Indonesian Journal of Data and Science

Volume 6 Issue 3
ISSN 2715-9930
https://doi.org/10.56705/ijodas.v6i3.301

Research Article

Classification of Employee Attendance Categories Using the Gradient Boosted Trees Algorithm

Mutia Safitri [1,*]; Sudin Saepudin [2]; Carti Irawan [3]; Mupaat [4]
[1] Universitas Nusa Putra, Sukabumi, 43152, Indonesia, mutia.safitri_si21@nusaputra.ac.id
[2] Universitas Nusa Putra, Sukabumi, 43152, Indonesia, sudin.saepudin@nusaputra.ac.id
[3] Universitas Nusa Putra, Sukabumi, 43152, Indonesia, carti.irawan@nusaputra.ac.id
[4] Universitas Nusa Putra, Sukabumi, 43152, Indonesia, mupaat@nusaputra.ac.id

Correspondence should be addressed to Mutia Safitri; mutia.safitri_si21@nusaputra.ac.id

Received 28 September 2025; Accepted 30 December 2025; Published 31 December 2025

© Authors 2025. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).
License: https://creativecommons.org/licenses/by-nc/4.0/ — Published by Indonesian Journal of Data and Science.

Abstract:
Employee attendance is a crucial factor in human resource management as it affects productivity and operational efficiency. However, the recording and analysis of employee attendance often encounter challenges, particularly in terms of the accuracy and effectiveness of the systems used. This study aims to develop an employee attendance classification model using the Gradient Boosted Trees algorithm to improve the accuracy of grouping attendance categories such as Present, Permission, Sick, Leave, and Absent into attendance level categories: High, Medium, and Low. The research method includes collecting employee attendance data throughout the year 2024. The model evaluation is carried out using metrics such as accuracy, precision, recall, and the confusion matrix. The results indicate that the developed model achieves an accuracy of 100.00%, with a mean precision of 100.00% and a mean recall of 100.00%.

Keywords: Employee Attendance, Gradient Boosted Trees Algorithm, Machine Learning, Data Mining.

1. Introduction:

In human resource management, the analysis of employee attendance data is a vital element that can enhance organizational productivity and efficiency [1]. Attendance and employee information systems are crucial components in a company's human resource management. Timely attendance and accurate employee information are essential to support the productivity and operational efficiency of a company [2]. Attendance refers to a recording process [3], and employee productivity and efficiency are among the key factors for a company to continuously improve and grow [4]. Therefore, monitoring and analyzing employee attendance patterns have become critical aspects of modern HR management.

In general, attendance recording has utilized technologies such as fingerprint scanners and RFID cards, allowing for faster and more accurate data collection. These challenges highlight the need for methods that can address such issues in an efficient and effective manner [5]. Many previous studies have primarily focused on the recording aspect of attendance without optimally applying machine learning-based classification methods to evaluate employee attendance levels in a structured manner. This highlights a research gap in the application of data mining techniques, especially classification algorithms, to enhance attendance data management.

As time progresses, technology continues to evolve, making it easier for humans to perform various tasks [6], one of which is data mining, which is the process of discovering patterns or interesting information from selected data using specific techniques or methods. The output of data mining can be used for future decision making [7]. The techniques, methods, or algorithms in data mining are very diverse; the selection of the appropriate method or algorithm greatly depends on the objectives and the overall KDD process [8]. One of the data mining processing

techniques is classification [9]. Classification is the process of extracting patterns from existing information in order to obtain the desired rules [10]. Classification is the process of finding a model and function that describes and distinguishes data into classes [11].

Gradient boosting is a supervised learning technique based on decision trees [12]. The GBT algorithm has been widely used in various classification fields, such as water quality prediction, regional status classification, cyber-attack (DDoS) detection, and others. The objective of this study is to develop a classification model capable of categorizing employee attendance types such as Present, Permission, Sick, Leave, and Absent into grouped attendance levels: High, Medium, and Low. This approach is expected to enable companies to conduct data-driven and systematic evaluations of employee attendance. The main contribution of this study lies in the implementation of the Gradient Boosted Tree (GBT) algorithm to construct an automated, accurate, and efficient employee attendance classification system. The proposed model is expected to assist organizations in accelerating attendance analysis, minimizing data manipulation, and serving as a foundation for more effective human resource management strategies.

2. Method

Research Stages

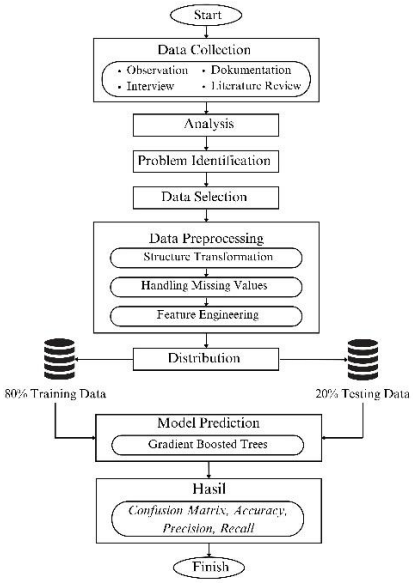


Figure 1. Research Stages

Data Source

This study uses a quantitative approach, which is an approach that describes or explains a problem that can be generalized because it is based on actual reality, as this research is related to and involves data [13].

The first stage carried out is data collection [14]. The data used in this study were obtained from two main sources: primary data and secondary data. The primary data in this study are data collected directly from the company's attendance system or internal application that records employee attendance. This study uses data obtained from [15] the recap stored by the HR department related to employee attendance. Attendance is an activity of data collection to determine the number of employee presences in a company [16].

In addition to primary data, this study also utilizes secondary data to complement the analysis. The sources of secondary data include literature and previous research.

Thus, the data source collected is employee attendance data totaling 100,354 records from January to December 2024. The types of data collected include: Employee ID (NIK), Name, Department, Area, Shift Group, Supervisor,

Position, Date, Category, Week, Month, and Year. These data were then evaluated to determine the accuracy obtained from the development and implementation of the model on new data [17].

Implementation of the Gradient Boosted Trees Algorithm

Gradient Boosted Trees is a classification algorithm that uses a boosting technique, which consists of a collection of several base models that are combined into a final, accurate model. With this technique, it can be proven that the Gradient Boosted Trees algorithm can produce a more accurate model compared to other classification algorithms [18]. The implementation stages in this study are as follows:

a. Data Selection: Data selection is the process of determining datasets or subsets of data that are relevant and significant for specific analytical or research purposes [19].

b. Data Preprocessing: Data preprocessing aims to transform raw data into high-quality data that is suitable for further processing [20]. Below is a summary table of preprocessing steps and their respective purposes**:**

Step	Description
Structure Transformation	Structure Transformation in this study aims to convert raw data into a more structured format. In this case, the attendance category column is transformed into several separate columns.
Handling Missing	Handling Missing Values is performed to fill in missing or empty values that could affect prediction results. Handling missing values before processing with various machine learning models has been shown to improve classification performance [21].
Feature Engineering	Feature Engineering involves creating new features from the existing raw data. In this study, feature engineering is carried out to build more representative features from the available employee attendance data.

The following formulas were used in the Feature Engineering stage:

Feature Name	Formula
Total Employee	Present + Permission + Sick + Leave + Absent
Effective Present	Present
%Present	(Effective Present / Total) × 100%
Attendance Level	If (%Present ≥ 90, “High”, If (%Present ≥ 75, “Medium”, “Low”))

c. Data Splitting: The data must be divided into training and testing sets to evaluate model performance. In this study, 80% of the data is used for training, and 20% is used for testing.

d. Implementation of the Gradient Boosted Trees Algorithm: After completing data preprocessing and splitting, the Gradient Boosted Trees algorithm is implemented using RapidMiner software. The process involves assigning the target label, selecting predictor features, training the model using the Gradient Boosted Trees operator, and evaluating performance using the Apply Model and Performance operators.

e. Analysis Results: Evaluation of the Gradient Boosted Trees algorithm is conducted to determine how well the model performs in classification. The confusion matrix is a table used to show the number of test data instances correctly and incorrectly classified, making it easier to evaluate the accuracy of a classification system [22]. The general form of the confusion matrix can be seen in the following figure:

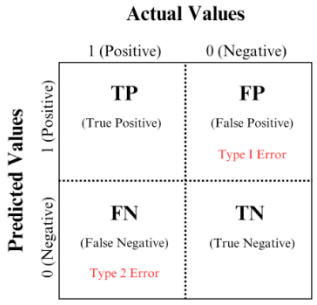


Figure 2. Confusion Matrix

There are three values used to measure the performance of the classification system, namely precision, recall, and accuracy. The value of precision refers to the sensitivity or correctness of the system in accurately identifying whether the data belongs to the negative or positive class.

Meanwhile, recall is the value that indicates the level of success or specificity in correctly retrieving information about data that belongs to either the negative or positive class [23].

The formula for accuracy is as follows:

Accuracy	(Number of correct predictions) / (Total Number of predictions) = (TP + TN) / (TP + TN + FP + FN)
Recall	TP / (TP + FN)
Precission	TP / (TP + FP)

Explanation:

- True Positive (TP): the condition where the model predicts positive and the result is indeed positive.
- True Negative (TN): the condition where the model predicts negative and the result is indeed negative.
- False Positive (FP): the condition where the model predicts positive but the result is actually negative.
- False Negative (FN): the condition where the model predicts negative but the result is actually positive.

3. Result and Discussion:

Result

In this study, the author uses RapidMiner software because the platform is equipped with various operators for data preprocessing, modeling, evaluation, and result visualization [24]. RapidMiner provides a user interface for designing analysis pipelines, which generates an XML file that describes the analysis process the user intends to apply to the data [25].

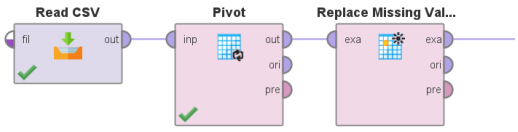


Figure 3. Process (1) Gradient Boosted Trees Algorithm

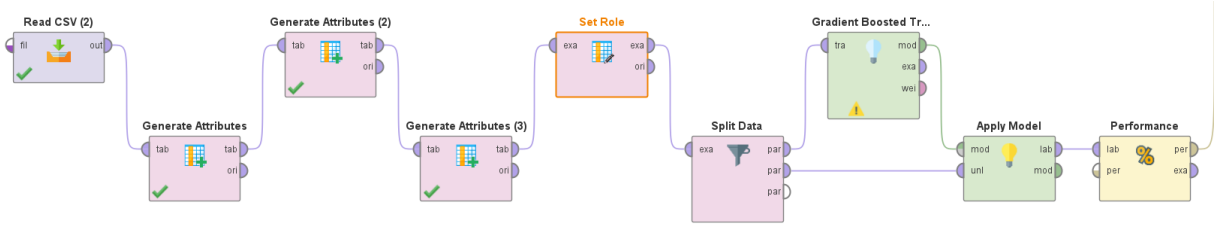


Figure 4. Process (2) Gradient Boosted Trees Algorithm

a. Data Selection

The initial stage in the classification process begins with reading the employee attendance dataset, which is stored in Comma Separated Values (CSV) format, using the Read CSV operator. Subsequently, data selection is performed to ensure that only relevant attributes are used in the model training process. Several attributes such as Employee ID, employee name, work area, shift group, supervisor name, position, and year are removed using the Exclude Column operator, as they are considered to have no direct influence on the classification of attendance levels. The removal of non-relevant attributes aims to reduce model complexity and prevent overfitting, while ensuring that only features with informative value toward the classification target are utilized. This process also helps accelerate the modeling and evaluation stages.

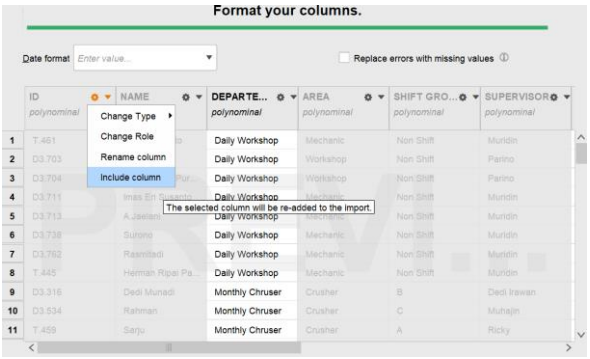


Figure 5. Process (1) Data Selection

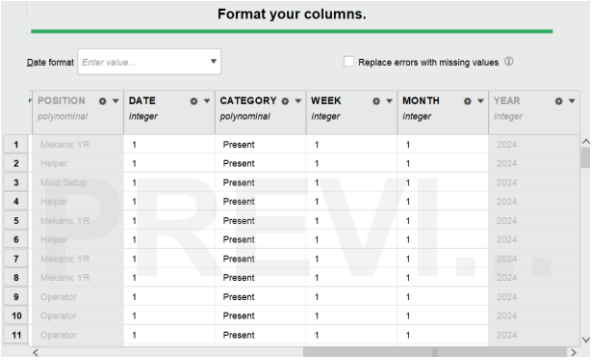


Figure 6. Process (2) Data Selection

b. Data Preprocessing

- Structure Transformation : The initial dataset was in a long format, where each data row represented a single daily attendance record for each employee. Attendance categories such as *Present*, *Permission*, *Sick*, *Leave*, and *Absent* were recorded in the same column. To enable more effective analysis, a structure transformation was performed to convert the data into a wide format by separating each attendance category into its own column. The result of the structure transformation process showed a significant change in the number of entries. The original dataset consisted of 100,354 rows, and the final result after transformation became 7,320 rows. This stage used the Pivot operator.

Row No.	DEPARTEM...	DATE	WEEK	MONTH	count(CATE...	count(CATE...	count(CATE...	count(CATE...	count(CATE...
1	Daily Workshop	1	1	1	?	?	?	8	?
2	Daily Workshop	1	5	2	?	?	?	8	?
3	Daily Workshop	1	9	3	?	?	?	7	?
4	Daily Workshop	1	14	4	?	?	?	6	?
5	Daily Workshop	1	18	5	?	?	?	6	?
6	Daily Workshop	1	22	6	?	?	?	6	?
7	Daily Workshop	1	27	7	?	?	?	8	?
8	Daily Workshop	1	31	8	?	?	?	8	?
9	Daily Workshop	1	36	9	?	?	?	8	?
10	Daily Workshop	1	40	10	?	?	?	8	?
11	Daily Workshop	1	44	11	?	?	1	7	?
12	Daily Workshop	1	49	12	?	?	?	8	?
13	Daily Workshop	2	1	1	?	?	1	7	?

ExampleSet (7,320 examples,0 special attributes,9 regular attributes)

Figure 7. Process Stucture Transformation

- Handling Missing Values: Missing values refer to conditions where a column lacks data due to unrecorded entries, input errors, or formatting mismatches during data export. The presence of empty values can negatively impact analysis results and the performance of classification models, especially during the training and prediction processes. In the transformed dataset, missing values were found in the attendance category columns such as *Present*, *Permission*, *Sick*, *Leave*, and *Absent*. This occurred because not all attendance categories appear in every combination of time and department attributes. For example, on a certain day, there may be no employees categorized as *Sick*, resulting in a null value in the *Sick* column for that row. In this study, missing values were replaced with the number zero (0). Zero was chosen because it logically represents the absence of an event in a specific attendance category for instance, no employees were on leave or sick on a particular day and in a particular department. This stage used the Replace Missing Values operator.

Row No.	DEPARTEM...	DATE	WEEK	MONTH	count(CATE...	count(CATE...	count(CATE...	count(CATE...	count(CATE...
1	Daily Workshop	1	1	1	0	0	0	8	0
2	Daily Workshop	1	5	2	0	0	0	8	0
3	Daily Workshop	1	9	3	0	0	0	7	0
4	Daily Workshop	1	14	4	0	0	0	6	0
5	Daily Workshop	1	18	5	0	0	0	6	0
6	Daily Workshop	1	22	6	0	0	0	6	0
7	Daily Workshop	1	27	7	0	0	0	8	0
8	Daily Workshop	1	31	8	0	0	0	8	0
9	Daily Workshop	1	36	9	0	0	0	8	0
10	Daily Workshop	1	40	10	0	0	0	8	0
11	Daily Workshop	1	44	11	0	0	1	7	0
12	Daily Workshop	1	49	12	0	0	0	8	0
13	Daily Workshop	2	1	1	0	0	1	7	0

ExampleSet (7,320 examples,0 special attributes,9 regular attributes)

Figure 8. Process Handling Missing Values

- Feature Engineering : The next stage in the preprocessing process is feature engineering, which involves creating new features designed to enhance the model's ability to learn patterns relevant to the classification task. Feature engineering is a crucial step, as well-crafted features can improve model accuracy and accelerate the training process. In this stage, the Generate Attributes operator was used by applying formulas (1), (2), (3), and (4).

	Total Employee	Effective Present	%Pres...	Attandance Level
	8	8	100	High
	8	8	100	High
	7	7	100	High
	6	6	100	High
	6	6	100	High
	6	6	100	High
	8	8	100	High
	8	8	100	High
	8	8	100	High
	8	8	100	High
	8	7	87	Medium
	8	8	100	High

Figure 9. Process Feature Engineering

c. Data Splitting

In this study, the Attendance Level attribute was set as the label (target/output) because it contains the information to be predicted by the classification model, namely the grouping into attendance categories: High, Medium, and Low. Attribute roles were assigned by classifying other attributes such as Total Employees, Effective Attendance, Present, Permission, Sick, Leave, Absent, and Attendance Percentage as predictors (inputs). After the labeling process, the data was divided into two main sets:

- Training data: 80% of the total data was used to build and train the classification model. In this study, 5,856 rows were used as training data.
- Testing data: 20% of the total data was used to evaluate the model's performance. In this study, 1,464 rows were used as testing data.

d. Implementation of the Gradient Boosted Trees Algorithm

After completing the data selection, preprocessing, and data splitting stages, the next step was to implement the Gradient Boosted Trees (GBT) algorithm to build the employee attendance classification model. This stage is the core of the machine learning process, where the system is trained to recognize patterns in the data and generate a predictive model capable of accurately classifying new data. This stage used the Gradient Boosted Trees, Apply Model, and Performance operators.

Discussion

After the model training process was completed, the next step was to evaluate the model's performance on the testing data to determine how well the model could classify the attendance level categories (*High*, *Medium*, *Low*) on previously unseen data.

accuracy: 100.00%

	true High	true Medium	true Low	class precision
pred. High	1315	0	0	100.00%
pred. Medium	0	115	0	100.00%
pred. Low	0	0	34	100.00%
class recall	100.00%	100.00%	100.00%	

Figure 10. Results of Gradient Boosted Trees

accuracy: 99.86%

	true High	true Medium	true Low	class precision
pred. High	1315	0	0	100.00%
pred. Medium	0	114	1	99.13%
pred. Low	0	1	33	97.06%
class recall	100.00%	99.13%	97.06%	

Figure 11. Results of Deep Learning

accuracy: 94.67%

	true High	true Medium	true Low	class precision
pred. High	1242	0	0	100.00%
pred. Medium	73	114	4	59.69%
pred. Low	0	1	30	96.77%
class recall	94.45%	99.13%	88.24%	

Figure 12. Results of Naïve Bayes

The classification of employee attendance categories using the Gradient Boosted Trees (GBT) algorithm produced excellent results, achieving 100% accuracy. All instances across the three classes High, Medium, and Low were correctly classified with no misclassifications. This performance is further supported by precision and recall values of 100% for each class, indicating that the model successfully identified all relevant patterns in the dataset. The strength of GBT lies in its ability to build strong learners through a series of decision trees, allowing it to handle complex variations in the attendance data with high precision.

Although this result is excellent, the perfect accuracy (100%) should be interpreted with caution, as it may indicate a case of overfitting a situation where the model performs exceptionally well on training data but may not generalize effectively to unseen data. This concern is reinforced by the absence of validation techniques such as k-fold cross-validation, which are typically used to evaluate a model's generalizability.

Moreover, the dataset may suffer from class imbalance, such as a disproportionately high number of records in the "High" category compared to "Medium" or "Low". This imbalance could introduce bias into the model. The study does not mention any resampling techniques or class weighting strategies to address this issue. A biased model may perform poorly when encountering underrepresented categories in real-world scenarios.

On the other hand, comparing the GBT model with other algorithms such as Deep Learning (99.86% accuracy) and Naïve Bayes (94.67% accuracy) adds substantial value to the study. Interestingly, although Deep Learning is known for its strength in modeling complex and non-linear patterns, GBT outperformed it in this case. This can be explained by several factors:

- The dataset is structured and tabular, which is ideal for tree-based models like GBT.GBT leverages a boosting mechanism, effectively combining multiple weak learners to build a more accurate final model.
- It automatically performs feature selection, making it particularly suitable for datasets with well-engineered attributes.
- Deep Learning typically requires larger datasets and more computational resources to outperform tree-based models, which may not have been optimal in this context.
- Meanwhile, Naïve Bayes, although efficient and fast, operates under the assumption of feature independence, which may not hold true for employee attendance data, thus resulting in lower accuracy.

4. Conclusion:

This study aimed to classify employee attendance levels (High, Medium, Low) based on attendance data using the Gradient Boosted Trees (GBT) algorithm. The evaluation results show that the GBT model achieved very high performance, with accuracy, precision, and recall reaching 100%. These results indicate that the algorithm was able to effectively and accurately recognize patterns within the attendance dataset.

However, such perfect performance should be interpreted with caution, as the model has not yet been tested on data from different periods or organizational contexts. Therefore, the model's ability to perform consistently under various conditions has not been fully established.

The limitations of this study include the use of data from a single source and time period, and the lack of real-world implementation within an operational attendance system. Additionally, issues related to class distribution and feature importance analysis were not explored in detail. For future research, it is recommended to:

- Apply the model to datasets from different departments, organizations, or timeframes to assess its adaptability.
- Develop and integrate the model into a digital attendance system, such as a web- or app-based platform.
- Conduct deeper analysis of feature importance to identify which attributes most influence classification outcomes, thereby improving both model efficiency and interpretability.

Acknowledgments:

The author would like to express sincere gratitude to PT XYZ for providing the attendance dataset used in this study. Appreciation is also extended to the Human Resources Department for granting access to the company's internal attendance system and for their assistance in data collection.

Special thanks are given to the author's academic supervisors and lecturers at Nusa Putra University, whose guidance and feedback have been invaluable throughout the research process. Lastly, heartfelt appreciation goes to family and colleagues for their continuous support and encouragement during the completion of this research.

References:

[1] F. P. Azizah et al., "Comparison of K-Means and Hierarchical Algorithms for Employee Attendance Data Clustering," JUTISI (Scientific Journal of Informatics Engineering and Information Systems), vol. 14, no. 1, doi: 10.35889/jutisi.v14i1.2644, 2025.

[2] D. Darmawan, R. Hidayat, and A. Kurniawan, "Development of Web-Based Employee Attendance and Information System," vol. 1, no. 6, pp. 928–933, 2024.

[3] S. Darma, Y. Yusman, and J. Hendrawan, "Employee Attendance Level Data Analysis Using K-Means Clustering at the Department of Public Works and Spatial Planning of Langkat Regency," Minfo Polgan, vol. 13, no. 1, August 2024, doi: https://doi.org/10.33395/jmp.v13i1.13958e- ISSN : 2797-3298p.

[4] H. A. Setyadi and S. Sundari, "Employee Attendance and Working Hours Management Information System for Payroll Calculation Completeness," Indones. J. Comput. Sci., vol. 1, no. 1, pp. 28–33, 2022, doi: 10.31294/ijcs.v1i1.1114.

[5] Jihan Asa Noyari et al., "Optimization of Campus Management Information System Performance Using Data Mining Techniques," MENTARI (Management, Education and Information Technology), vol. 3, no. 1, pp. 52−63, September 2024.

[6] W. A. Wahyuni and S. Saepudin, "Application of Clustering Data Mining to Group Various Brands of Washing Machines," SISMATIK (National Seminar on Information Systems and Informatics Management, Nusa Putra University), vol. 1, no. 1, pp. 306–313, 2021.

[7] F. Sodik Pamungkasa et al., "Comparison of Supervised Learning Classification Methods on Bank Customer Data Using Python," PRISMA, vol. 3, pp. 689–694, 2020.

[8] S. Suliman, "Implementation of Data Mining on Student Academic Performance Based on Social Life and Socio-Economy Using K-Means Clustering Algorithm," Simkom, vol. 6, no. 1, pp. 1–11, 2021, doi: 10.51717/simkom.v6i1.48.

[9] Panji Bimo Nugroho Setio et al., "Classification Using Decision Tree Based on C4.5 Algorithm," PRISMA 2020, vol. 3, pp. 64–71.

[10] Ainurrohmah, "Accuracy of Classification Algorithms in RapidMiner and Weka Software," PRISMA 2021, vol. 4, pp. 493–499, 2021.

[11] N. Hidayati et al., "Application of Clustering and Classification Algorithms on the Importance Level of Learning Systems at Open University," SWABUMI JOURNAL, vol. 8, no. 2, September 2020, pp. 134~142.

[12] S. Elsa Suryana et al., "Application of Gradient Boosting with Hyperopt to Predict Bank Telemarketing Success," GAUSSIAN, vol. 10, no. 4, 2021.

[13] Atika Juhaedah Alifah, Sudin Saepudin, and Carti Irawan, "Implementation of K-Means Clustering Algorithm in Analyzing Public Satisfaction Toward Public Services," Journal of Informatics Engineering (JUTIF), vol. 5, no. 4, August 2024, pp. 487–496, DOI: https://doi.org/10.52436/1.jutif.2024.5.4.2125.

[14] T. Zulhaq Jasman et al., "Analysis of Gradient Boosting, AdaBoost, and CatBoost Algorithms in Water Quality Classification," Journal of Informatics Engineering and Information Systems, vol. 8, no. 2, August 2022.

[15] A. N. Z. Hidayah and A. F. Rozi, "Application of Data Mining in Determining the Best Employee Performance Using the C4.5 Algorithm," JISAI, vol. 1, no. 2, May 2021.

[16] S. Wahyuni and M. Sulaeman, "Application of Deep Learning Algorithm for Face Detection Attendance System at PT Karya Komponen Presisi," J. Inform. SIMANTIK, vol. 7, no. 1, pp. 5–6, 2022.

[17] A. Pebdika, R. Herdiana, and D. Solihudin, "Classification Using Naive Bayes Method to Determine Candidates for PIP Recipients," JATI (Journal of Informatics Engineering Students), vol. 7, no. 1, pp.

[18] E. Firasari et al., "Combination of K-NN and Gradient Boosted Trees for Social Assistance Program Acceptance Classification," J. Inf. Technol. and Computer Science, vol. 7, no. 6, pp. 1231–1236, 2020, doi: 10.25126/jtiik.202073087.

[19] Arya Wijaya, Ahmad Faqih, Dodi Solihudin, Cep Lukman Rohmat, and Sandy Eka Permana, "Application of Association Rules Using Apriori Algorithm to Identify Purchase Patterns," JATI (Journal of Informatics Engineering Students), vol. 7, no. 6, 2023.

[20] Fauzan and Didi Juardi, "Application of Data Mining on Food and Beverage Sales Using Naive Bayes Algorithm," Scientific Journal of Informatics (JIF), vol. 9, no. 2, 2021.

[21] A. F. Nugraha, Y. Pristyanto, and I. Pratama, "Handling Missing Values to Improve Machine Learning Model Performance on Telemarketing Data," Pseudocode Journal, vol. 7, no. 2, September 2020.

[22] R. Nurhidayat and K. E. Dewi, "Application of K-Nearest Neighbor Algorithm and N-Gram Feature Extraction in Aspect-Based Sentiment Analysis," Komputa: Scientific Journal of Computing and Information, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.

[23] M. Azhari, Z. Situmorang, and R. Rosnelly, "Comparison of Accuracy, Recall, and Precision of Classification Algorithms: C4.5, Random Forest, SVM, and Naive Bayes," Budidarma Informatics Media Journal, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.

[24] D. A. Maqfiroh and Z. Fatah, "K-Nearest Neighbor for Weather Condition Prediction Using RapidMiner," Journal of Information Systems and Technology (SANDI), vol. 6, no. 2, November 2024.

[25] Fauziah and A. S. Ramadhantya, "Using RapidMiner to Predict Student Graduation with Naive Bayes Algorithm," vol. 10, no. 1, June 2024.