

Research Article

# Analysis of Generation Z Parenting Styles and Children's Educational Awareness Using Decision Tree and K-Means Methods

Novi Syahfitri\*, Sudi Suryadi, Budianto Bangun, Angga Putra Juledi

Faculty of Science and Technology, Universitas Labuhanbatu, Sumatera Utara, Indonesia

\*Corresponding Author: [novisyahfitri522@gmail.com](mailto:novisyahfitri522@gmail.com) | Phone: +6282174879322

## ABSTRACT

This study analyzes how Generation Z parents' parenting styles and technology supervision influence children's educational awareness. The research applied a data mining approach within the Knowledge Discovery in Databases (KDD) framework using Decision Tree and K-Means algorithms. Data were collected through questionnaires from 10 parents in Sumberjo Village. The Decision Tree results show that technology use and supervision provide the highest information gain, indicating that they are the most influential factors in determining children's educational awareness. K-Means clustering with  $K = 3$  also shows that groups characterized by better technology supervision and stronger parenting patterns tend to have higher educational awareness. Validation using RapidMiner produced results that were consistent with the manual calculations, confirming that the analytical model is valid for the dataset used in this study.

**Keywords:** Parenting style; Generation Z; Educational awareness; Decision Tree; K-Means

## 1. INTRODUCTION

The rapid development of information and communication technology has changed the way people communicate, learn, and carry out everyday activities (Singgalen, 2021). These changes influence not only economic and social life but also family life and the way parents' guide children in the learning process (Kharis et al., 2023). In the digital era, parenting is increasingly connected with the use of technological devices, online information, and social media (Singgalen, 2021). As a result, parents are expected not only to provide care and affection, but also to manage children's digital behavior wisely (Kharis et al., 2023).

Generation Z grew up in a highly connected environment and is often described as a digital-native generation (Kharis et al., 2023). Parents who belong to or are closely associated with this digital environment face distinctive challenges when guiding children (Kharis et al., 2023). Easy access to smartphones, entertainment platforms, and online information can support learning, but it can also reduce direct interaction and supervision if not managed carefully (Atalya Angelus Leza et al., 2024). For this reason, parenting style and technology supervision are expected to play an important role in shaping children's educational awareness.

Educational awareness refers to a child's understanding of the value, goals, and benefits of learning activities (Atalya Angelus Leza et al., 2024). Children with stronger educational awareness usually show greater curiosity, more discipline, and better motivation to participate in formal and informal learning (Atalya Angelus Leza et al., 2024). In family settings, this awareness is influenced by communication patterns, parental involvement, and the consistency of guidance provided at home (Maruli Tua Silaen, 2023). Therefore, the quality of parenting remains essential even when children have broad access to digital resources.

This study aims to identify the parenting-related factors that influence children's educational awareness by using Decision Tree and K-Means methods (Hartati et al., 2021). The Decision Tree algorithm is employed to determine the most influential attribute, while K-Means is used to group respondents based on similar parenting and technology-supervision characteristics. The study focuses on questionnaire data collected from parents in Sumberjo Village.

## 2. RESEARCH METHOD

### 2.1 Research Design and Analytical Framework

This research used a data mining approach based on the Knowledge Discovery in Databases (KDD) framework (Purwati et al., 2023; Alam et al., 2022). The stages consisted of problem identification, data collection, data preprocessing, Decision Tree analysis, K-Means clustering, and interpretation of the results (Purwati et al., 2023; Kharis et al., 2023). Questionnaire data were gathered from parents with school-age children in Sumberjo Village. The dataset was then prepared for classification and clustering analysis (Purwati et al., 2023).

Decision Tree was applied to determine which attribute most strongly influenced children's educational awareness (Informatika et al., 2021). In this study, entropy and information gain were used to compare the contribution of parenting style and technology supervision (Informatika et al., 2021). K-Means clustering was used to group respondents based on similar characteristics so that patterns in parenting and educational awareness could be interpreted more clearly (Fatonah & Pancarani, 2022).



**Figure 1.** Knowledge Discovery in Databases (KDD)

Source: Research, 2025

Knowledge Discovery in Database (KDD) is a comprehensive process that covers a series stage for extracting information meaningful from large data sets (Purwati et al., 2023). This process starts with data selection, data cleaning, transformation, and the main stage, namely data mining, which aims to find pattern, connection, or structure hidden Which previously unidentified (Kharis et al., 2023).

### 2.2 Population, Sample, and Variables

The population in this study consisted of parents with children in Sumberjo Village. The sample used in the analysis contained 10 questionnaire responses. The main variables were parenting style, technology use and supervision, and children's educational awareness. In the Decision Tree model, parenting style and technology supervision acted as predictor variables, whereas children's educational awareness was the target variable (Informatika et al., 2021). In the K-Means model, the same attributes were transformed into numerical form for clustering purposes (Hartati et al., 2021).

**Table 1.** Sample data used in the study.

No	Child's Name	Parenting Style	Technology Use & Supervision	Educational Awareness
1	Ahmad Rizky	Low	Controlled	Good
2	Siti Aulia	High	Not controlled	Good
3	Muhammad Farhan	High	Controlled	Insufficient
4	Nabila Zahra	High	Controlled	Good
5	Andi Pratama	High	Controlled	Good
6	Putri Maharani	High	Controlled	Good
7	Fajar Nugroho	High	Not controlled	Insufficient
8	Dinda Ayuningtyas	Low	Controlled	Insufficient
9	Rafi Akbar	Low	Not controlled	Good
10	Alya Ramadhani	Low	Not controlled	Insufficient

Source: Research, 2025

The **Table 1** represents sample data for the research, with a total of 10 children. The data has been structured and includes three main attributes: parenting style, technology use and supervision, and children's educational awareness. Each data set represents a different situation, thus demonstrating a variety of parenting patterns and diverse technology use.

### 2.3 Data Transformation and Validation

For K-Means clustering, categorical variables were converted into numerical values: parenting style (High = 1, Low = 0), technology supervision (Controlled = 1, Not controlled = 0), and educational awareness (Good = 1, Insufficient = 0). The number of clusters was set to  $K = 3$  to identify three dominant parenting patterns. RapidMiner was used as a supporting tool to validate the manual calculations and visualize the resulting classification and clustering processes (Rafi Nahjan et al., 2023).

## 3. RESULTS AND DISCUSSION

### 3.1 Decision Tree Analysis

The target variable, children's educational awareness, consisted of 6 observations in the good category and 4 observations in the insufficient category, which produced a total entropy value of 0.971. Information gain was then calculated for each predictor attribute in accordance with the Decision Tree procedure (Informatika et al., 2021). The results show that technology use and supervision provided a higher information gain than parenting style, indicating that it is the most influential attribute in the dataset.

Because technology supervision produced the highest information gain, it became the root node of the Decision Tree model. This result indicates that the quality of supervision over children's technology use plays a more decisive role than parenting style alone. Nevertheless, parenting style still acts as a supporting factor that strengthens positive educational awareness when direct supervision is also present.

$$\begin{aligned} &\text{Information Gain:} \\ &\text{Gain (S, A)} = \text{Entropy (S)} - \sum \left( \frac{S_i}{S} \times \text{Entropy (S}_i) \right) \\ &\text{Gain (S, Style Parenting)} = 0,971 - 0,951 = 0,020 \\ &\text{Gain (S, Supervision Technology)} = 0.971 - 0.878 = 0.093 \end{aligned}$$

**Table 2.** Information gain summary for the Decision Tree model.

Attribute	Information Gain
Parenting Style	0.020
Technology Use & Supervision	0.093

Source: Research, 2025

The Information Gain value, the Technology Usage & Monitoring attribute has the highest gain value (0.093) compared to Parenting Style (0.020). Thus, the Technology Usage & Supervision attribute is worthy of being selected as the root node (root of the decision tree) in the Decision Tree model.

### 3.2 K-Means Clustering Results

The K-Means process grouped the data into three clusters based on parenting style, technology supervision, and educational awareness. The results show that Cluster 1 represents respondents with strong parenting and controlled technology supervision, Cluster 2 represents low parenting but still controlled technology supervision, and Cluster 3 represents weak technology supervision with relatively lower educational-awareness outcomes.

These results reinforce the Decision Tree findings. Cluster patterns indicate that controlled technology supervision tends to be associated with better educational awareness, especially when accompanied by stronger parenting practices (Hartati et al., 2021; Fatonah & Pancarani, 2022). In contrast, limited supervision is more closely associated with suboptimal educational awareness, even when some supportive parenting traits are present.

The implementation of the K-Means algorithm shows that technology use and supervision become the main factors shaping the cluster structure (Hartati et al., 2021). Children who are in the environment with supervision good technology tend own awareness more education high, especially when supported by style optimal parental care.

**Table 3.** K-Means cluster assignment results.

No	Child's Name	Parenting Style	Tech Supervision	Awareness	Cluster
1	Ahmad Rizky	Low	Controlled	Good	C2
2	Siti Aulia	High	Not controlled	Good	C3
3	Muhammad Farhan	High	Controlled	Insufficient	C1
4	Nabila Zahra	High	Controlled	Good	C1
5	Andi Pratama	High	Controlled	Good	C1
6	Putri Maharani	High	Controlled	Good	C1
7	Fajar Nugroho	High	Not controlled	Insufficient	C3
8	Dinda Ayuningtyas	Low	Controlled	Insufficient	C2
9	Rafi Akbar	Low	Not controlled	Good	C3
10	Alya Ramadhani	Low	Not controlled	Insufficient	C3

Source: Research, 2025

Based on the **Table 3**, the research data that has been transformed into numerical form can be seen in the following table.

**Table 4.** Centroid interpretation of each K-Means cluster.

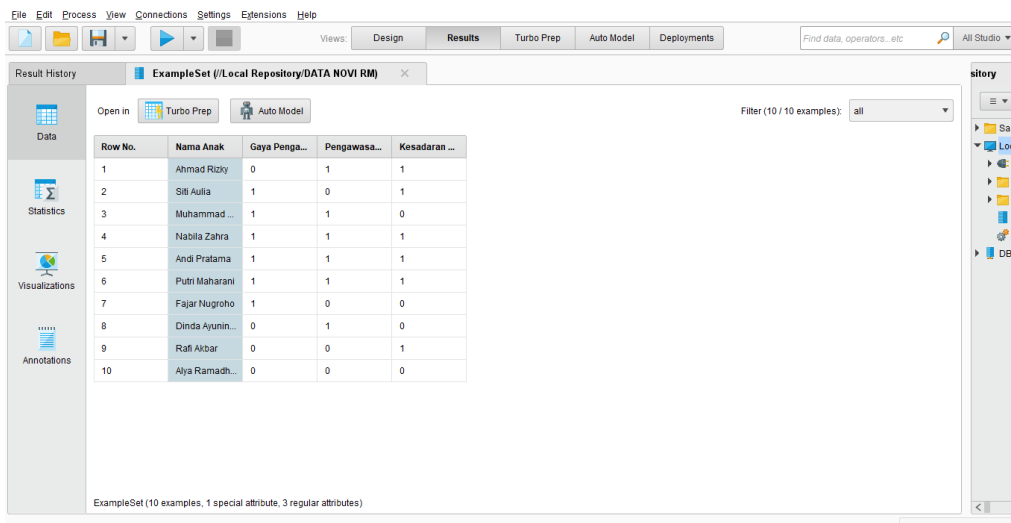
Cluster	Parenting Style (X1)	Tech Supervision (X2)	Awareness (X3)	Interpretation
C1	1.00	1.00	0.75	Strong parenting, controlled technology, higher awareness
C2	0.00	1.00	0.50	Low parenting, but supervision remains controlled
C3	0.25	0.00	0.50	Weak technology supervision, awareness not optimal

Source: Research, 2025

### 3.3 RapidMiner Validation

RapidMiner was used to validate the analytical results obtained manually. The software-generated outputs showed the same general patterns as the manual calculations. In the Decision Tree model, technology use and supervision appeared as the root node. In the Decision Tree model, technology use and supervision appeared as the root node, which is consistent with the logic of information gain-based classification (Informatika et al., 2021).

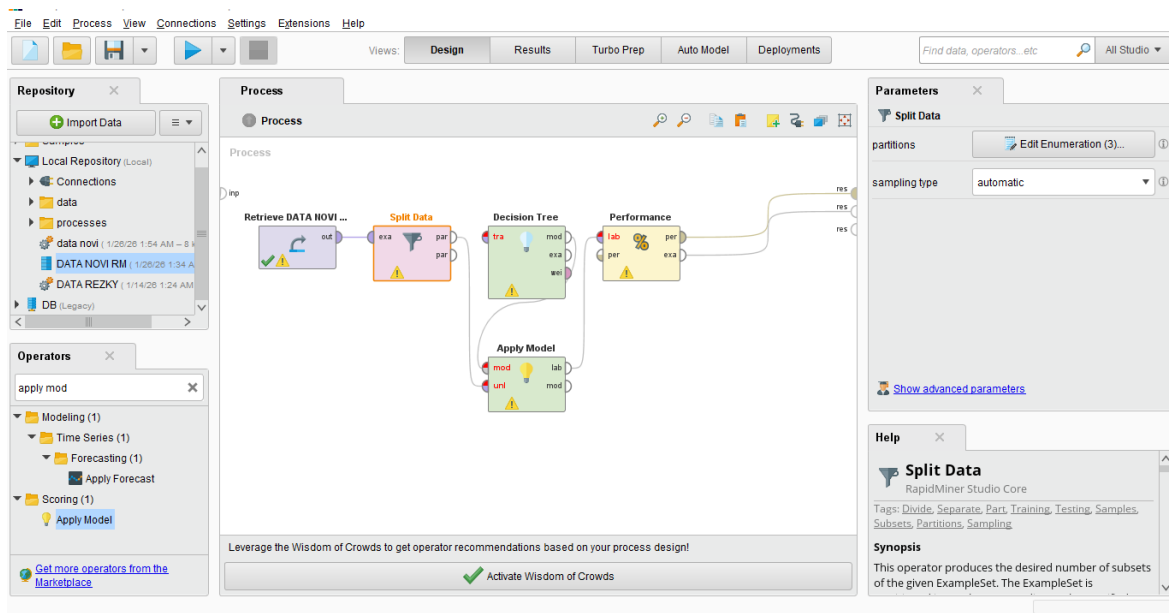
The small differences observed in software-generated centroid values can be explained by the automatic iteration process performed by RapidMiner. However, these minor variations do not change the substantive interpretation of the results. Therefore, the use of RapidMiner supports the validity and consistency of the model built from the dataset.



**Figure 2.** RapidMiner dataset after numerical transformation

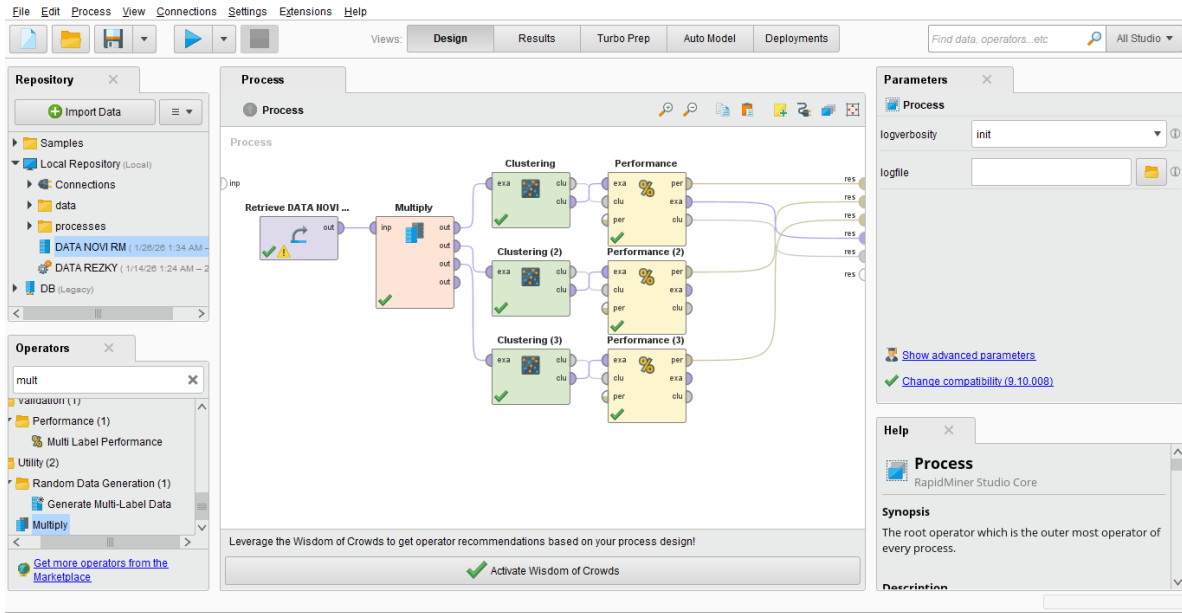
Source: Research, 2025

The results of data processing using RapidMiner shows consistent output with manual calculations using Microsoft Excel. Similarities result This seen good in implementation algorithm Decision Tree and K-Means, which uses the same data, variables, and parameters.



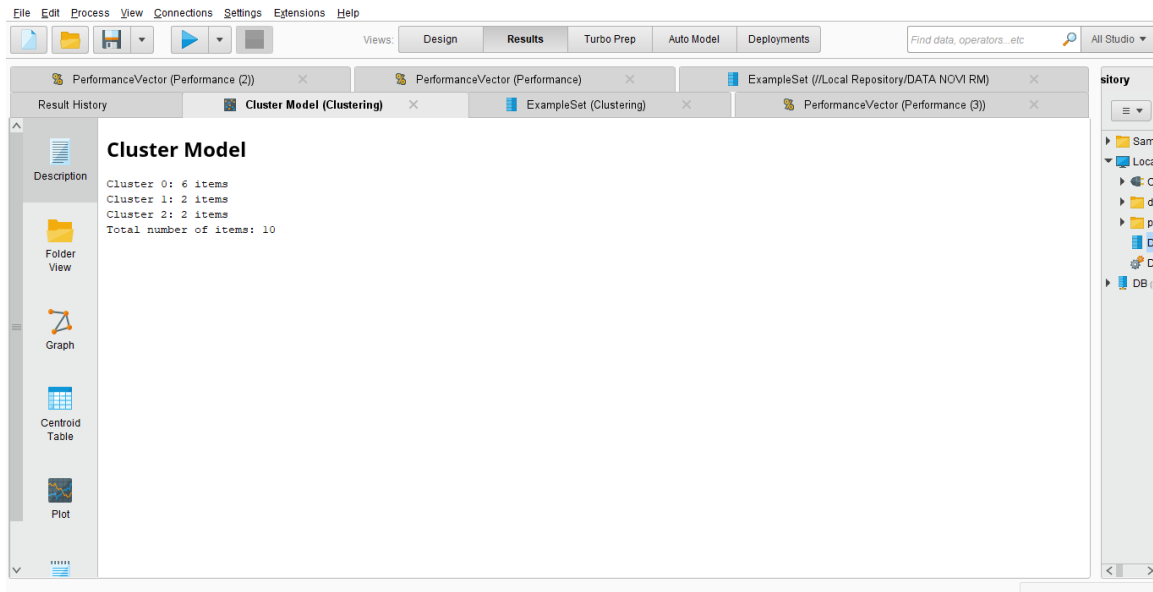
**Figure 3.** Decision Tree process in RapidMiner  
Source: Research, 2025

In Decision Tree, RapidMiner automatically automatic count entropy and information gain values based on input data. Modeling results show that variables Use and Supervision Technology selected as root node, because own highest information gain value. Findings This in accordance with results manual calculations in Excel which also show that variables the most influential to Children's Education Awareness. With Thus, the structure tree the decisions generated by RapidMiner strengthen results manual analysis.



**Figure 4.** K-Means process in RapidMiner  
Source: Research, 2025

In K-Means clustering, RapidMiner produced a distribution of clusters and centroid values that were aligned with the Excel calculations (Rafi Nahjan et al., 2023). Temporary that, in K-Means Clustering, RapidMiner produces distribution clusters and centroid values are aligned with results Excel calculation. Amount cluster ( $K = 3$ ), members each cluster, as well as centroid characteristics show the same pattern, namely that cluster with supervision good technology tend own awareness education older children height. Difference the centroid value that appears only nature small and caused by the iteration process automatic in RapidMiner, however No change interpretation results in a way overall.



**Figure 5.** Cluster model output in RapidMiner  
Source: Research, 2025

In a way overall, the similarity between the RapidMiner results and the Excel calculations shows that the process of data transformation, calculation, and modeling was carried out correctly (Poerwandono & Perwitosari, 2023). RapidMiner plays a role as tool help that validates results manual calculations, at the same time make things easier visualization and interpretation data patterns. Therefore, that is, the use of RapidMiner in study This can stated valid, accurate, and supportive results analysis obtained via Excel.

## 4. CONCLUSION

This study shows that technology use and supervision are the most influential factors in determining children's educational awareness in the analyzed dataset. The Decision Tree model produced a higher information gain for technology supervision than for parenting style, indicating that the management of children's digital activities is central to the classification outcome. The K-Means results also demonstrate that better technology supervision is associated with higher educational awareness, particularly when it is combined with stronger parenting practices. RapidMiner validation confirms that the overall patterns generated by the manual calculations are consistent and reliable (Rafi Nahjan et al., 2023). Because the sample size is limited to 10 observations, the findings should be interpreted as exploratory and may be strengthened by future studies with larger datasets and additional variables.

## ACKNOWLEDGEMENTS

The authors express gratitude to the respondents in Sumberjo Village and to Labuhanbatu University for supporting this study.

## REFERENCES

- AT, H., AA, E.-SM, & Eid, M. M. (2023). A comparative analysis of methods for detecting and diagnosing breast cancer based on data mining. *Journal of Artificial Intelligence and Metaheuristics*, 4(2), 8-17. <https://doi.org/10.54216/jaim.040201>
- Adedoyin, F. F., Satrovic, E., & Kehinde, M. N. (2022). The anthropogenic consequences of energy consumption in the presence of uncertainties and complexities: Evidence from World Bank income clusters. *Environmental Science and Pollution Research*, 29(16), 23264-23279. <https://doi.org/10.1007/s11356-021-17476-5>
- Alam, S., Ghofur, M., & Masripah, N. (2022). Classification of COVID-19 vaccine screening data with the Naive Bayes algorithm using the Knowledge Discovery in Database method. *Journal of Computer Networks, Architecture and High Performance Computing*, 4(2), 177-185. <https://doi.org/10.47709/cnahpc.v4i2.1584>
- Alsaadi, E. M. T. A., Khlebus, S. F., & Alabaichi, A. (2022). Identification of human resource analytics using machine learning algorithms. *Telkomnika (Telecommunication Computing Electronics and Control)*, 20(5), 1004-1015. <https://doi.org/10.12928/telkomnika.v20i5.21818>

- Atalya Angelus Leza, M., Widya Utami, N., & Anugrah Cahya Dewi, P. (2024). Prediction of student performance at SMAS Catholic Saint Joseph Denpasar based on discipline and parents' economic level using the Knowledge Discovery in Database method and multiple linear regression. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 373-379. <https://doi.org/10.36040/jati.v8i1.8754>
- Fatonah, N. S., & Pancarani, T. K. (2022). Comparative analysis of clustering algorithms for mapping toddler nutritional status at Pasir Jaya Community Health Center. *Konvergensi*, 18(1), 1-9. <https://doi.org/10.30996/konv.v18i1.5497>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*. <http://arxiv.org/abs/2203.05794>
- Hartati, T., Nurdiawan, O., & Wiyandi, E. (2021). Analysis and application of the K-Means algorithm in the promotion strategy of the Suaka Bahari Maritime Academy campus. *Journal of Maritime Transportation Technology Science*, 3(1), 1-7. <https://doi.org/10.51578/j.sitektransmar.v3i1.30>
- Hasibuan, S. A., Sihombing, V., & Nasution, F. A. (2023). Analysis of community satisfaction levels using the neural network method in data mining. *Sinkron*, 8(3), 1724-1735. <https://doi.org/10.33395/sinkron.v8i3.12634>
- Informatika, T., Informatika, I., & Lampung, D. (2021). Comparison of the IRIS dataset with RapidMiner and Orange applications using classification algorithms. 158-163.
- Kharis, S. A. A., Zili, A. H. A., Putri, A., & Robiansyah, A. (2023). Analysis of trends in Indonesian public interest in artificial intelligence in welcoming Society 5.0 using Google Trends. *G-Tech: Jurnal Teknologi Terapan*, 7(4), 1345-1354. <https://doi.org/10.33379/gtech.v7i4.3091>
- Maruli Tua Silaen. (2023). Classification of student personality characteristics based on the Big Five personalities using the K-Nearest Neighbor method. *Journal of Informatics and Electronic Engineering*, 6(1), 121-129. <https://doi.org/10.36595/jire.v6i1.860>
- Poerwandono, E., & Perwitosari, J. (2023). Application of data mining for employee performance assessment at PT Riksa Dinar Djaya using the Naive Bayes classification method. *Journal of Science and Technology*, 5(1). <https://doi.org/10.55338/saintek.v5i1.1416>
- Pratama, R. K., & Mandita, F. (2024). Closed labor recruitment classification using the K-Nearest Neighbor algorithm. *Indexia*, 6(1), 42. <https://doi.org/10.30587/indexia.v6i1.7224>
- Purwati, N., Pedliyansah, Y., Kurniawan, H., Karnila, S., & Herwanto, R. (2023). Comparison of Apriori and FP-Growth data mining methods to identify sales patterns. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(2), 155-161. <https://doi.org/10.30591/jpit.v8i2.4876>
- Rafi Nahjan, M., Nono Heryana, & Apriade Voutama. (2023). Implementation of RapidMiner with the K-Means clustering method for sales analysis in Oh Cell stores. *TEKINKOM*, 7(1), 101-104. <https://doi.org/10.36040/jati.v7i1.6094>
- Setiawan, A., Rabbi, A., & Gumilang, Y. S. A. (2024). Image processing for strawberry fruit sorting based on ripeness using the K-Nearest Neighbors algorithm. *Blend Sains Jurnal Teknik*, 2(4), 322-328. <https://doi.org/10.56211/blendsains.v2i4.551>
- Singgalen, Y. A. (2021). Selection of methods and algorithms in sentiment analysis on social media: A systematic literature review. *Jurnal Sistem Informasi dan Informatika*, 3(2), 278-302. <https://doi.org/10.33557/journalisi.v3i2.125>
- Syafrizal, S., Afdal, M., & Novita, R. (2023). Sentiment analysis of PLN Mobile application reviews using the Naive Bayes classifier and K-Nearest Neighbor algorithms. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 10-19. <https://doi.org/10.57152/malcom.v4i1.983>
- Wahyudi, A., Ovelia Tampubolon, S., Afrilia Putri, N., Ghassa, A., Rasywir, E., & Kisbianty, D. (2022). Application of the Naive Bayes classifier data mining algorithm to determine customer purchase interest toward INDIHOME. *Journal of Informatics and Computer Engineering*, 2(2), 240-247. <https://doi.org/10.33998/jakakom.2022.2.2.111>