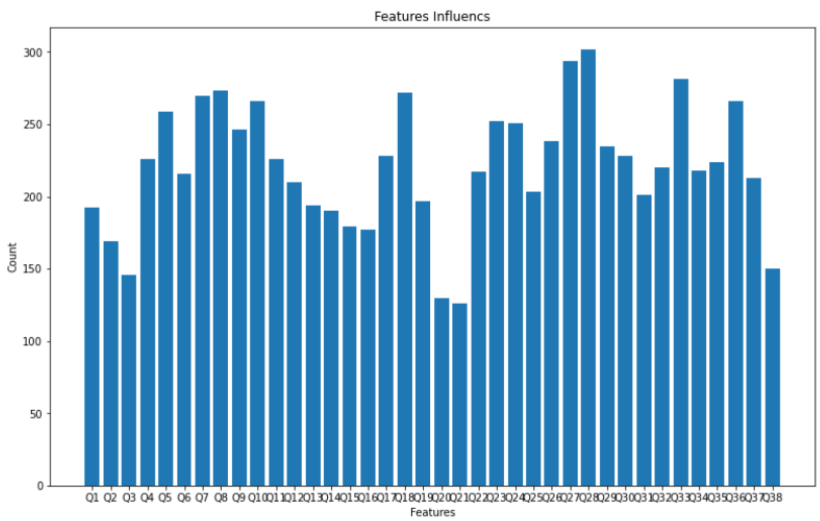


Figure: 6. Features influences

8. We have used the maximum function method on the data frame influ to get the question number of which factor is most rated/influenced by the workers.

```
influ[influ["Max_Count"]== max(influ["Max_Count"])]
```

	Feature	Max_Count
27	Q28	302

9. Also, we have plotted the data frame to visualize the influence of each questionnaire as a bar chart.

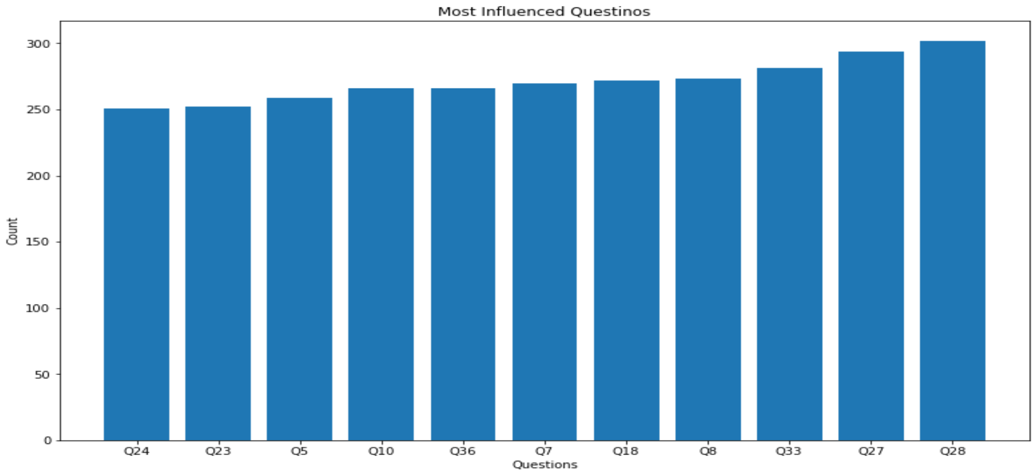


Figure: 7. Influenced questions

Conclusion

In conclusion, this research paper delved into the application of Exploratory Data Analysis (EDA) and machine learning techniques within the domain of Behavior Based Safety (BBS) to analyze safety dimension datasets derived from worker reviews in industrial settings. We used a thorough methodology that included statistical analyses, parameter evaluations, and visualization techniques to glean insights into safety perceptions and behaviors. We did this by utilizing Python programming and reliable data analysis libraries. Our study sought to identify opportunities for improving safety protocols, provide evidence-based decision-making processes, and promote a safety-centric culture within organizations through the careful analysis of worker feedback and behavioral trends. The outcomes demonstrated how well EDA uncovered the underlying features of safety dimension datasets and how machine learning models could predict safety behavior based on specific dimensions. The interrelationships between safety dimensions and their influence on perceptions of workplace safety were better understood through the use of descriptive statistics, visual aids, and linear regression modeling. The results underscore the critical function of data analytics in interpreting intricate datasets, propelling concrete progress in occupational safety, and placing the welfare of employees at the forefront of changing work settings. This study lays the groundwork for future efforts to use data-driven methods to address new safety issues and foster an excellence in safety culture in a variety of industrial settings.

References

[1] Yu, Kai, Qinggui Cao, Changzhen Xie, Nannan Qu, and Lujie Zhou. "Analysis of intervention strategies for coal miners' unsafe behaviours based on analytic network process and system dynamics." *Safety Science* 118 (2019): 145-157.

[2] Yu, Heyao, Jay Neal, Mary Dawson, and Juan M. Madera. "Implementation of behaviour-based training can improve food service employees’ handwashing frequencies, duration, and effectiveness." *Cornell Hospitality Quarterly* 59, no. 1 (2018): 70-77.

[3] Houette, Beate, and Natascha Mueller-Hirth. "Practices, preferences, and understandings of rewarding to improve safety in high-risk industries." *Journal of safety research* 80 (2022): 302-310.

[4] Soltanmohammadlou, Nazi, Sanaz Sadeghi, Carol KH Hon, and Fariba Mokhtarpour-Khanghah. "Real-time locating systems and safety in construction sites: A literature review." *Safety Science* 117 (2019): 229-242.

[5] Fu, Hanliang, Gunasekaran Manogaran, Kuang Wu, Ming Cao, Song Jiang, and Aimin Yang. "Intelligent decision-making of online shopping behaviour based on the internet of things." *International Journal of Information Management* 50 (2020): 515-525.

[6] Kalteh, Haji Omid, Seyyed Bagher Mortazavi, Eesa Mohammadi, and Mahmood Salesi. "The relationship between safety culture and safety climate and safety performance: a systematic review." *International journal of occupational safety and ergonomics* 27, no. 1 (2021): 206-216.

[7] Zhou, Cheng, Hanbin Luo, Weili Fang, Ran Wei, and Lieyun Ding. "Cyber-physical-system-based safety monitoring for blind hoisting with the internet of things: A case study." *Automation in Construction* 97 (2019): 138-150.

[8] Goh, Yang Miang, Chalani U. Ubeynarayana, Karen Le Xin Wong, and Brian HW Guo. "Factors influencing unsafe behaviours: A supervised learning approach." *Accident Analysis & Prevention* 118 (2018): 77-85.

[9] Chen, Hainan, Xiaowei Luo, Zhuang Zheng, and Jinjing Ke. "A proactive workers' safety risk evaluation framework based on position and posture data fusion." *Automation in Construction* 98 (2019): 275-288.

[10] Omidi, Leila, Seyed Abolfazl Zakerian, Jebraeil Nasl Saraji, Esmaeil Hadavandi, and Mir Saeed Yekaninejad. "Prioritization of Human Factors Variables in the Management of Major Accident Hazards in Process Industries Using Fuzzy AHP Approach." *Health Scope* 7, no. 4 (2018).

[11] Zhou, Tuqiang, and Junyi Zhang. "Analysis of commercial truck drivers' potentially dangerous driving behaviours based on 11-month digital tachograph data and multilevel modelling approach." *Accident Analysis & Prevention* 132 (2019): 105256.

[12] Xia, Nini, Mark A. Griffin, Xueqing Wang, Xing Liu, and Dan Wang. "Is there agreement between worker self and supervisor assessment of worker safety performance? An examination in the construction industry." *Journal of safety research* 65 (2018): 29-37.

[13] Fang, Weili, Lieyun Ding, Peter ED Love, Hanbin Luo, Heng Li, Feniosky Pena-Mora, Botao Zhong, and Cheng Zhou. "Computer vision applications in construction safety assurance." *Automation in Construction* 110 (2020): 103013.

[14] Dong, Shuang, Heng Li, and Qin Yin. "Building information modelling in combination with real-time location systems and sensors for safety performance enhancement." *Safety Science* 102 (2018): 226-237.

[15] Hakanen, J.J.; Perhoniemi, R.; Toppinen-Tanner, S. Positive gain spirals at work: From job resources to work engagement, personal initiative and work-unit innovativeness. J. Vocat. Behav. 2008, 73, 78–91.

[16] Hinze, J., Hallowell, M., Baud, K., 2013a. Construction-safety best practices and relationships to safety performance. J. Constr. Eng. Manage. 139(10), 04013006.

[17] Li, H., Lu, M., Hsu, S.C., Gray, M., Huang, T., 2015. Proactive behaviour-based safety management for construction safety improvement. Saf. Sci. 75(6), 107–117.

[18] Martínez-Córcoles, M., Gracia, F., Tomás, I., Peiró, J.M., 2011. Leadership and employees' perceived behaviours in a nuclear power plant: a structural equation model. Saf. Sci. 49, 1118–1129.

[19] Nemanich, L.A.; Keller, R.T. Transformational leadership in an acquisition: A field study of employees. Leadersh. Q. 2007, 18, 49–68.

[20] Øyvind Dahl, Olsen, E., 2013. Safety compliance on offshore platforms: a multi-sample survey on the role of perceived leadership involvement and work climate. Saf. Sci., 54(4), 17–26.

[21] Sampson, J.M., DeArmond, S., Chen, P.Y., 2014. Role of safety stressors and social support on safety performance. Saf. Sci. 64(3), 137–145.

[22] Tuncel, S., Lotlikar, H., Salem, S., Daraiseh, N., 2006. Effectiveness of behaviour based safety interventions to reduce accidents and injuries in workplaces: critical appraisal and meta-analysis. Theor. Issues Ergon. Sci. 7(3), 191–209.

[23] Wang, J., Zou, P.X.W, Li, P.P., 2016. Critical factors and paths influencing construction workers' safety risk tolerances. Accident Anal. Prev. 93, 267–279.

[24] Subramani, T., and S. R. Rajiv. "Improving construction efficiency and productivity of industry using SPSS." *International Journal of Application or Innovation in Engineering & Management (IJAIEM)* 5, no. 5 (2016): 239-250.

[25] Kabil, GV Arockia, and V. Sundararaju. "Behaviour Based Safety in Workplace." *International Journal of Research in Engineering, Science and Management* 2, no. 12 (2019): 327-333.

[26] Perera, H., & Costa, L, "Personality Classification Of Text Through Machine Learning And Deep Learning: A Review (2023)," International Journal for Research in Advanced Computer Science and Engineering, 9(4), 6–12. https://doi.org/10.53555/cse.v9i4.2266.

[27] Bogumil M. Konopka, Felicja Lwow, Magdalena Owczarz, Łukasz Łaczmański, "Exploratory Data Analysis of a Clinical Study Group: Development of a Procedure for Exploring Multidimensional Data," *PLOS ONE*, [Online] https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC6107146/pdf/pone.0201950.pdf, August23, 2018, pp. 1-21.

[28] Matthew Ntow-Gyamfi and Sarah Serwaa Boateng, "Credit Risk and Loan Default among Ghanaian Banks: An Exploratory Study," *Management Science Letters*, Vol. 3, 2013, pp.753–762.

[29] John T. Behrens, "Principles and Procedures of Exploratory Data Analysis," *Psychological Methods*, 1997, Vol. 2, No. 2, pp.131-160.