Jurnal Elektronik Ilmu Komputer Udayana
Volume 13, No 2. November 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Comparative Analysis of Algorithm of Naive Bayes and C4.5 on Employee Performance Classification

Muhammad Sadikin[a1], Rini Oktari Batubara [a2]

[a,b] Information Systems Department, University of Potensi Utama
Medan, Indonesia
[1]dicky.aries.3@gmail.com(Corresponding author)
[2] rini.admmedan@gmail.com

Abstract

Employee performance appraisal is one of the supporting factors to increase the productivity of an agency's performance. Performance appraisal is needed to determine the level of success of each employee's performance and is used to assess success in accordance with the objectives set in order to realize the mission and vision of a company. The employee performance appraisal system that runs at the Potensi Utama of University is outlined in the Standard Operating Procedure (SOP) document. There are 10 criteria that serve as benchmarks for assessment, namely work discipline, administrative discipline, reliability, maturity and personal integrity, morale, communication and cooperation, adherence to work principles, quantity of work results, quality of work results, and concern for the organization. The purpose of this study was to determine the level of accuracy of the two methods used by using Weka 3.8 tools using performance evaluation data mining Cross Validation 2,5,6 and 10 Fold Validation as a comparison

Keywords: *Data Mining, Classification, Naïve Bayes, C4.5, Employee Performance Appraisal*

1. Introduction

Employee performance appraisal is one of the supporting factors to increase the productivity of an agency's performance, therefore with performance appraisal it will be known the achievements of each employee, this is because the level of synergy between employees and the company is very high[1]. The employee performance appraisal system that runs at the Potensi Utama of University is outlined in the Standard Operating Procedure (SOP). In the employee performance report, there are several criteria that serve as benchmarks for assessment, namely work discipline, administrative discipline, reliability, personal maturity and integrity, morale, communication and cooperation, adherence to work principles, quantity of work results, quality of work results, and concern. against the organization. While the resulting output, there are 3 classes, namely Special, Good and Enough.

The procedure that runs in assessing employee performance is carried out based on the weighted average calculation process for each criterion, but the classification for determining the highest level to the lowest level on the performance appraisal criteria has not been applied so that an accuracy value has not been obtained in the final result of the accumulated calculation in giving. weight.

Based on this description, the authors conducted a comparative analysis of employee performance appraisal classifications with the aim of obtaining information on employee performance classification levels as a recommendation in determining the highest level to the lowest level through a decision tree that can affect employee performance appraisal. Then the algorithm calculation process is used to obtain the calculation results of the accuracy value in the classification of employee performance appraisals through two algorithms, namely the Naive Bayes Algorithm and C4.5.

Research using these two algorithms, the authors choose based on several studies that have been reviewed, with the results of the first review, namely the superiority of the Naïve Bayes algorithm in research conducted by Amra and Maghari; 2017 states that the application of the Naïve Bayes algorithm is able to provide computational time effectiveness in processing training data and the accuracy of the results given to training data is also good[2]. Then another research conducted by Jadhav and Channe; 2016 said that the Naïve Bayes algorithm also has good performance, requiring short computation time to process training data[3]. also said that the C4.5 algorithm has the advantage of providing high accurate results, with the representation decision tree easy to understand and

understandable[4]. Then according to researchers Kostopoulos.et all; 2019 where the C4.5 algorithm is one of the decision tree model algorithms that has been successfully applied to solving various problems and has proven to be quite effective in classification and regression which shows a number of major advantages, such as simplicity, flexibility and interpretability[5].

2. Reseach Methods

The research framework is used in solving problems in research. The stages in the research framework are data acquisition (collecting data) Employee Performance Reports, dividing data into 2 parts, namely training data and testing data, conducting training data experiments on each algorithm with the aim of finding a Classification Model of the algorithms used, then doing an experiment with testing data and comparing the results of the testing data experiment against the two algorithms used. testing using the Naïve Bayes algorithm, and the C4.5 algorithm, with the help of Weka tools to find comparisons of accuracy results between the two algorithms. The research framework can be seen in Figure 1.

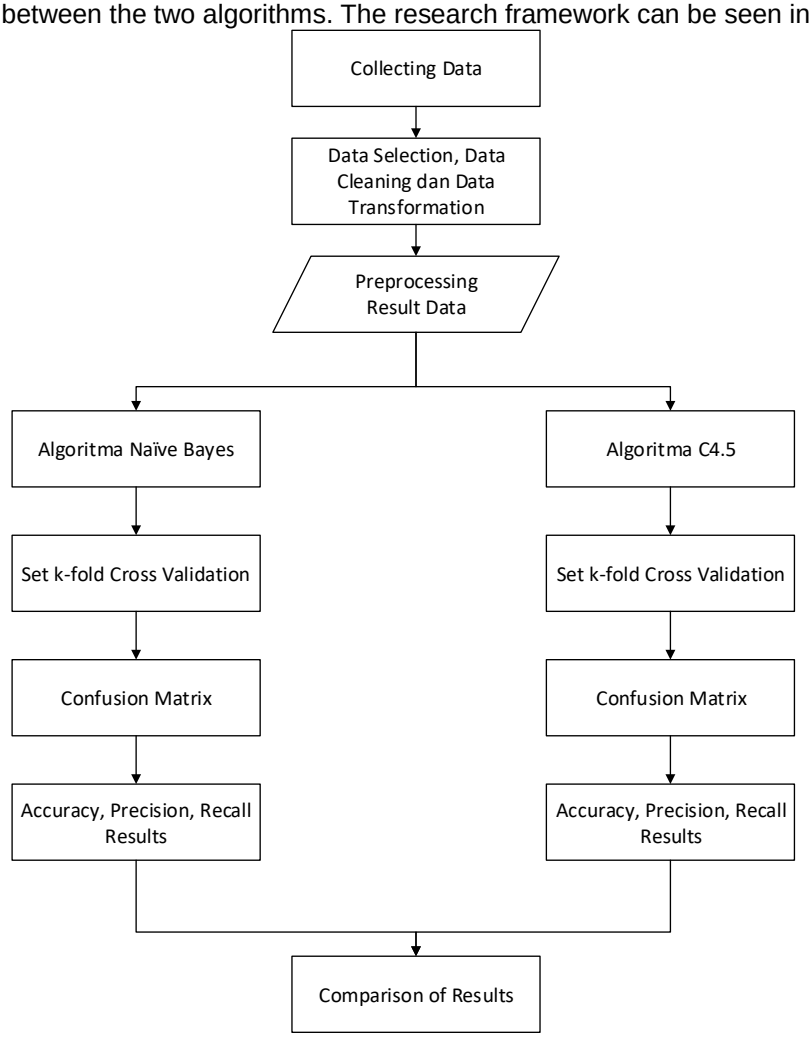


Figure 1. Research Framework

DATA COLLECTION

This stage is the initial stage that is carried out before the data testing process is carried out. The data collection process carried out in this study was carried out by means of interviews and literature studies. After the data obtained is then collected and analyzed. Then the data that has been analyzed are grouped so that later the writer will be easy to analyze the next data. In this study the data used are data such as:

1. Data on the number of employees at the Potensi Utama of University.
2. Data on the performance appraisal report of the Potensi Utama of University employees xample data 50.
3. There are 10 variables used in determining employee performance in this study, namely work discipline, administrative discipline, reliability, maturity and personal integrity, morale,

p-ISSN: 2301-5373
e-ISSN: 2654-5101

communication and cooperation, adherence to work principles, quantity of work results, quality of work results, and concern for the organization.

4. The weight value for each of the assessment variables used in determining employee performance.

PREPROCESSING DATA MINING

Preprocessing is one of the important stages for data in the mining process. The data used in the mining process is not always in ideal conditions for processing. Sometimes in this data there are various problems that can interfere with the results of the mining process itself, such as missing value, redundant data, or data formats that are not in accordance with the system. Therefore, to overcome these problems, a preprocessing stage is needed. Preprocessing is one of the stages of eliminating problems that can interfere with the results of data processing. In this study, the Prepocessing stage consisted of Data Selection, Data Cleaning and Data Transformation.

The research data will then be processed using the concept of data mining using the Naïve Bayes and C4.5 algorithms, starting from the data selection process, data cleaning, data transformation, calculations using the Naïve Bayes and C4.5 algorithms to produce new knowledge. The preprocessing stages in data mining are as follows:

1. Data Selection

Data Selection is the selection of data used for the data mining process. The selected data that will be used for the data mining process is stored in a file and separated from the operational database[6].

Employee performance report data at University of Potensi Utama is the result of a summary given by each division to the HRD (Vice Chancellor II). Sample employee performance report data at University of Potensi Utama can be seen in Table I.

Table 1. Data Selection

No.	Inisial	Jabatan	Dimensi 1 (B=15)	Dimensi 2 (B=10)	Dimensi 3 (B=10)	Dimensi 4 (B=5)	Dimensi 5 (B=10)	Dimensi 6 (B=5)	Dimensi 7 (B=5)	Dimensi 8 (B=15)	Dimensi 9 (B=15)	Dimensi 10 (B=10)	Total BxN	Nilai Akhir	Class
1	SEWR	Staf Adminitr	8,5	9	8,5	9	7	6	8	8,5	8,5	7	812,5	81,25	Istimewa
2	IR	Staf Rektorat	9	8,5	9	8,5	7	6	7	9	8	9	832,5	83,25	Istimewa
3	SWP	Staf Rektorat	9	9	8,5	8,5	7	6	8	8,5	8	8	820	82	Istimewa
4	CP	Staf Rektorat	9	9	9	9	6	8,5	8,5	8	8,5	8,5	837,5	83,75	Istimewa
5	MS	Staf Rektorat	9	9	9	9	6	8,5	8,5	8	8	8	825	82,5	Istimewa
6	HH	Staf BAAK	8,5	8	8,5	8	8	6	8	8,5	8,5	8	817,5	81,75	Istimewa
7	VAG	Staf BAAK	8,5	8	8,5	8	8	6	8	8,5	8	8,5	815	81,5	Istimewa
8	SL	Staf BAAK	8,5	7	8,5	8	7	6	8	9	8	8	797,5	79,75	Baik
9	DSB	Staf BAAK	8,5	7	8,5	7	7	6	7	8	8,5	8,5	785	78,5	Baik
10	EA	Staf BAAK	8,5	8	8,5	8	6	8,5	8,5	8	8	8	797,5	79,75	Baik
11	EZR	Staf BAAK	8,5	8	8,5	8	6	8,5	8	8,5	8,5	8,5	815	81,5	Istimewa
12	SA	Staf BAAK	8,5	8	8,5	8	6	8,5	8	8,5	8	8	802,5	80,25	Istimewa
13	RIPA	Staf BAAK	8,5	8	8,5	8	6	8,5	8	8	8	8	795	79,5	Baik
14	DS	Staf BAAK	8	8	8,5	8	8,5	8	8	8	8	8	810	81	Istimewa
15	FSW	Staf BAAK	8	8	9	8	8	9	9	8	8	8,5	825	82,5	Istimewa
16	IA	Staf BAAK	8	8	8	9	9	9	9	9	8	8	840	84	Istimewa
17	NS	Staf BAAK	7	7	7	8,5	8,5	8,5	8,5	7	8,5	8,5	775	77,5	Baik
18	MZ	Staf BAAK	8	8	8	8,5	8,5	8,5	9	7	8,5	7	797,5	79,75	Baik
19	ZS	Staf BAAK	8	8	8	8	6	8,5	8	8,5	8	8	790	79	Baik
20	RA	Administrasi	7	7	7	7	6	8,5	7	7	7	7	697,5	69,75	Cukup
21	JA	Administrasi	7	7	7	7	6	7	7	7	8,5	7	712,5	71,25	Baik
22	RU	Administrasi	7	7	7	7	6	7	7	7	7	8,5	705	70,5	Baik
50	DM	Administrasi	7	7	7	7	6	7	7	7	7	7	690	69	Cukup

2. Data Selection

Data Cleaning is the process of eliminating unnecessary data, redundant data. The data that is eliminated in the data cleaning process is eliminating the Employee Name and Employee Position variables. The Cleaning Results of the employee performance report data at the University of Potensi Utama can be seen in Table 2[7]

Table 2. Data Cleaning

No	Dimensi 1 (B=15)	Dimensi 2 (B=10)	Dimensi 3 (B=10)	Dimensi 4 (B=5)	Dimensi 5 (B=10)	Dimensi 6 (B=5)	Dimensi 7 (B=5)	Dimensi 8 (B=15)	Dimensi 9 (B=15)	Dimensi 10 (B=10)	Total BxN	Nilai Akhir	Class
1	8,5	9	8,5	9	7	6	8	8,5	8,5	7	812,5	81,25	Istimewa
2	9	8,5	9	8,5	7	6	7	9	8	9	832,5	83,25	Istimewa
3	9	9	8,5	8,5	7	6	8	8,5	8	8	820	82	Istimewa
4	9	9	9	9	6	8,5	8,5	8	8,5	8,5	837,5	83,75	Istimewa
5	9	9	9	9	6	8,5	8,5	8	8	8	825	82,5	Istimewa
6	8,5	8	8,5	8	8	6	8	8,5	8,5	8	817,5	81,75	Istimewa
7	8,5	8	8,5	8	8	6	8	8,5	8	8,5	815	81,5	Istimewa
8	8,5	7	8,5	8	7	6	8	9	8	8	797,5	79,75	Baik
9	8,5	7	8,5	7	7	6	7	8	8,5	8,5	785	78,5	Baik
10	8,5	8	8,5	8	6	8,5	8,5	8	8	8	797,5	79,75	Baik
11	8,5	8	8,5	8	6	8,5	8	8,5	8,5	8,5	815	81,5	Istimewa
12	8,5	8	8,5	8	6	8,5	8	8,5	8	8	802,5	80,25	Istimewa
13	8,5	8	8,5	8	6	8,5	8	8	8	8	795	79,5	Baik
14	8	8	8,5	8	8,5	8	8	8	8	8	810	81	Istimewa
15	8	8	9	8	8	9	9	8	8	8,5	825	82,5	Istimewa
16	8	8	8	9	9	9	9	9	8	8	840	84	Istimewa
17	7	7	7	8,5	8,5	8,5	8,5	7	8,5	8,5	775	77,5	Baik
18	8	8	8	8,5	8,5	8,5	9	7	8,5	7	797,5	79,75	Baik
19	8	8	8	8	6	8,5	8	8,5	8	8	790	79	Baik
20	7	7	7	7	6	8,5	7	7	7	7	697,5	69,75	Cukup
21	7	7	7	7	6	7	7	7	8,5	7	712,5	71,25	Baik
22	7	7	7	7	6	7	7	7	7	8,5	705	70,5	Baik
50	7	7	7	7	6	7	7	7	7	7	690	69	Cukup

3. Data Transformation and Data Reduction

Data transformation is the process of classifying data, the data to be classified are work discipline, administrative discipline, reliability, maturity and personal integrity, work spirit, communication and cooperation, steadfastness to work principles, quantity of work results, quality of work results, to concern for the organization. The basis for creating transformation data is based on the weight assessment stated in the Standard Operating Procedure (SOP) at University of Potensi Utama.

Data Transformation is a data adjustment step by converting all attribute values into characters which aim to make the clustering process easier[8].

Table 3. Attribute Value Conversion

1. Disiplin Kerja		**2. Disiplin Administrasi**	
Kelompok Bobot	**Klasifikasi**	**Kelompok Bobot**	**Klasifikasi**
x <= 45	Tidak_Disiplin	x <= 30	Tidak_Disiplin
45,1 <= x <= 90	Kurang_Disiplin	30,1 <= x <= 60	Kurang_Disiplin
90,1 <= x <= 120	Cukup_Disiplin	60,1 <= x <= 80	Cukup_Disiplin
x > 120	Disiplin	x > 80	Disiplin
3. Keandalan		**4. Kedewasaan dan Integritas Pribadi**	
Kelompok Bobot	**Klasifikasi**	**Kelompok Bobot**	**Klasifikasi**
x <= 30	Tidak_Bisa_Dipercaya	x <= 15	Tidak_Konsisten
30,1 <= x <= 60	Kurang_Bisa_Dipercaya	15,1 <= x <= 30	Kurang_konsisten
60,1 <= x <= 80	Cukup_Bisa_Dipercaya	30,1 <= x <= 40	Cukup_konsisten
x > 80	Bisa_Dipercaya	x > 40	Konsisten
5. Semangat Kerja		**6. Komunikasi dan kerja sama**	
Kelompok Bobot	**Klasifikasi**	**Kelompok Bobot**	**Klasifikasi**
x <= 30	Tidak_Semangat	x <= 15	Tidak_Inisiatif
30,1 <= x <= 60	Kurang_Semangat	15,1 <= x <= 30	Kurang_Inisiatif
60,1 <= x <= 80	Cukup_Semangat	30,1 <= x <= 40	Cukup_inisiatif
x > 80	Semangat	x > 40	Inisiatif
7. Keteguhan pada Prinsip Kerja		**8. Kuantitas Hasil Kerja**	
Kelompok Bobot	**Klasifikasi**	**Kelompok Bobot**	**Klasifikasi**
x <= 15	Tidak_Menguasai	x <= 45	Tidak_Tercapai
15,1 <= x <= 30	Kurang_Menguasai	45,1 <= x <= 90	Jarang_Tercapai
30,1 <= x <= 40	Cukup_Menguasai	90,1 <= x <= 120	Cukup_Tercapai
x > 40	Sangat_Menguasai	x > 120	Selalu_Tercapai
9. Kualitas Hasil Kerja		**10. Kepedulian Terhadap Organisasi**	
Kelompok Bobot	**Klasifikasi**	**Kelompok Bobot**	**Klasifikasi**
x <= 45	Tidak_Memuaskan	x <= 30	Tidak_Perduli
45,1 <= x <= 90	Jarang_Memuaskan	30,1 <= x <= 60	Kurang_Perduli
90,1 <= x <= 120	Cukup_Memuaskan	60,1 <= x <= 80	Cukup_Perduli
x > 120	Selalu_Memuaskan	x > 80	Sangat_Perduli

p-ISSN: 2301-5373
e-ISSN: 2654-5101

The last stage of data preprocessing is data reduction. This is the process of reducing data to reduce the volume of data without losing important information. This can speed up processing and analysis.

Table 5. Transfomation Results

Dimensi 1 (B=15)	Dimensi 2 (B=10)	Dimensi 3 (B=10)	Dimensi 4 (B=5)	Dimensi 5 (B=10)	Dimensi 6 (B=5)	Dimensi 7 (B=5)	Dimensi 8 (B=15)	Dimensi 9 (B=15)	Dimensi 10 (B=10)	Class
Disiplin	Disiplin	Bisa_Dipercaya	Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Selalu_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istimewa
Disiplin	Disiplin	Bisa_Dipercaya	Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istimewa
Disiplin	Disiplin	Bisa_Dipercaya	Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istimewa
Disiplin	Disiplin	Bisa_Dipercaya	Konsisten	Kurang_Semangat	Inisiatif	Sangat_Menguasai	Cukup_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istimewa
Disiplin	Disiplin	Bisa_Dipercaya	Konsisten	Kurang_Semangat	Inisiatif	Sangat_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istimewa
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Selalu_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istimewa
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istimewa
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Cukup_Semangat	Kurang_inisiatif	Cukup_Menguasai	Cukup_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Baik
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Inisiatif	Sangat_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Inisiatif	Cukup_Menguasai	Selalu_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istimewa
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Inisiatif	Cukup_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istimewa
Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Inisiatif	Cukup_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
Cukup_Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Semangat	Cukup_inisiatif	Cukup_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istimewa
Cukup_Disiplin	Cukup_Disiplin	Bisa_Dipercaya	Cukup_Konsisten	Cukup_Semangat	Inisiatif	Sangat_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istimewa
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Konsisten	Semangat	Inisiatif	Sangat_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istimewa
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Konsisten	Semangat	Inisiatif	Sangat_Menguasai	Cukup_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Baik
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Konsisten	Semangat	Inisiatif	Sangat_Menguasai	Cukup_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Baik
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Inisiatif	Cukup_Menguasai	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Inisiatif	Cukup_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Cukup
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Cukup_inisiatif	Cukup_Menguasai	Cukup_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Baik
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Cukup_inisiatif	Cukup_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Baik
Cukup_Disiplin	Cukup_Disiplin	Cukup_Bisa_Dipercaya	Cukup_Konsisten	Kurang_Semangat	Cukup_inisiatif	Cukup_Menguasai	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Cukup

Data mining is the process of finding patterns, new knowledge from big data sets, also known as knowledge discovery in databases [9]. According to Ketui.et all; 2019 that data mining is a process of finding patterns and relationships from a large number of datasets with various methods such as machine learning, statistics, and database systems[6]. This study analyzed the performance appraisal of employees by applying the naïve Bayes algorithm and C4.5 to determine accuracy by applying the k-fold evaluation model Cross Validation and Confusion Matrix.

NAÏVE BAYES ALGORITHM

The Naive Bayes algorithm is a simple probability classifier that calculates a set of probabilities by calculating the frequency and combination of values in a given data set[10].
This method requires only a small amount of training data to determine the parameter estimates required in the classification process. NBC often performs much better in the most complex of real-world situations than might be expected[11].
The general form of the Naïve Baye Algorithms can be seen in equation II.1. [12]

$$\mathbf{P}\ (\mathbf{C}|\mathbf{X}) = \frac{\mathbf{P}(\mathbf{X}|\mathbf{C}).\mathbf{P}(\mathbf{c})}{\mathbf{P}(\mathbf{X})} \quad \text{..... (II.1)}$$

Where :

X	: Data with an unknown class
C	: The data hypothesis is a specific class
P (C | X)	: The probability of hypothesis C under condition X (posteriori probability)
P (c)	: Probability hypothesis c (prior probability)
P (x | c)	: Probability X based on the conditions in hypothesis C
P (x)	: Probability X (Doctor prior probability)

ALGORITHM C4.5

The decision tree has become a well-known method of data mining. The increasing use of the method is due to a number of benefits, among others, is the fact that it is easy to understand and interpret[13]. Decision trees also require minimal data preparation effort, are capable of handling both numerical and categorical data, and they perform very well with large data sets in a relatively short amount of time[14].

The C4.5 algorithm steps for building a decision tree, among others[12]:

1. Select attribute as root, attribute selection as root is based on the highest gain value of existing attributes. To calculate the highest gain value, use the following II.2.1 equation:

   Gain (S, A) = Entropy (S) - $\sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$ (II.2.1)

   Where:
   A = Attribute
   S = Case Set

n = number of partitions attribute A
Si = Proportion of Si to S
S = number of cases in S

Meanwhile, to calculate the Entropy value can be seen in equation II.2.2 below:

Entropy (S) = $\sum_{i=0}^{n} -\boldsymbol{pi} * \boldsymbol{log_2 pi}$ (II.2.2)

Where:

S = Case set
n = number of partitions S
Pi = the proportion of Si to S

2. Create a branch for each member of the Node.
3. Check if the entropy value of any Node member is zero. If present, determine which leaves are formed. If all Node member entropy values are zero, the process stops.
   If any member of the Node has an entropy value greater than zero, repeat the process from the beginning with Node as a condition until all members of the Node are zero. Node is the attribute that has the highest gain value of the existing attributes.

EVALUATION OF DATA MINING CLASSIFICATION PERFORMANCE

Classification system performance describes how well the system classifies data. Evaluation is intended to test the data mining classification model to determine system performance. According to[15] The methods that are quite popular for evaluating the performance of data mining classifications are:

1. k-fold Cross Validation
   Cross Validation using cross validation will be carried out as many experiments as k. Each experiment will use one test data and the k-1 part will become training data, then the test data will be exchanged with a single training data so that for each trial different test data will be obtained[16]. Training data is data that will be used to test learning while data is data that has not been used as learning and will be used as data for testing the truth or accuracy of learning outcomes.

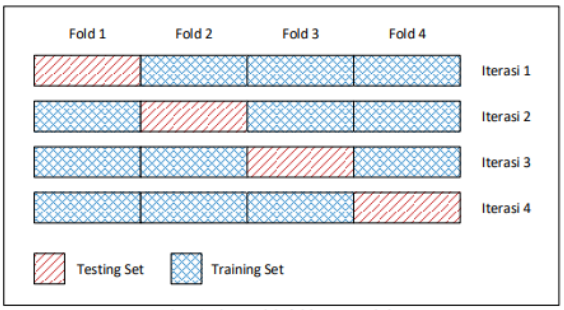


Figure 1. Illustration of k-fold cross validation

2. Confusion Matrix
   In data mining, to measure the performance of the resulting model, one of which uses a confusion matrix (accuracy). Confusion matrix is a method used to calculate accuracy in the concept of data mining. Precision or confidence is the proportion of positive predicted cases that are also true positives on the actual data. Recall or sensitivity is the proportion of true positive cases predicted to be positive correctly[17].
   The confusion matrix contains information about the actual class and predictive class of a classification process. Basically confusion matrix compares the results of the classification performed by a system with the actual classification results[17]. Table 6 shows the classification confusion matrix for the two classes.

Table 6. The Confusion Matrix Model

Actual	Classified as	
	+	-
+	True positives (A)	False negatives (B)
-	False positives (C)	True negatives (D)

Accuracy calculation with confusion matrix table is as follows:

Accuracy = (A + D) / (A + B + C + D)or

Accuracy = $\frac{\boldsymbol{Total\ Correctly}}{\boldsymbol{\Sigma\ Actual}}$

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Precision is defined as the ratio of the relevant items selected to all selected items. Precision can be interpreted as a match between requests for information and answers to these requests. The formula for precision is:

Precision = A / (C + A), or

Precision of class $= \frac{Correctly\ Predicted}{\Sigma\ Predicted}$

Weight Average Precision $= \Sigma\ \frac{Actual\ Class}{Total\ Precision}\ x\ Precision\ of\ class$

Recall is defined as the ratio of the relevant items selected to the total number of relevant items available. Recall is calculated by the formula:

Recall = A / (A + D) or

Recall of class $= \frac{Correctly\ Classified}{\Sigma\ Actual}$

Weight Average Recall $= \Sigma \frac{Actual\ Class}{Total\ Recall}\ x\ Recall\ of\ class$

Precision and Recall can be rated in numeric form by using a percentage calculation (1-100%) or by using a number between 0-1. The recommendation system will be considered good if the precision and recall values are high.

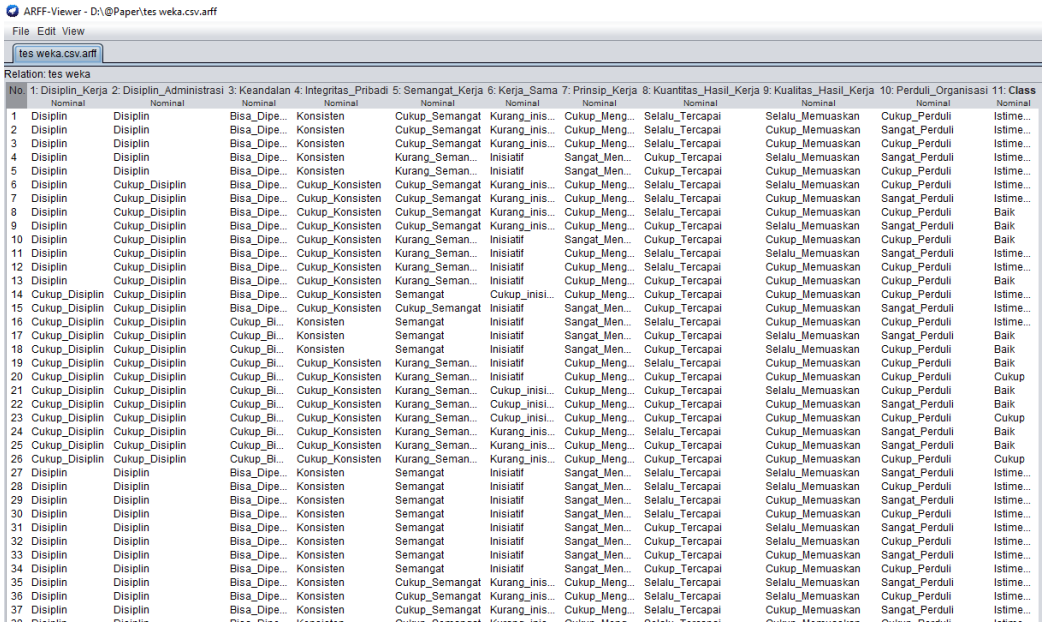
3. Result and Discussion

Experiments were carried out by selecting the Naïve Bayes algorithm and C4.5 in classifying employee performance in order to determine the classification model for the training data used. Testing was carried out using one of the popular data mining tools to solve many problems in data mining, namely WEKA version 3.8[18].

Waikato Environment for Knowledge Analysis (Weka) is a suite of machine knowledge software. Weka contains a compilation of visualization tools and algorithms for data analysis and predictive modeling, with a graphical user interface[19].

The steps taken in carrying out the data testing process to determine the classification of employee performance are as follows:

1. Collect employee data and carry out weight conversion to data.
2. Configure the excel file extension which will be presented in arff format.
3. Perform the calculation process (data execution) to determine the accuracy of the two methods used.


ARFF-Viewer - D:\@Paper\tes weka.csv.arff

File Edit View

tes weka.csv.arff

Relation: tes weka

No	1: Disiplin_Kerja	2: Disiplin_Administrasi	3: Keandalan	4: Integritas_Pribadi	5: Semangat_Kerja	6: Kerja_Sama	7: Prinsip_Kerja	8: Kuantitas_Hasil_Kerja	9: Kualitas_Hasil_Kerja	10: Perduli_Organisasi	11: Class
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istime...
2	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istime...
3	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
4	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Kurang_Seman...	Inisiatif	Sangat_Men...	Cukup_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istime...
5	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Kurang_Seman...	Inisiatif	Sangat_Men...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
6	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istime...
7	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istime...
8	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
9	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Baik
10	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Kurang_Seman...	Inisiatif	Sangat_Men...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
11	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Kurang_Seman...	Inisiatif	Cukup_Meng...	Selalu_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istime...
12	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Kurang_Seman...	Inisiatif	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
13	Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Kurang_Seman...	Inisiatif	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
14	Cukup_Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Semangat	Cukup_inisi...	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
15	Cukup_Disiplin	Cukup_Disiplin	Bisa_Dipe...	Cukup_Konsisten	Cukup_Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istime...
16	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
17	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Baik
18	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Baik
19	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Inisiatif	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Baik
20	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Inisiatif	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Cukup
21	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Cukup_inisi...	Cukup_Meng...	Cukup_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Baik
22	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Cukup_inisi...	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Baik
23	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Cukup_inisi...	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Cukup
24	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Baik
25	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Kurang_inis...	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Baik
26	Cukup_Disiplin	Cukup_Disiplin	Cukup_Bi...	Cukup_Konsisten	Kurang_Seman...	Kurang_inis...	Cukup_Meng...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Cukup
27	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Selalu_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istime...
28	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Selalu_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istime...
29	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istime...
30	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Selalu_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
31	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istime...
32	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istime...
33	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istime...
34	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Semangat	Inisiatif	Sangat_Men...	Cukup_Tercapai	Cukup_Memuaskan	Cukup_Perduli	Istime...
35	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Selalu_Memuaskan	Sangat_Perduli	Istime...
36	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Selalu_Memuaskan	Cukup_Perduli	Istime...
37	Disiplin	Disiplin	Bisa_Dipe...	Konsisten	Cukup_Semangat	Kurang_inis...	Cukup_Meng...	Selalu_Tercapai	Cukup_Memuaskan	Sangat_Perduli	Istime...

Figure 2. Converting CSV to Arff

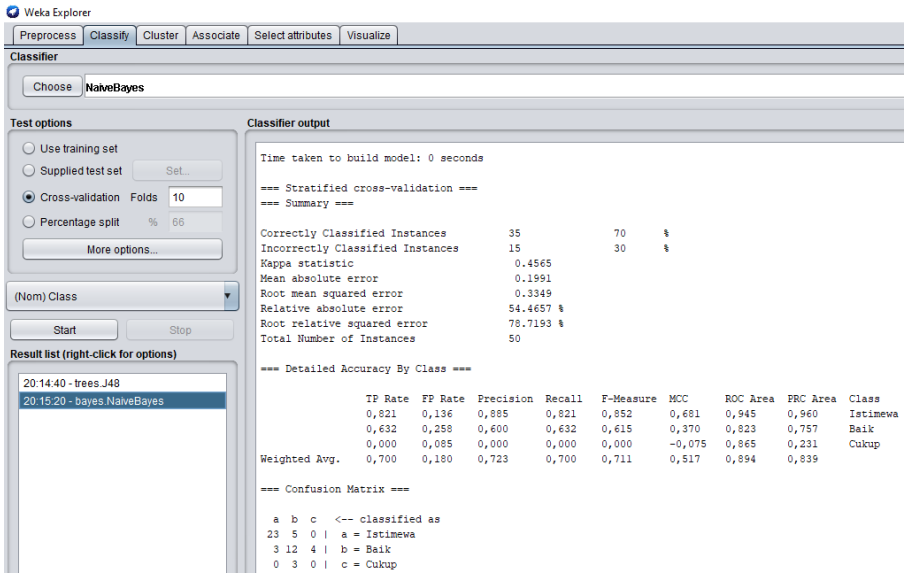


Figure 3. Experiment with Naïve Bayes Algorithm

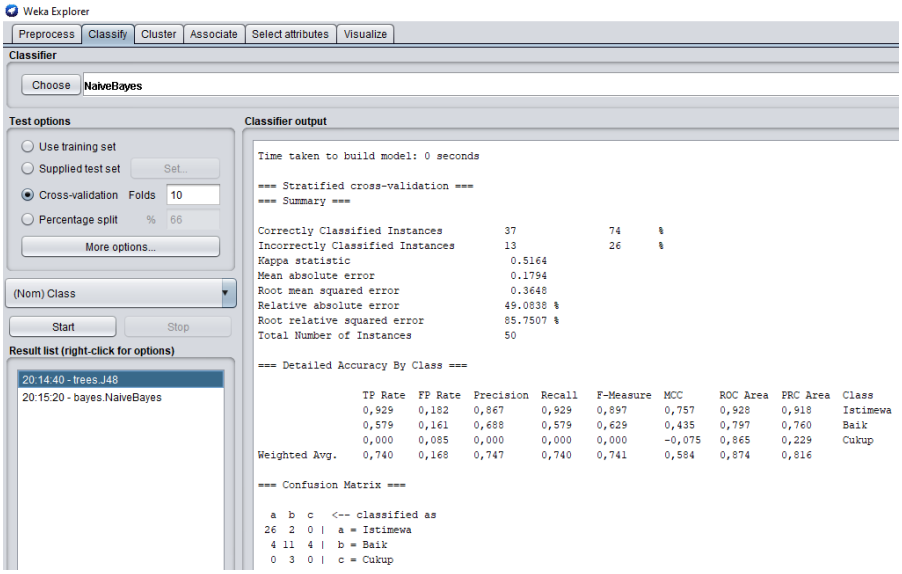


Figure 4. Experiments with the C4.5 Algorithm

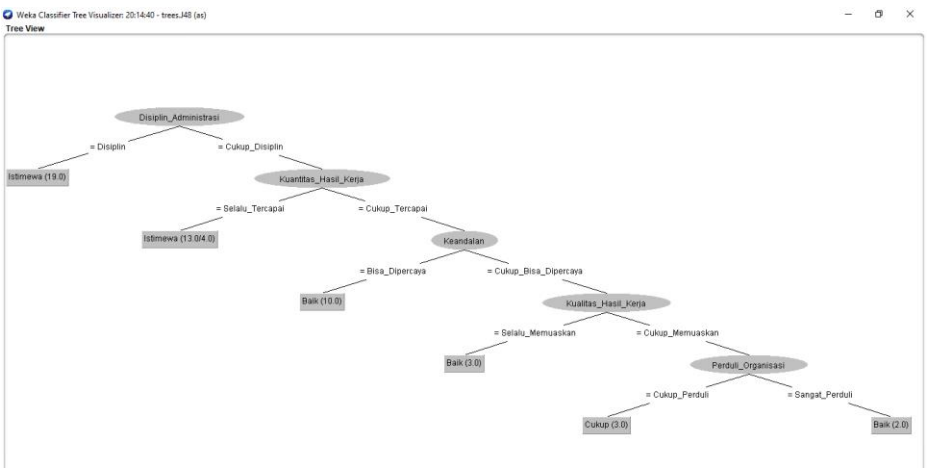


Figure 5. Decision Tree Results C4.5

The rules generated from the decision tree of the determining factors for Employee Performance Appraisal based on Figure 6 are as follows:

p-ISSN: 2301-5373
e-ISSN: 2654-5101

1. If Administrative Discipline = Discipline, Then Employee Performance = Special;
2. If Administrative Discipline = Discipline, Then Quantity of Work = Always Achieved, Then Employee Performance = Special;
3. If Administrative Discipline = Discipline, Then Quantity of Work = Sufficiently Achieved, Then Reliability = Can be trusted, Then Employee Performance = Good;
4. If Administrative Discipline = Discipline, Then Quantity of Work = Fairly Achieved, Then Reliability = Sufficiently Trustworthy, Then Quality of Work = Always Satisfactory, Then Employee Performance = Good;
5. If Administrative Discipline = Discipline, Then Quantity of Work = Sufficiently Achieved, Then Reliability = Sufficiently Trustworthy, Then Quality of Work = Sufficiently Satisfactory, Then Care for Organization = Enough to Care, Then Employee Performance = Enough;
6. If Administrative Discipline = Discipline, Then Quantity of Work = Sufficiently Achieved, Then Reliability = Sufficiently Trustworthy, Then Quality of Work = Sufficiently Satisfactory, Then Care for Organization = Very Concerned, Then Employee Performance = Good;

TESTING AND EVALUATION OF THE MODEL

The model obtained from the two methods of Naïve Bayes and C4.5 is then tested using k-fold cross validation. Cross validation is a simple form of statistical technique. The number of standard folds to predict the error rate of data is to use 2-fold, 5-fold, 6-fold and 10-fold Cross Validation.[20]. The distribution of Data Training and Data Testing will be randomized automatically by tools.

The evaluation model used in this study is a Confusion Matrix model. The confusion matrix contains information about the actual class and predictive class of a classification process. Basically confusion matrix compares the results of the classification performed by a system with the actual classification results[15].

Table 7. Experiment Results with K-Cross Validation

Cross Validation	2 K-Fold		5 K-Fold		6 K-Fold		10 K-Fold	
Algoritma	Naïve Bayes	C4.5	Naïve Bayes	C4.5	Naïve Bayes	C4.5	Naïve Bayes	C4.5
Accuracy	76	74	68	76	78	80	70	74
Presisi	0,798	0,607	0,705	0,773	0,798	0,794	0,723	0,747
Recall	0,760	0,740	0,680	0,760	0,780	0,800	0,700	0,740

From the table above, it can be seen that the algorithm with the best accuracy, precision and recall values lies in 6 Cross Validation, so that the confusion matrix table is found as shown in table 8.

Table 8. Table Confusion Matrix (6 Cross Validation)

Naïve Bayes		Actual Values			
		Istimewa	Baik	Cukup	Total Actual
Predicted Values	Istimewa	24	4	0	28
	Baik	2	14	3	19
	Cukup	0	2	1	3
	Total Predicted	26	20	4	50

C4.5		Actual Values			
		Istimewa	Baik	Cukup	Total Actual
Predicted Values	Istimewa	28	0	0	28
	Baik	4	12	3	19
	Cukup	0	3	0	3
	Total Predicted	32	15	3	50

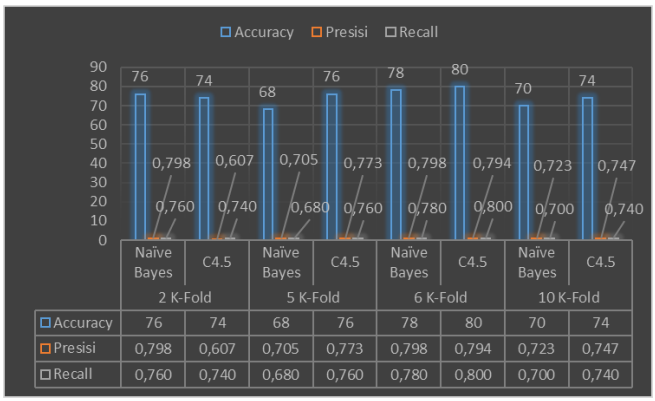
From the table above, we can calculate the accuracy, precision and recall values with the calculation of confusion matrix for the Naïve Bayes Algorithm:

1. Presision of class $= \frac{Correctly\ Predicted}{\Sigma\ Predicted}$
   against Class A $= \frac{24}{26} = 0{,}923$
   against Class B $= \frac{14}{20} = 0{,}7$
   against Class C $= \frac{1}{4} = 0{,}25$
2. Recall of class $= \frac{Correctly\ Classified}{\Sigma\ Actual}$
   against Class A $= \frac{24}{28} = 0{,}857$
   against Class B $= \frac{14}{19} = 0{,}737$
   against Class C $= \frac{1}{3} = 0{,}333$
3. Accuracy $= \frac{Total\ Correctly}{\Sigma\ Actual} = \frac{24+14+1}{50} = 0{,}78x\ 100 = 78\%$

4. Weight Average Precision $= \Sigma \frac{Actual\ Class}{Total\ Precision}\ x\ Precision\ of\ class$
$= \left(\frac{28}{50}\ x\ 0{,}923\right) + \left(\frac{19}{50}\ x\ 0{,}7\right) + \left(\frac{3}{50}\ x\ 0{,}25\right)$
$= 0{,}517 + 0{,}266 + 0{,}015$
$= 0{,}798$
5. Weight Average Recall $= \Sigma \frac{Actual\ Class}{Total\ Recall}\ x\ Recall\ of\ class$
$= \left(\frac{28}{50}\ x\ 0{,}857\right) + \left(\frac{19}{50}\ x\ 0{,}737\right) + \left(\frac{3}{50}\ x\ 0{,}333\right)$
$= 0{,}480 + 0{,}280 + 0{,}020$
$= 0{,}78$

While the validation test with confusion matrix calculation for the C4.5 Algorithm is as follows:

1. Presision of class $= \frac{Correctly\ Predicted}{\Sigma\ Predicted}$
against Class A $= \frac{28}{32} = 0{,}875$
against Class B $= \frac{12}{15} = 0{,}8$
against Class C $= \frac{0}{3} = 0$
2. Recall of class $= \frac{Correctly\ Classified}{\Sigma\ Actual}$
against Class A $= \frac{28}{28} = 1$
against Class B $= \frac{12}{19} = 0{,}631$
against Class C $= \frac{0}{3} = 0$
3. Accuracy $= \frac{Total\ Correctly}{\Sigma\ Actual} = \frac{28+12+0}{50} = 0{,}8x\ 100 = 80\%$
4. Weight Average Precision $= \Sigma \frac{Actual\ Class}{Total\ Precision}\ x\ Precision\ of\ class$
$= \left(\frac{28}{50}\ x\ 0{,}875\right) + \left(\frac{19}{50}\ x\ 0{,}8\right) + \left(\frac{3}{50}\ x\ 0\right)$
$= 0{,}49 + 0{,}304 + 0$
$= 0{,}794$
5. Weight Average Recall $= \Sigma \frac{Actual\ Class}{Total\ Recall}\ x\ Recall\ of\ class$
$= \left(\frac{28}{50}\ x\ 1\right) + \left(\frac{19}{50}\ x\ 0{,}631\right) + \left(\frac{3}{50}\ x\ 0\right)$
$= 0{,}56 + 0{,}240 + 0$
$= 0{,}800$



	2 K-Fold		5 K-Fold		6 K-Fold		10 K-Fold	
	Naïve Bayes	C4.5	Naïve Bayes	C4.5	Naïve Bayes	C4.5	Naïve Bayes	C4.5
Accuracy	76	74	68	76	78	80	70	74
Presisi	0,798	0,607	0,705	0,773	0,798	0,794	0,723	0,747
Recall	0,760	0,740	0,680	0,760	0,780	0,800	0,700	0,740

Figure 6. Comparison Graph of Accuracy, Precision, and Recall Results

4. Conclusion

The results of data mining processing experiments on employee performance appraisal data showed that the Naïve Bayes algorithm obtained an accuracy value of 78%, with a Precision value of 0.798, and a Recall value of 0.780 with the highest Cross Validation of 6 Fold Cross Validation. While the C4.5 algorithm is able to provide accuracy values by obtaining an accuracy value of 80% of the data, with a Precision value of 0.794, and a Recall value of 0.800 with the highest Cross Validation of 6 Fold Cross Validation. The results of the comparison of the accuracy of the two algorithms can be seen in Figure 7.

p-ISSN: 2301-5373
e-ISSN: 2654-5101

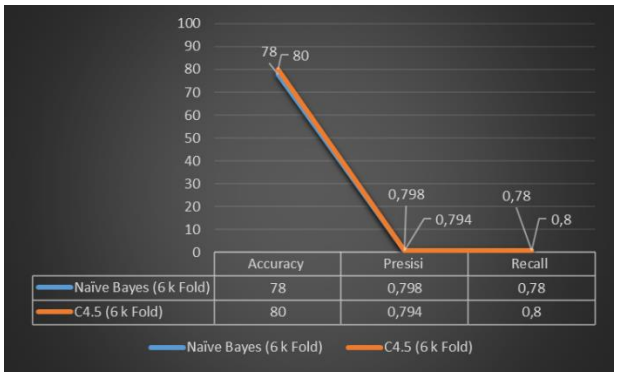


Figure 7. Comparison Graph of Accuracy, Precision, and Recall Results for 6 Cross Validation

Based on the results of research that has been carried out in the case of employee performance appraisal by applying the Naïve Bayes algorithm and C4.5, the following conclusions can be drawn

1. In predicting Employee Performance Appraisal using Naïve Bayes and C4.5 modeling methods with a total of 50 data, that the Decision Tree (C4.5) method is better at predicting Employee Performance Appraisal at Potensi Utama of University.
2. The implementation of the two classification algorithms was successfully carried out, where the two algorithms provided a fairly good accuracy performance, namely an average> = 70%. Modeling with the Naïve Bayes Algorithm obtains the highest accuracy value with 6 Fold Cross Validation obtaining the highest average accuracy rate of 78% with a Precision value of 0.798 and and Recall of 0.78. While modeling with the C4.5 Algorithm is able to provide good accuracy values by obtaining an accuracy value of 80% with a Precision value of 0.794 and Recall of 0.80

References

[1] A. S. Lather, R. Malhotra, P. Saloni, P. Singh, and S. Mittal, “Prediction of employee performance using machine learning techniques,” *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3373477.3373696.

[2] I. A. A. Amra and A. Y. A. Maghari, “Students performance prediction using KNN and Naïve Bayesian,” *ICIT 2017 - 8th Int. Conf. Inf. Technol. Proc.*, no. December 2018, pp. 909–913, 2019, doi: 10.1109/ICITECH.2017.8079967.

[3] A. Alam, D. A. F. Alana, and C. Juliane, “Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes,” *Sinkron*, vol. 8, no. 4, pp. 2641–2650, 2023, doi: 10.33395/sinkron.v8i4.12998.

[4] R. A. Saputra, S. Wasiyanti, and D. Pribadi, “Information Gain Pada Algoritma C4.5 Untuk Klasifikasi Penerimaan Bantuan Pangan Non Tunai (Bpnt),” *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 25, 2021, doi: 10.21927/ijubi.v4i1.1757.

[5] G. Kostopoulos, N. Fazakis, S. Kotsiantis, and K. Sgarbas, “Multi-objective Optimization of C4.5 Decision Tree for Predicting Student Academic Performance,” *10th Int. Conf. Information, Intell. Syst. Appl. IISA 2019*, pp. 1–4, 2019, doi: 10.1109/IISA.2019.8900771.

[6] I. Shingari and D. Kumar, “Predicting Student Performance Using Classification Data Mining Techniques,” *Int. J. Comput. Sci. Eng.*, vol. 6, no. 7, pp. 43–48, 2018, doi: 10.26438/ijcse/v6i7.4348.

[7] Y. Andini, J. T. Hardinata, and Y. P. Purba, “Penerapan Data Mining Terhadap Tata Letak Buku Di Perpustakaan Sintong Bingei Pematangsiantar Menggunakan Metode Apriori,” *J. TIMES*, vol. 11, no. 1, pp. 9–15, 2022, doi: 10.51351/jtm.11.1.2022661.

[8] N. Ketui, W. Wisomka, and K. Homjun, “Using classification data mining techniques for students performance prediction,” *ECTI DAMT-NCON 2019 - 4th Int. Conf. Digit. Arts, Media Technol. 2nd ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng.*, pp. 359–363, 2019, doi: 10.1109/ECTI-NCON.2019.8692227.

[9] A. M.O., O. A.O., and W. W.F., “Teachers’ Performance Evaluation in Higher Educational Institution using Data Mining Technique,” *Int. J. Appl. Inf. Syst.*, vol. 10, no. 7, pp. 10–15, 2016, doi: 10.5120/ijais2016451524.

[10] D. Fahrudy *et al.*, “Intelligent System for Classification of Student Personality With Naive Bayes Algorithm,” *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 1, pp. 1–9, 2022, doi: 10.31598/sintechjournal.v5i1.969.

[11] A. Rawat and A. Choubey, “A Survey on Classification Techniques For Text Mining,” *Int. Conf.*

Adv. Comput. Commun. Syst., vol. 2, no. 3, pp. 436–443, 2016, doi: 10.17148/IJARCCE.2016.53143.

[12] Y. A. Gerhana, I. Fallah, W. B. Zulfikar, D. S. Maylawati, and M. A. Ramdhani, "Comparison of naive Bayes classifier and C4.5 algorithms in predicting student study period," *J. Phys. Conf. Ser.*, vol. 1280, no. 2, 2019, doi: 10.1088/1742-6596/1280/2/022022.

[13] M. Kartiwi, T. Gunawan, T. Arundina, and M. Omar, "Decision tree optimization for Sukuk rating prediction," *J. Fundam. Appl. Sci.*, vol. 10, no. 2S, pp. 323–332, 2018, doi: 10.4314/jfas.v10i2s.25.

[14] M. Sadikin, R. Rosnelly, and T. Surya Gunawan, "Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 1100–1109, 2020, doi: http://dx.doi.org/10.30865/mib.v4i4.2434.

[15] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4 . 5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: https://ejournal.upm.ac.id/index.php/energy/article/view/111/449

[16] Siswanto, Abdussomad, W. Gata, N. K. Wardhani, G. Gata, and B. H. Prasetyo, "The feasibility of credit using C4.5 algorithm based on particle swarm optimization prediction," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, pp. 416–421, 2019, doi: 10.23919/EECSI48112.2019.8977074.

[17] T. Rosandy, "Perbandingan Metode Naive Bayes Classifier dengan Metode Decision Tree Untuk Menganalisa Kelancaran Pembiayaan," *J. TIM Darmajaya*, vol. 02, no. 01, pp. 52–62, 2016.

[18] V. R. Sajja, P. J. Lakshmi, D. B. Naik, and H. K. Kalluri, "Prediction of student performance using decicion tree classifier," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 247–255, 2019, doi: 10.35940/ijeat.E7520.088619.

[19] M. S.Menaka and D. G.Kesavara, "A STUDY ON E-LEARNING SYSTEM TO ANALYSE STUDENT PERFORMANCE USING DATA MINING," *JETIR*, vol. 6, no. 4, pp. 75–81, 2019, doi: 10.13140/RG.2.2.33861.40164.

[20] A. N. Kholifah and N. Insani, "Analisis Klasifikasi Pada Nasabah Kredit Koperasi X Menggunakan Decision Tree C4. 5 Dan Naive Bayes," *J. Pendidik. Mat. dan Sains*, vol. 5, no. 6, pp. 1–8, 2016.