

## Training Evaluation Analysis Using Text Mining

<sup>1</sup>Mas Akhmad Nainunis, <sup>2</sup>Imam Yuadi

<sup>1</sup>Master Human Resource Development, Post Graduate School, Airlangga University

<sup>2</sup>Department of Information and Library Science, Airlangga University

Email: <sup>1</sup>mas.akhmad.nainunis-2022@pasca-unair.ac.id, <sup>2</sup>imam.yuadi@fisip.unair.ac.id

---

### Article Info

#### Article history:

Received Oct 25<sup>th</sup>, 2023

Revised Dec 30<sup>th</sup>, 2023

Accepted Jan 25<sup>th</sup>, 2024

---

#### Keyword:

Association Rules

FP Growth

Text Mining

Training Evaluation

---

### ABSTRACT

Training evaluation is an evaluation of the results of training that has been carried out. This evaluation includes technical and non-technical factors which are very important for the company to pay attention to when implementing training in the future. Many companies only implement training evaluations as a formality and only include evaluations that are limited by choice, such as closed questionnaires, training evaluations using open questionnaires can provide the freedom to provide positive or negative input that can be of concern to the company. This research aims to find out the words or topics that appear most frequently in open comments on training evaluation results by using the FP Growth algorithm and association rules to find out the relationship between topics or words from the training evaluation results. They are applied to 516 open-ended comments submitted via the post-training questionnaire. The research results showed that 15 association rules were created using Rapidminer using the FP-Growth algorithm with a minimum support of 0.02 and a minimum confidence of 0.5. All rules have a lift value > 1 which indicates that all rules are valid or have a strong association relationship. This research can determine the pattern of comments or suggestions given by workers regarding training evaluation.

*Copyright © 2024 Puzzle Research Data Technology*

---

### Corresponding Author:

Mas Akhmad Nainunis,

Master Human Resources Development, Post Graduate School,

Airlangga University,

Surabaya

Email: mas.akhmad.nainunis-2022@pasca.unair.ac.id

DOI: <http://dx.doi.org/10.24014/ijaidm.v7i1.27607>

---

## 1. INTRODUCTION

Training is a form of effort to increase employee competency so that work is more effective and efficient [1], proven in several research results that can measure the impact of training on improving employee performance [2][3], ability to provide services [4] and employee engagement [5][6]. Trained resources not only have competence in the form of knowledge and skills but also have a positive attitude [2]. Through training and human resource development, organizations or companies can still innovate and be competitive in achieving their strategic vision and mission [7].

Often the training carried out in an organization or company only has a tactical or short-term impact, the training is unable to provide long-term effects, therefore feedback from training participants is needed for continuous improvement regarding the quality of training implementation. Human resource development through training cannot be separated from the quality of the training itself. Several previous studies examined the positive impact of employee satisfaction on training satisfaction and organizational citizenship behavior (OCB) [8][9], employee commitment [10][11], employee loyalty [12][13] and the negative impact on turnover intention [10][12]. Apart from that, high training satisfaction is also related to positive perceptions of knowledge-sharing practices [14][15], thereby preserving tacit knowledge within the organization. The knowledge-sharing process ultimately supports the development of individuals and the organization as a whole. Participants' satisfaction with the training is shown through their assessments and the feedback given by participants after attending the training.

Results of evaluation of participant satisfaction with the implementation of training at PT. XYZ often provides contradictory results when in reality the training program satisfaction index reaches the "very good" standard, but the same training program receives criticism, suggestions, and improvements in the form of open comments. This makes determining what improvements are needed to improve the quality of training increasingly complex.

On the other hand, feedback from customers in the form of open comments is an important point and can contribute to improving service quality [16][17]. Customer feedback based on several studies can be categorized into: appreciation, suggestions, and criticism [18]. Analysis of feedback provides insight into the level of satisfaction, experience, and expectations of service users which can encourage continuous improvement [17], maintain quality, and generate innovations in meeting service user expectations [19]. However, large volumes of feedback cannot be analyzed manually because it results in inconsistencies in interpreting the feedback [20], while the use of conventional methods of content analysis on large amounts of feedback text has weaknesses related to accuracy and complicated coding procedures [16], therefore an analytical method is needed that is capable of extracting information from large amounts of data with a minimal error rate.

Although research exploring the evaluation of participant reactions is still very limited, several relevant previous studies used similar approaches and methods but with different topics and data collection techniques, namely about terrorism and harsh words on Twitter using the FP Growth algorithm [21][22]. Both studies used the same approach and methods as this research, but with different data collection techniques, where this research took data from the results of an open questionnaire and did not take data from social media platforms like previous studies, and this research was conducted in a ceramic manufacturing company.

The implementation of an evaluation model that has not provided steps to improve training, as well as the need to analyze participants' open comments, the number of which continues to increase along with the volume of training implementation, as well as limited research on training evaluation using text mining, encourage researchers to conduct studies related to evaluating training participants' reactions through open comments to determine aspects of improvement in the implementation of training based on the topics or words that most frequently appear from open comments from training participants using text mining. So that the topics or words that appear most often in open comments can be used as considerations for companies to improve the quality of training in the future.

It is hoped that the results of this research can provide a practical contribution to training organizers, namely the HR Department in the training and development section, in utilizing the results of evaluating training participants' reactions to determine aspects that need to be improved in training. This research also provides procedures for developing the best model for classifying participant reaction comments so that practitioners gain holistic insight into participant satisfaction and opinions regarding the implementation of training, especially in training with a large number of participants or a high level of implementation frequency. Apart from practical benefits, this research is expected to contribute to gaps in the literature and insight related to text mining on the topic of evaluating participants' post-training reactions. In this study, section II will discuss the research method, section III deals with research results and discussion, and section IV deals with conclusions obtained from this research.

## 2. RESEARCH METHOD

The stages in this research process use the Cross Industry Standard Process for Data Mining (CRISP-DM) model, a data mining methodology prepared by a consortium of companies founded by the European Commission in 1996 and designated as a standard process in data mining. By referring to the fact that data mining in the CRISP-DM model has six phases [23], the stages of this research can be explained in Figure 1.

### 1. Business Understanding

The first stage is to understand the goals and needs from a business perspective, and then translate this knowledge into defining problems in data mining. Next, plans and strategies will be determined to achieve these goals. The problem in this research is how companies can evaluate training using data mining using open comment data from training participants, and the goal the company wants to achieve is to find topics that can be taken into consideration for improving training implementation in the future.

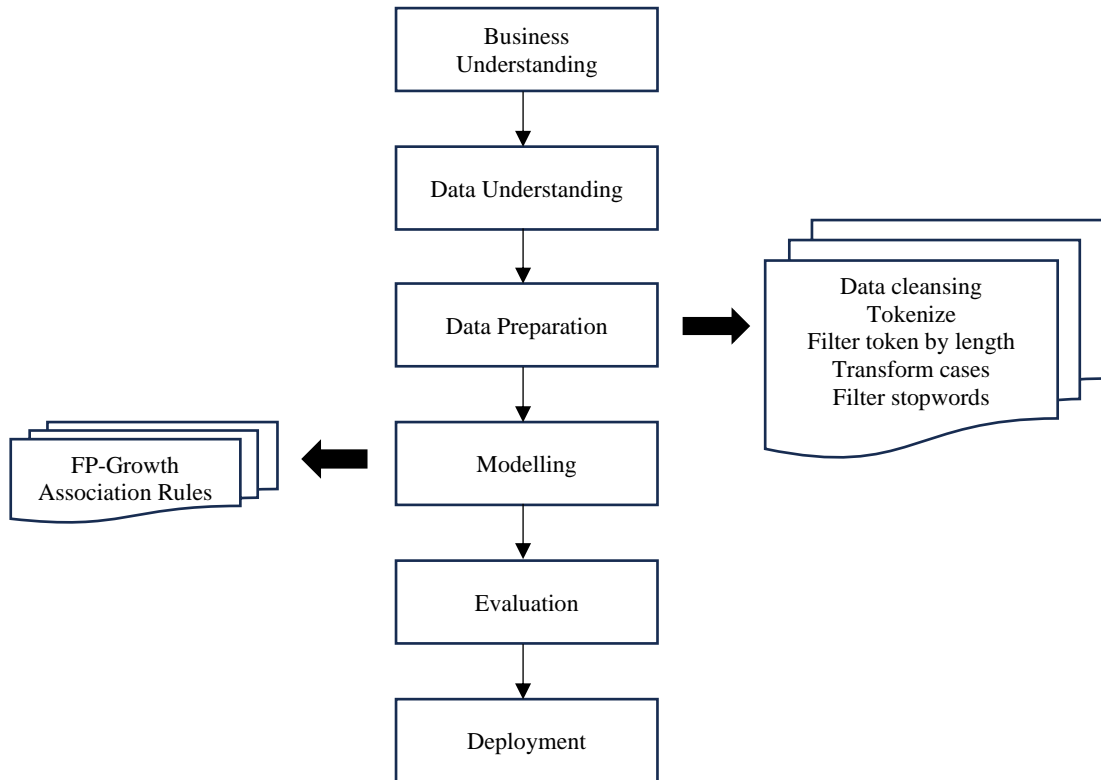
### 2. Data Understanding

This stage begins with data collection which will then continue with a process to gain a deep understanding of the data, identify data quality problems, or detect interesting parts of the data that can be used to hypothesize hidden information. In this research, the data used is open-ended

questionnaire data filled in by employees PT. XYZ as many as 517 respondents after training conducted in Mojokerto, in October 2023.

**Table 1.** Rating Results vs Open Comments Post Training at PT. XYZ

Training Program	Rating (Scale 1-5)	Comments
Team Building & Positive Mindset	4,36 (very good)	More varied material Less relevant material There are several places where the service is lacking, especially the toilets, perhaps they could be added and the transportation is also inadequate



**Figure 1.** Research Method

3. Data Preparation

This stage includes all activities to build the final dataset (data that will be processed in the modeling stage) from raw data. This stage can be repeated several times. This stage also includes selecting tables, records, and data attributes, including the process of cleaning and transforming the data to then be used as input in the modeling stage. At this stage, several sub-stages are carried out as follows:

- a. Data cleansing  
From the data obtained through an open questionnaire, cleaning was carried out first, such as empty comment data and symbols.
- b. Transform case  
In this sub-stage or sub-process, all letters in the text are changed to lowercase, so that it can help with consistency of analysis, reduce the dimensions or size of the data, and prevent unnecessary differences due to differences in lowercase and uppercase types.
- c. Tokenization  
In this sub-stage or sub-process, the data is broken down into smaller units, namely sentences into single words or in the form of fragments of words.
- d. Filter stopwords  
In this sub-stage or sub-process, data in the form of words that are not needed or relevant to the research will be deleted, leaving only words that are important and relevant in this research.
- e. Filter token by length  
In this sub-stage or sub-process, words that have less than four characters are removed.

#### 4. Modeling

At this stage, various modeling techniques will be selected and applied and several parameters will be adjusted to obtain optimal values. Specifically, several different techniques can be applied to the same data mining problem. On the other hand, there are modeling techniques that require special data formats. So at this stage, it is still possible to return to the previous stage. This research uses the FP Growth algorithm and Association Rules.

##### a. FP-Growth

FP-Growth is a frequent itemsets search algorithm obtained from FP-tree by exploring the tree from bottom to top [24]. The FP-Growth algorithm is a development of the a priori algorithm. This algorithm determines frequent itemsets that end in a certain suffix by using the divide and conquer method to break down the problem into smaller sub problems [25]. FP-Growth uses the concept of tree building in frequent itemset searches, not using generate candidates like the a priori algorithm, so it is faster. In this research, using the FP-Growth algorithm, we will obtain the words that appear most frequently in a collection of text data taken from an open questionnaire about training evaluation.

##### b. Association Rules

Association rules is a method that aims to look for patterns that often appear in many transactions, where each transaction consists of several items [26]. The parameters used in the association rules are support and confidence values. Support is a measure that shows the level of itemset dominance of the entire transaction, with the following formula 1.

$$Support = \frac{(X \cup Y) count}{n} \quad (1)$$

Where  $(X \cup Y) count$  is the number of transactions containing x and y, and n is the total transactions. Another parameter is confidence, which is a measure of how valid an Association Rule is or not. An association rule with confidence equal to or greater than the minimum confidence value can be said to be a valid association rule. To determine the confidence value, you can use the following formula 2.

$$Confidence = \frac{(X \cup Y) count}{X.count} \quad (2)$$

Where X count is the number of occurrences of item X in all transactions. Apart from that, the correlation in a rule is not only measured by the support and confidence values but also by the correlation between the two itemsets. One measure of correlation used in research is using the lift value, which can be measured using the following formula 3 [24].

$$Lift(A, B) = \frac{P(A \cup B)}{P(A).P(B)} \quad (3)$$

Where the value is the confidence value of itemsets A with B, and P(A) is the number of transactions containing A. If the lift value is less than 1 then A has a negative correlation with B, which means there is no interrelationship with each other. And if the lift value is greater than 1 then A has a positive correlation with B. Meanwhile, if the lift value is equal to 1 then A and B are independent.

##### c. Evaluation

At this stage, the model has been formed and is expected to have the good quality from a data analysis point of view. At this stage, an evaluation will be carried out on the effectiveness and quality of the model before use and determine whether the model can achieve the goals set at the initial stage (Business Understanding).

##### d. Deployment

At this stage, the knowledge or information that has been obtained will be organized and presented in a special form so that it can be used by users. The deployment stage can take the form of creating a simple report or implementing a repeatable data mining process within the company. In many cases, the deployment phase involves consumers, in addition to data

analysts, because it is critical for consumers to understand what actions must be taken to use the model that has been created.

### 3. RESULTS AND ANALYSIS

Of the 517 datasets used by researchers to conduct research related to the analysis of post-training open comments at PT. XYZ using the CRISP-DM model, obtained the following research results:

#### 3.1. Word Cloud

The Wordlist results that appear most frequently in the open questionnaire data collection are presented in Figure 2.

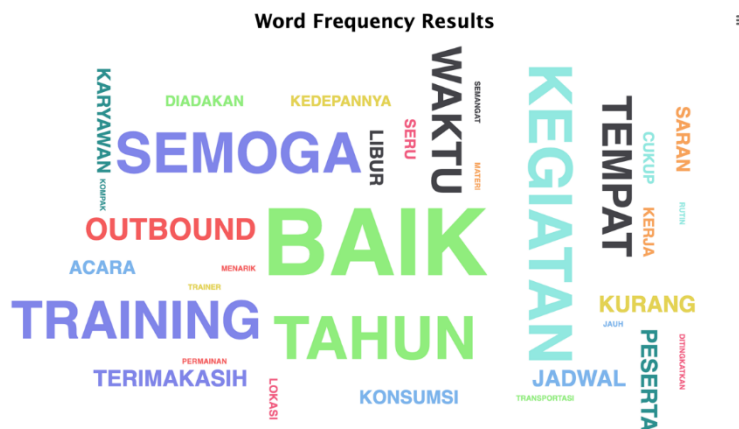


Figure 2. Word cloud work results appear most frequently

From these results it was found that in the data collection extracted through an open post-training questionnaire, the word "BAIK" was the one that appeared most frequently. This can be seen in the letters of the word being printed at the largest size. It can be concluded that the implementation of the training is generally assessed good by the training participants, apart from that the word "KURANG" also appears in this set of documents, although the number is smaller (smaller font size) compared to the word "BAIK", this shows that there is still another focus that is assessed by the training participants who still need to be improved, for example "KONSUMSI", "TRANSPORTASI" which also appeared so there needs to be an increase in these factors in the future. Apart from that, there are also other concerns such as "LOKASI" and "JADWAL", as well as the hope of the training participants to hold more training in the future with the appearance of the words "SEMOGA" and "KEDEPANNYA".

#### 3.2. FP Growth Result

With a minimum support value = 0.05 or 5%, the results obtained from the FP Growth algorithm are related to the words or topics that appear most often in open comments related to post-training evaluation, as table 2.

Table 2. FrequentItemSets (FP Growth) Results

Support Value	Item 1	Item 2
0.190	Baik	
0.164	Tahun	
0.162	Semoga	
0.139	Kegiatan	
0.135	Training	
0.077	Tempat	
0.074	Waktu	
0.068	Tahun	Semoga
0.062	Kurang	
0.056	Konsumsi	
0.054	Saran	
0.054	Bagus	
0.050	Seru	
0.050	Libur	

From the FP Growth results, it was found that the word that appeared most frequently in open comments after the training was "Baik", with a support value of 0.190 or 19%.

### 3.3. Association Rules Results

With minimum support of 2% and a confidence value of 50%, the association relationship values related to the itemsets that appear most frequently are related to post-training suggestions or comments conducted in October 2023, as table 3.

**Table 3.** Association rules based on support value (Top 8)

Support Value	Premises	Conclusion
0.033	Libur	Jadwal
0.033	Jadwal	Libur
0.031	Kedepannya	Semoga
0.029	Libur	Dinas
0.029	Dinas	Libur
0.023	Cukup	Baik
0.023	Tahun, Kegiatan	Semoga
0.023	Transportasi	Baik

From the results of association rules based on support values as in Table 3, the following are obtained, the word "Libur" is related to the word "Jadwal" with a support value of 0.033 or 3.3%. And the word "Kedepannya" is related to the word "Semoga" with a support value of 0.031 or 3.33%. The word "Libur" is related to the word "Dinas" with a support value of 0.029 or 2.9%. And the word "Cukup" with the word "Baik", the words "Tahun, Kegiatan" with the word "Semoga", and the word "Transportasi" with the word "Baik", each of which is related to the same support value, namely 0.023 or 2.3 %.

**Table 4.** Association rules based on confidence value (Top 8)

Confidence Value	Premises	Conclusion
1	Dinas	Libur
1	Jadwal, Dinas	Libur
0.800	Transportasi	Baik
0.733	Dinas	Jadwal
0.733	Dinas	Libur, Jadwal
0.733	Libur, Dinas	Jadwal
0.708	Jadwal	Libur
0.706	Tahun, Kegiatan	Semoga

From the results of association rules based on confidence values as in table 4, the following are obtained: The word "Dinas" is related to the word "Libur", and the words "Jadwal, Dinas" are also related to the word "Libur", with a confidence value of 1 or 100%. And the word "Transportasi" is related to the word "Baik" with a confidence value of 0.800 or 80%, and the word "Dinas" is related to the word "Jadwal", and to the words "Libur, Jadwal" with the same confidence value of 0.733 or 73.3%. The word "Jadwal" is related to the word "Libur" with a confidence value of 0.708 or 70.8%, and the words "Tahun, Kegiatan" are related to the word "Semoga" with a confidence value of 0.706 or 70.6%. From the confidence value, it is found that each word is related to each other with a fairly high confidence value.

**Table 5.** Association rules based on lift value (Top 8)

Lift Value	Premises	Conclusion
22.302	Libur, Jadwal	Dinas
22.302	Dinas	Libur, Jadwal
19.885	Libur	Dinas
19.885	Dinas	Libur
19.885	Jadwal, Dinas	Libur
15.797	Libur, Dinas	Jadwal
15.797	Libur	Jadwal
14.085	Libur, Jadwal	Dinas

From the results of the association rules based on the lift value as in Table 5, the following are obtained: The word "Libur, Jadwal" has a relationship or association with the word "Dinas" with the highest lift value of 22,302. And the word "Libur" has a relationship or association with the word "Dinas" with lift value of 19,885. The words "Jadwal, Dinas" have a relationship or association with the word "Libur" with a lift value of 19,885. And the word "Libur, Dinas" has a relationship or association with the word "Jadwal" with lift value of 15,797. The word "Libur" has a relationship or association with the word "Jadwal" with a lift value

of 15,797. And the word "Jadwal" has a relationship or association with the word "Libur" with a lift value of 14,085. Because the lift value is  $> 1$ , it can be said that each of these words has a strong association relationship. The following is the graph visualization results for association rules as a figure 3.

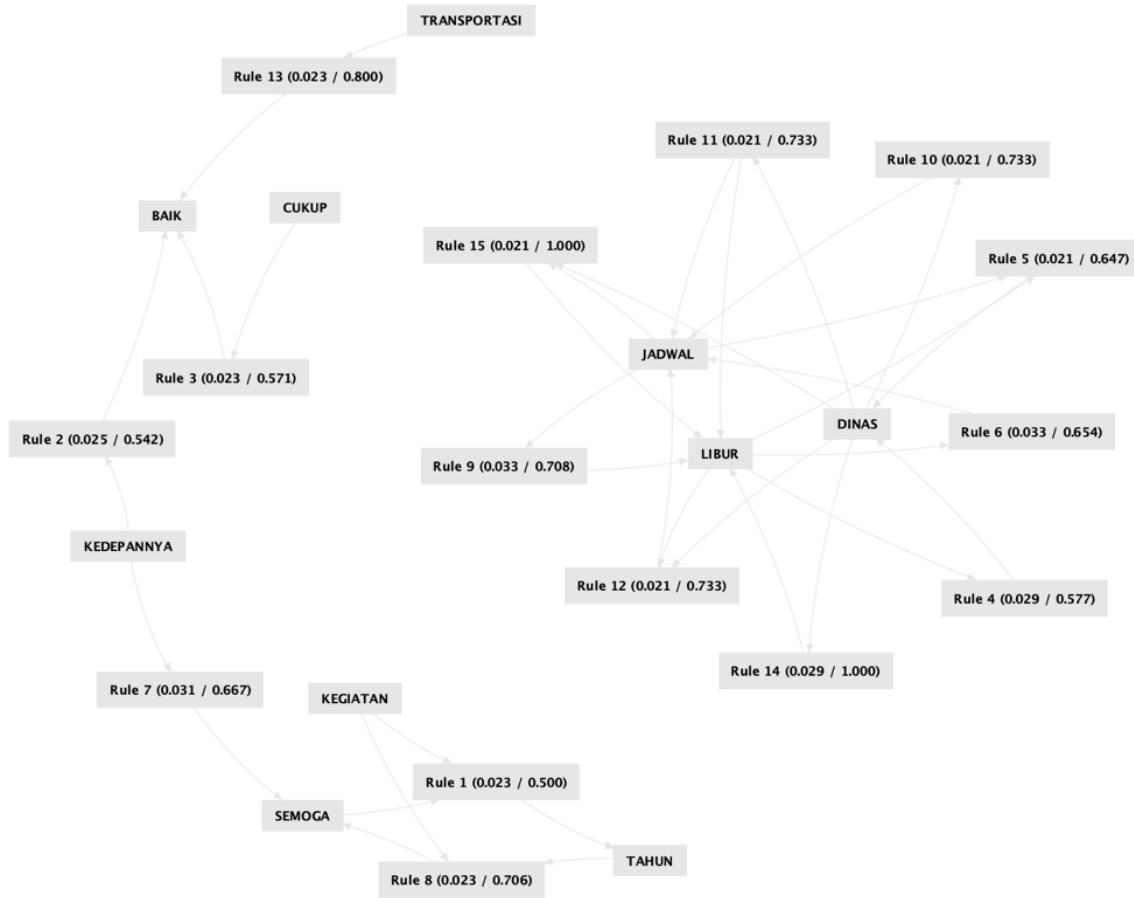


Figure 3. Association rules visualization results

From Figure 3 it can be seen that there are 15 association rules as follows: Rule 1: The relationship between the words "Semoga, Kegiatan" and the word "Tahun" with a support value of 0.023 and a confidence value of 0.500 or 50%. Rule 2: The relationship between the word "Kedepannya" and the word "Baik" with a support value of 0.025 and a confidence value of 0.542 or 54.2%. Rule 3: The relationship between the word "Cukup" and the word "Baik" with a support value of 0.023 and a confidence value of 0.571 or 57.1%. Rule 4: The relationship between the word "Libur" and the word "Dinas" with a support value of 0.029 and a confidence value of 0.577 or 57.7%. Rule 5: The relationship between the words "Jadwal, Libur" and the word "Dinas" with a support value of 0.021 and a confidence value of 0.647 or 64.7%. Rule 6: The relationship between the word "Libur" and the word "Jadwal" with a support value of 0.033 and a confidence value of 0.654 or 65.4%. Rule 7: The relationship between the word "Kedepannya" and the word "Semoga" with a support value of 0.031 and a confidence value of 0.667 or 66.7%. Rule 8: The relationship between the words "Tahun, Kegiatan" and the word "Semoga" with a support value of 0.023 and a confidence value of 0.706 or 70.6%. Rule 9: The relationship between the word "Jadwal" and the word "Libur" with a support value of 0.033 and a confidence value of 0.708 or 70.8%. Rule 10: The relationship between the word "Dinas" and the word "Jadwal" with a support value of 0.021 and a confidence value of 0.733 or 73.3%. Rule 11: The relationship between the words "Jadwal, Libur" and the word "Dinas" with a support value of 0.021 and a confidence value of 0.733 or 73.3%. Rule 12: The relationship between the words "Libur, Dinas" and the word "Jadwal" with a support value of 0.021 and a confidence value of 0.733 or 73.3%. Rule 13: The relationship between the word "Transportasi" and the word "Baik" with a support value of 0.023 and a confidence value of 0.800 or 80%. Rule 14: The relationship between the word "Dinas" and the word "Libur" with a support value of 0.029 and a confidence value of 1,000 or 100%. Rule 15: The relationship between the words "Jadwal, Dinas" and the words "Libur" with a support value of 0.021 and a confidence value of 1,000 or 100%.

From the results of the analysis of open comments regarding training evaluations carried out by PT. XYZ using text mining analysis, with the FP Growth and Association Rules algorithm, it was found that the

majority of training participants felt satisfied in carrying out training activities, this can be seen in table 3.1 where the word "Baik" is the word that appears most often, several training participants also have hope that future training can be carried out again. Apart from that, several suggestions for improvement, such as implementation so that it is not carried out during official holidays, are the main things that can be input for the company for the next training activity because the words "Dinas", "Libur" and "Jadwal" are words that have a strong association. with a high lift value. This can certainly be a positive input and contribution for the company to make continuous improvements, especially in the field of training in the future.

#### 4. CONCLUSION

In this study, researchers analyzed post-training evaluations through open comments in the form of a questionnaire survey in Indonesian. Text analysis was applied to 516 comments obtained from training participants via an open-ended questionnaire. From text analysis, 15 association rules were obtained using Rapidminer with the FP-Growth algorithm with a minimum support of 0.02 and a minimum confidence of 0.5. As shown in Table 3.4, all association rules have a lift value greater than 1, which indicates that all of them are valid or have a strong association relationship. So that this research can find out the pattern of comments or suggestions given by relevant training participants that can make a positive contribution to the implementation of training, organizations, or companies.

In the future, this research can be developed not only regarding training evaluations carried out after participants have undergone training but can also be implemented in pre-training where participants before undergoing training, besides that it is possible to apply different support and confidence values to a larger data set to ensure the consistency of the results reported in the research remains valid. The results of this research can be an excellent reference for organizations or companies in conducting training evaluations on various existing topics by utilizing input and expectations of training participants in the form of text or open questionnaires, so that they can provide more comprehensive improvements for organizations or companies.

#### REFERENCES

- [1] A. Elnaga and A. Imran, 'The Effect of Training on Employee Performance', 2013. [Online]. Available: [www.iiste.org](http://www.iiste.org)
- [2] N. P. Z. Thwin, Y. N. Soe, A. Bhaumik, and Midhunchakkarayarthi, 'Impact of Training and Development on Employee Performance: A Study of Myanmar Internet Service Provider Industry', *Proceedings on Engineering Sciences*, vol. 4, no. 3, pp. 291–300, 2022, doi: 10.24874/PES04.03.007.
- [3] A. Cik, M. Asdar, A. I. Anwar, and S. Efendi, 'Impact of Training and Learning Organization on Employee Competence and Its Implication on Job Satisfaction and Employee Performance of Bank in Indonesia', 2021. [Online]. Available: [www.psychologyandeducation.net](http://www.psychologyandeducation.net)
- [4] K. Nama, B. Daweti, M. Lourens, and T. Chikukwa, 'The impact of training and development on employee performance and service delivery at a local municipality in South Africa', *Problems and Perspectives in Management*, vol. 20, no. 4, pp. 42–51, 2022, doi: 10.21511/ppm.20(4).2022.04.
- [5] M. S. Nawaz, M. Hassan, S. Hassan, S. Shaukat, and M. A. Asadullah, 'Impact of employee training and empowerment on employee creativity through employee engagement: Empirical evidence from the manufacturing sector of Pakistan', *Middle East J Sci Res*, vol. 19, no. 4, pp. 593–601, 2014, doi: 10.5829/idosi.mejsr.2014.19.4.13618.
- [6] M. Ali Jareko, T. Mushtaque, and H. Ali Mahesar, 'Impact of Working Environment and Training & Development on Organization Performance Through Mediating Role of Employee Engagement and Job Satisfaction', 2017. [Online]. Available: <https://www.researchgate.net/publication/316911148>
- [7] M. A. Memon, R. Sallaeh, M. N. R. Baharom, S. Md Nordin, and H. Ting, 'The relationship between training satisfaction, organizational citizenship behaviour, and turnover intention: A PLS-SEM approach', *Journal of Organizational Effectiveness*, vol. 4, no. 3, pp. 267–290, Sep. 2017, doi: 10.1108/JOEPP-03-2017-0025.
- [8] C. Priyana and D. Aseanty, 'The Effect of Training Satisfaction to Turnover Intention Mediated by Organizational Citizenship Behavior and Job Satisfaction in Employees of Bank Syariah in Jakarta', *Webology*, vol. 17, no. 2, pp. 551–567, Dec. 2020, doi: 10.14704/WEB/V17I2/WEB17051.
- [9] M. Rubel and A. Rahman, 'Effect of Training and Development on Organizational Citizenship Behavior (OCB): An Evidence from Private Commercial Banks in Bangladesh Effect of Training and Development on Organizational Citizenship Behavior (OCB): An Evidence from Private Commercial Banks in Bangladesh Effect of Training and Development on Organizational Citizenship Behavior (OCB): An Evidence from Private Commercial Banks in Bangladesh', *Type: Double Blind Peer Reviewed International Research Journal Publisher: Global Journals Online Year*, vol. 18, 2018, [Online]. Available: <https://www.researchgate.net/publication/332058791>
- [10] C. Jaworski, S. Ravichandran, A. C. Karpinski, and S. Singh, 'The effects of training satisfaction, employee benefits, and incentives on part-time employees' commitment', *Int J Hosp Manag*, vol. 74, pp. 1–12, Aug. 2018, doi: 10.1016/j.ijhm.2018.02.011.



- [11] E. Ocen, K. Francis, and G. Angundaru, 'The role of training in building employee commitment: the mediating effect of job satisfaction', *European Journal of Training and Development*, vol. 41, no. 9, pp. 742–757, 2017, doi: 10.1108/EJTD-11-2016-0084.
- [12] Q. Albtosh, A. H. Ngah, and Y. M. Yusoff, 'Training satisfaction relative to turnover intention: the mediating role of employee loyalty', *Industrial and Commercial Training*, 2022, doi: 10.1108/ICT-06-2021-0047.
- [13] J. Paek, 'The Effect of Education and Training on Service Quality, Customer Satisfaction and Loyalty in Airline Industry', 2018.
- [14] I. Buonomo, M. Piccinini, P. Benevene, G. Blasutig, and S. Cervai, 'Job training satisfaction and knowledge sharing in IT company: a case study', *Journal of Workplace Learning*, vol. 34, no. 8, pp. 677–690, Oct. 2022, doi: 10.1108/JWL-02-2022-0016.
- [15] J. Ben Mansour, A. Naji, and A. Leclerc, 'The relationship between training satisfaction and the readiness to transfer learning: The mediating role of normative commitment', *Sustainability (Switzerland)*, vol. 9, no. 5, May 2017, doi: 10.3390/su9050834.
- [16] F. Vencovský, 'Service Quality Evaluation Using Text Mining: A Systematic Literature Review', in *Lecture Notes in Business Information Processing*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 159–173. doi: 10.1007/978-3-030-61140-8\_11.
- [17] M. Zibarzani *et al.*, 'Customer satisfaction with Restaurants Service Quality during COVID-19 outbreak: A two-stage methodology', *Technol Soc*, vol. 70, Aug. 2022, doi: 10.1016/j.techsoc.2022.101977.
- [18] Bieman Chris, Handschuh Siegfried, Freitas Andre, Meziane Farid, and Metais Elisabeth, 'Natural Language Processing and Information Systems', 2015. doi: 10.1007/978-3-319-19581-0.
- [19] M. Zhang, B. Fan, N. Zhang, W. Wang, and W. Fan, 'Mining product innovation ideas from online reviews', *Inf Process Manag*, vol. 58, no. 1, Jan. 2021, doi: 10.1016/j.ipm.2020.102389.
- [20] F. V. Ordenes, B. Theodoulidis, J. Burton, T. Gruber, and M. Zaki, 'Analyzing Customer Experience Feedback Using Text Mining: A Linguistics-Based Approach', *J Serv Res*, vol. 17, no. 3, pp. 278–295, 2014, doi: 10.1177/1094670514524625.
- [21] R. B. Bahaweres and D. A. Nugrahanti, 'Implementation of Text Association Rules about Terrorism on Twitter in Indonesia', in *2022 10th International Conference on Cyber and IT Service Management, CITSM 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/CITSM56380.2022.9935864.
- [22] H. Margono, X. Yi, and G. K. Raikundalia, 'Mining Indonesian Cyber Bullying Patterns in Social Networks', 2014. [Online]. Available: <http://www.tweetadder.com/download>
- [23] D. T. Larose, 'Data Mining Methods and Model', 2006.
- [24] P.-N. Tan, M. Steinbach, and V. Kumar, 'Introduction to Data Mining Instructor's Solution Manual', 2006.
- [25] J. Han, M. Kamber, and J. Pei, 'Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)', 2011.
- [26] S. Zhang, C. Zhang, and Q. Yang, 'Data Preparation for Data Mining', *Applied Artificial Intelligence*, vol. 17, no. 5–6, pp. 375–381, May 2003, doi: 10.1080/713827180.

## BIBLIOGRAPHY OF AUTHORS



Mas Akhmad Nainunis ([mas.akhmad.nainunis-2022@pasca.unair.ac.id](mailto:mas.akhmad.nainunis-2022@pasca.unair.ac.id)) was born in Gresik. I am a student at Airlangga University, Post Graduate School, Majoring Human Resources Development with Data Analytic concentration. I am interested with the data science, so my research focus is related to machine learning, data mining and deep learning. I also work in the manufacturing industry as human resources, so the research really helps me not only as an academic but also a practitioner.



Imam Yuadi ([imam.yuadi@fisip.unair.ac.id](mailto:imam.yuadi@fisip.unair.ac.id)) is an Associate Professor in the Department of Information and Library Science, at Universitas Airlangga. His doctoral degree was from Information Management, National Chiao Tung University Taiwan in 2017. He has been conducting several kinds of research about digital libraries, digital forensics, bibliometric analysis, and data science.