SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

K-Means and *Naive Bayes* Algorithms for Evaluation of Education Personnel Performance Based on SPMI Standards

Eko Wahyudi [1)*] , Rian Farta Wijaya [2)] , Khairul [3)]
[1,2,3)] Master of Information Technology Study Program, Panca Budi Development University, Indonesia
[1)] ekow7952@gmail.com , [2)] rianfartawijaya@dosen.pancabudi.ac.id , [3)]khairul@dosen.pancabudi.ac.id

Submitted : July 17, 2024 | **Accepted** : 25 July 2024 | **Published** : July 2024

Abstract: This research compares the K-Means and Naive Bayes algorithms in evaluating the performance of educational staff based on SPMI standards at STMIK Triguna Dharma. The main objective is to identify the effectiveness of the two algorithms in grouping performance evaluation data and determine the advantages and disadvantages of each method. Primary data was obtained through surveys and interviews, while secondary data came from institutional archives. The K-Means algorithm shows 100% accuracy with the ability to group educational staff into very good, good, quite good, poor and poor performance categories. Meanwhile, the Naive Bayes algorithm shows 91% accuracy, with 100% precision results for the "good" and "fairly good" categories. These results indicate that K-Means is more effective in grouping educational staff based on performance evaluation compared to Naive Bayes. This research makes a significant contribution in the field of evaluating the performance of educational staff and offers insights for a more effective implementation of SPMI in higher education.

Keywords: Performance evaluation, educational staff, K-Means, Naive Bayes, SPMI.

INTRODUCTION

The university education system is one of the main pillars in a country's intellectual development. Higher education is a separate education system with smaller work units, namely study programs, involving facilities and infrastructure components, students (students), curriculum teaching materials and human resources (Samudin, 2023) . Then within the work unit there is an organizational structure, such as components that support the continuity of the function and mission of higher education, namely educational staff. Higher education education personnel can be divided into two main categories, namely lecturers and staff. This research study discusses one of the educational staff appointed by the STMIK Triguna Dharma Family Development Foundation, consisting of administrative employees, library officers, financial officers, student affairs officers and technical employees who support university operations and administration.

The role of educational staff at universities is very important, because educational staff are required to be more focused on realizing their performance targets. One of the keys to managing higher education performance can be implemented through the Internal Quality Assurance System (SPMI) higher education performance indicators. According to Law no. 12 of 2012 article 52 which explains Quality Assurance in Higher Education emphasizes that Quality Assurance in Higher Education is a systemic activity to improve the quality of higher education in a planned and sustainable manner. However, the Internal Quality Assurance System in developing a university to achieve standardization requires an internal quality assurance system (SPMI) pattern, namely planning, implementation, evaluation, control and improvement (Arifudin, 2019) .

This means that every university must follow the standardization of educational processes implemented in Indonesia. National Higher Education Standards are contained in Minister of Education and Culture Regulation no. 3 of 2020. Then in 2023 the Ministry of Education and Culture issued Permendikbudristek 53 of 2023 article 2 paragraph (2) which explains that the Higher Education Standards as intended in paragraph (1) consist of SN-Dikti and Higher Education Standards set by universities. Internal Quality Assurance System Standards (SPMI) are one of the Employee Performance Evaluation Standards as Higher Education Standards set by universities which serve as guidelines for measuring *the Key Performance Indicators* of educational staff. *Key Performance Indicator* evaluation of educational staff is carried out by means of a Performance Assessment in each work unit.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Performance assessment is the process of the work of personnel in an institution through performance assessment instruments, performance assessment to evaluate personnel's work performance by comparing it with standard performance standards (Saputra et al., 2022) .

In essence, performance appraisal (work performance) can be considered a tool to verify that individuals meet predetermined performance standards (Harahap, 2019) . Performance standards are carried out by means of Performance evaluations are usually carried out periodically at certain time intervals in most institutions. Evaluations are carried out once or twice a year. In general, workers are first evaluated towards the end of the preparation period. Evaluating new employees several times during their first year of work is also common practice (Al-Muhtadi & Sumiati, 2023) . The aim of evaluating educational staff is to improve performance in order to achieve the vision and mission of higher education, achieve SPMI standards, and improve the quality of education. The quality of education is not only determined by facilities and curriculum but also by the performance of the educational staff who are part of higher education institutions. However, in the implementation of the evaluation of educational staff, there are still many teaching staff who have not carried out the main tasks and functions of the unit in accordance with STMIK Triguna Dharma procedures. Therefore, to overcome the problem, it is necessary to have an approach that can be used by data mining science, such as a comparison of the *K-Means* and *Naïve Bayes algorithms*.

In the literature, one approach to the *K-Means algorithm* is a non-hierarchical grouping method that partitions objects into several groups *(clusters)* (Sidik et al., 2023) . *Naive Bayes* is an algorithm used in *text mining*. The *Naive Bayes* algorithm predicts future events based on previous experience so it is known as *Bayes' Theorem* (Duei Putri et al., 2022) . Therefore, the purpose of the algorithm comparison is to compare the effectiveness of *K-Means* and *Naïve Bayes* in evaluating data presenting educational staff performance assessments and determining the strengths or weaknesses of each educational staff performance evaluation.

LITERATURE REVIEW

Data Mining

Data Mining is a process of analyzing hidden data patterns according to various perspectives for categorization into useful information, which is collected in general areas, data *warehouses* for efficient analysis, data mining algorithms, facilitating business decision making, other information (Muhammad Arhami, 2020) . Objective from data mining is For find truth or conclusion through analysis data , with objective For find pattern And equality in gathering data Which has been sequenced correctly (Risqi Ananda et al., 2023) . Data mining has a number of objective main , namely:

1. As a means of providing an *explanatory explanation* of a condition or phenomenon.
2. As a means of confirming (*confirmatory*) a hypothesis or assumption through data analysis.
3. As a means of predicting *an* event or phenomenon that may occur in the future through data analysis.
4. as a *prescriptive* means to provide recommendations or suggestions regarding an event or phenomenon that occurs.

Data mining is a step in the *Knowledge Discovery in Database* (KKD) process (Sharif, 2019) . *Knowledge Discovery in Database* (KDD) is a process that aims to find knowledge in large data sets. This process is often referred to as data mining (Setianingsih & Ali, 2023) .

K-Means Algorithm with *Clustering Technique*

Clustering is a process where a number of data are grouped into groups that have similar characteristics and these data have the same characteristics within the group, and simultaneously differentiate themselves from data that is combined in other groups (Andini et al., 2022) . Then one of the commonly used *clustering methods is the K-means method.* According to (Ningsi et al., 2021) the *K-means* method is an approach to data analysis or data mining method that is *unsupervised* . Then the *K-means method* is also the oldest and most widely used clustering algorithm in various small to medium applications because of its ease of implementation (Dr. Suyanto, 2019) . The steps for grouping data using the *K-means method* are as follows (Adi & Susetyo, 2020) :

1. Determine the number *of clusters* , for example k = 3
2. Allocate data into groups randomly
3. Calculate the group center (*centroid* /average) from the data in each group. The *centroid* location for each group is taken from the average (*mean*) of all data values for each feature. If M represents the amount of data in a group, i represents the ith feature in a group, and p represents the dimension of the data, then the equation for calculating the centroid of the ith feature uses equation 1.

$$\mathbf{Ci} = \frac{1}{M}\sum_{J=1}^{M} \boldsymbol{xj} \quad (1)$$

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

4. Allocate each data to the nearest *centroid* /average. There are several ways that can be used to measure the distance of data to the center of the group, including Euclidean. Distance measurements in Euclidean distance space can be found using equation 2.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (2)$$

5. Returning to step 3 is necessary if there is data that is still changing groups, there is a change in the centroid value that exceeds a predetermined threshold, or if the change in the value of the objective function is still above a predetermined threshold value.

Naïve Bayes algorithm

Naive Bayes (NB) algorithm is a simple method of classification based on probability theory. This method was first put forward by a British scientist named Thomas Bayes (Sovisii et al., 2022) . Bayes' theorem is a principle in probability theory related to conditional probability (Supriyadi, 2023) . In Bayes' theorem, conditional probability can be explained by a formula expressed as (Muhammad Arhami, 2020) :

$$\mathbf{P(H\,|} \frac{P\,(X\mid H)^{*}P(H)}{P(X)} \quad (3)$$

Internal Quality Assurance System (SPMI)

The quality of education is one of the important factors that determines the quality of the graduates produced (Saiful Anwar Matondang, 2023) . Minister of Research, Technology and Higher Education Regulation No. 62 of 2016 (Article 1) states that the quality of higher education is the conformity between the implementation of higher education and educational standards consisting of National Higher Education Standards and Higher Education Standards set by higher education institutions (Kemenristekdikti, 2016) . The Higher Education Quality Assurance System is a mechanism that aims to control and ensure the quality of the provision of higher education by universities in order to achieve established quality standards. This is a necessity regulated in Law no. 12 of 2012 concerning Higher Education and Regulation of the Minister of Research, Technology and Higher Education Number 62 of 2016. For STMIK Triguna Dharma Higher Education, the implementation of the Higher Education Quality Assurance System is not only a legal necessity, but also a strategic need. Therefore, internal quality assurance functions to support the achievement of academic targets, such as the suitability of academic degree classifications and the validity of information about academic quality (Fadhli, 2020) .

Additional Standards (Employee Performance Assessment Evaluation Standards)

Employee Performance Evaluation Standards are an important basis for improving the quality and effectiveness of employee performance in higher education. With this additional standard, STMIK Triguna Dharma confirms its commitment to developing superior and competent human resources. The implementation of this standard reflects STMIK Triguna Dharma's commitment to assessing and improving the performance of all higher education personnel, including administrative staff, security personnel and public services. Thus, performance evaluation does not only focus on academic aspects, but also involves all elements that support campus operations as a whole. Then Educational Personnel is also a job category involved in various supporting functions in educational institutions such as schools and universities. The role of STMIK Triguna Dharma educational staff includes various administrative, technical and operational tasks that support the smooth running of the academic and operational activities of the educational institution.

METHOD

This research can be categorized as quantitative research which uses a mixed approach, with primary and secondary data. Primary data obtained directly from education personnel, such as through surveys or interviews, provides an in-depth view of their performance based on the SPMI Standards. Meanwhile, secondary data, which may come from institutional archives or previous performance reports, provides additional context and information. By integrating these two types of data, this research can present a more comprehensive analysis, making it possible to understand not only individual performance but also contextual and historical factors that may influence compliance with the educational quality standards set by SPMI. This mixed approach enriches research with a broader and deeper perspective, allowing more holistic conclusions to be drawn regarding the evaluation of the performance of educational staff. Then in this research, primary data was collected by filling in a performance assessment criteria aspect form by education staff in the STMIK Triguna Dharma tertiary environment. The form contains criteria aspects related to performance assessment evaluation and implementation of Employee Performance Evaluation Evaluation Standards:

Table 1. Aspects of Performance Assessment Criteria

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

CODE	CRITERIA
K1	Personal Work Results
K2	Job Knowledge
K3	Initiative
K4	Mental Dexterity
K5	Attitude
K6	Work Discipline
K7	Teamwork Results
K8	Instruction Compliance
K9	Loyalty

Then, in accordance with the research objectives that have been set, the researcher wants to conduct a comparative analysis of the *K-Means* and *Naive Bayes algorithms* to evaluate the performance of educational staff based on SPMI standards . The conceptual framework above can be described as follows :

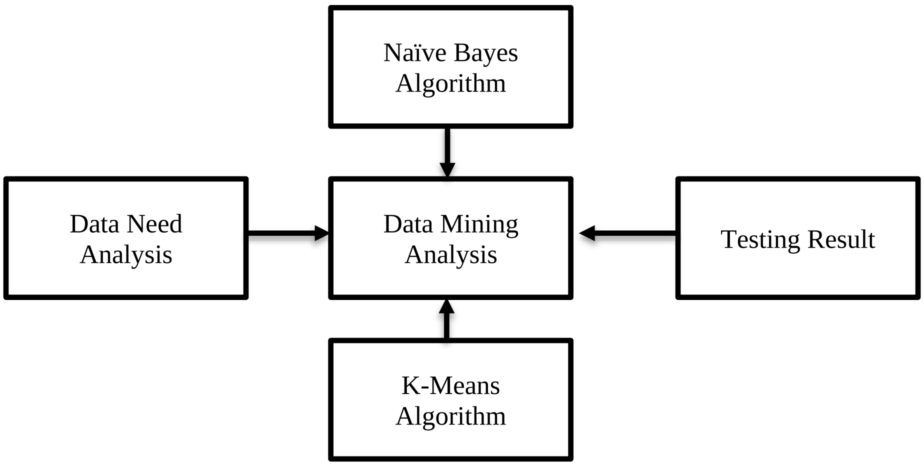

Figure 4. Research Conceptual Framework

Based on a conceptual framework of data mining which is produced based on data needs analysis for evaluating the performance of educational staff. This data mining is carried out by processing educational staff evaluation assessment data based on standard employee evaluation indicator criteria set by universities, namely quality assurance institutions, and analyzing the data using a comparison of the *Naïve Bayes algorithm* and the *K-Means algorithm* . Then determine the test statistics used to measure the significance of the differences between the results of the two algorithms.

RESULT

K-Means Algorithm Calculation Process

K-means algorithm calculation process , it is necessary to collect data with a total of 77 educational staff data which is carried out by a data mining process. The aim of the initial stage of data preparation is to ensure that the data used is appropriate to the problem to be solved, ensure the correctness of the data, and ensure the appropriate data format. The following is the initial stage of data preparation, namely data initialization.

1. Initialization of Education Personnel Performance Evaluation Data

Table 2. Educational Personnel Evaluation Assessment Data

Education Personnel Data	K1	K2	K3	K4	K5	K6	K7	K8	K9
P1	85	75	80	85	80	80	82	80	91

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

P2	88	89	85	85	85	75	85	76	75
P3	50	62	50	67	60	50	68	50	68
P4	80	85	80	80	83	86	82	80	86
P5	88	89	85	85	90	87	85	89	85
P6	75	77	65	80	80	85	80	65	81
P7	89	86	85	90	90	90	89	90	88
P8	80	85	70	80	85	65	75	65	70
P9	55	89	75	85	90	90	86	90	80
P10	85	88	87	85	86	85	83	89	80
…	…	…	…	…	…	…	…	…	…
P77	85	86	90	90	90	86	80	88	84

2. Determining the Initial Centroid

In applying the number of clusters (K), 5 *clusters have been determined* . After determining the number *of clusters* , the next step is to determine the initial center point of each *cluster* , which is called the centroid. Below is the centroid point that has been selected.

Table 3. Initial Centroid Data

CENTROID	**DATA**	**K1**	**K2**	**K3**	**K4**	**K5**	**K6**	**K7**	**K8**	**K9**	**ACHIEVEMENTS OF EDUCATIONAL PERFORMANCE PERFORMANCE**
C1	P23	95	95	90	92	94	93	95	95	95	Very Good Education Staff
C2	P21	80	82	83	84	86	90	80	85	85	Good Education Staff
C3	P6	75	77	65	80	80	85	80	65	88	Educational Staff Are Quite Good
C4	P52	67	65	60	67	65	63	62	67	68	Educational Staff Are Not Good
C5	P3	50	62	50	67	60	50	68	50	68	Bad Education Staff

3. *K-Means* Calculation Process

Calculate the distance between data and centroid using the *Euclidean formula* . Data will be assigned as a member of the closest cluster based on this distance calculation. The process of calculating the distance between variables from each data sample and the centroid can be explained as follows:

a. With Centroid P23 (95, 95, 90, 92, 94, 93, 95, 95, 95)

- Distance between P23 and point P23

$=\sqrt{\sum_{i=0}^{n}(x1-x2)^2}$

$=\sqrt{\begin{matrix}(95\text{-}95)^2+(95\text{-}95)^2+(90\text{-}90)^2+(92\text{-}92)^2+\\(94\text{-}94)^2+(93\text{-}93)^2+(95\text{-}95)^2+(95\text{-}95)^2+(95\text{-}95)^2\end{matrix}}$

= 0.00

b. With Centroid P21 (80, 82, 83, 84, 86, 90, 80, 85, 85)

- Distance between P23 and point P21

$=\sqrt{\sum_{i=0}^{n}(x1-x2)^2}$

$=\sqrt{\begin{matrix}(80\text{-}95)^2+(82\text{-}95)^2+(83\text{-}90)^2+(84\text{-}92)^2+\\(86\text{-}94)^2+(90\text{-}93)^2+(80\text{-}95)^2+(85\text{-}95)^2+(85\text{-}95)^2\end{matrix}}$

= 31.70

c. With Centroid P6 (75, 77, 65, 80, 80, 85, 80, 65, 88)

- Distance between P23 and point P6

$=\sqrt{\sum_{i=0}^{n}(x1-x2)^2}$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$= \sqrt{\begin{matrix}(75\text{-}95)^2 + (77\text{-}95)^2 + (65\text{-}90)^2 + (80\text{-}92)^2 + \\ (80\text{-}94)^2 + (85\text{-}93)^2 + (80\text{-}95)^2 + (65\text{-}95)^2 + (88\text{-}95)^2\end{matrix}}$

= 55.44

d. With Centroid P52 (67, 65, 60, 67, 65, 63, 62, 67, 68)

- Distance between P23 and point P52

$=\sqrt{\sum_{i=0}^{n}(x1-x2)^2}$

$= \sqrt{\begin{matrix}(67\text{-}95)^2 + (65\text{-}95)^2 + (65\text{-}90)^2 + (67\text{-}92)^2 + \\ (65\text{-}94)^2 + (63\text{-}93)^2 + (62\text{-}95)^2 + (67\text{-}95)^2 + (68\text{-}95)^2\end{matrix}}$

= 86.90

e. With Centroid P3 (50, 62, 50, 67, 60, 50, 68, 50, 68)

- Distance between P23 and point P3

$=\sqrt{\sum_{i=0}^{n}(x1-x2)^2}$

$= \sqrt{\begin{matrix}(50\text{-}95)^2 + (62\text{-}95)^2 + (50\text{-}90)^2 + (67\text{-}92)^2 + \\ (60\text{-}94)^2 + (50\text{-}93)^2 + (68\text{-}95)^2 + (50\text{-}95)^2 + (68\text{-}95)^2\end{matrix}}$

= 108.75

same calculation process up to object 77. The results of the literacy calculation 1 can be seen in the table below. The shortest distance is the calculation result that is closest to the cluster center. Meanwhile, WCV (*Within Cluster Variation*) results from the calculation of the closest distance to the cluster center.

Table 4. 1st Iteration Calculation Results

Education Personnel Data	**Distance to Point C1**	**Distance to Point C2**	**Distance to Point C3**	**Distance to Point C4**	**Distance to Point C5**	**JT Value**	**J.T**	**JT^2 value**
P1	37.74	16.88	24.84	52.54	74.45	16.88	C2	285.00
P2	37.75	23.45	34.32	56.29	77.37	23.45	C2	550.00
P3	108.75	81.31	60.32	30.28	0.00	0.00	C5	0.00
P4	34.84	9.43	23.60	53.70	75.78	9.43	C2	89.00
P5	21.14	13.56	38.11	66.85	89.94	13.56	C2	184.00
P6	55.44	29.44	7.00	40.10	58.37	7.00	C3	49.00
P7	16.76	16.37	40.10	71.35	93.77	16.37	C2	268.00
…	…	…	…	…	…	…	…	…

From table 2 there are *cluster memberships* as follows:

- C1
  {P11, P14, P15, P20, P23, P24, P25, P29, P31, P32, P34, P35, P37, P44, P48, P49, P53, P54, P57, P58, P59, P60, P61, P63, P66 }
- C2
  {P1, P2, P4, P5, P7, P9, P10, P12, P13, P16, P17, P21, P22, P26, P27, P36, P38, P39, P40, P41, P42, P43, P45, P46, P47 , P51, P55, P56, P62, P64, P65, P67, P68, P69,P71, P72, P73, P75, P77}
- C3
  {P6, P8, P28, P33}
- C4
  {P18, P19, P30, P50, P52, P70, P74, P76}
- C5
  {P3}

Then recalculate the ratio value by comparing the BCV (*Between Cluster Variation*) and WCV (*Within Cluster Variation) values* . The following is the calculation process for BCV (*Between Cluster Variation*) and WCV (*Within Cluster Variation)* .

- Calculate the BCV value with the formula adding up the results of the distance between each centroid.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

BCV = 0.00 + 31.70 + 55.44 + 86.90 + 108.75
= 282.80

- Calculating the WCV value is by squaring the distance calculation results and adding up all the exponents of the distance calculation results.
  WCV = 113 + 182 + 78 + 471 + 0 + 59 +……+ 0
  = 20636
- Calculate the ratio value by dividing the BCV value by WCV

BCV/WCV = 282.80/20636
= 0.013704

Then the next step to enter iteration-2 is to create a new centroid center. Like the table below :

Table 5. New Centroid

CENTROID	K1	K2	K3	K4	K5	K6	K7	K8	K9
C1	91	92	91	91	90	91	91	92	91
C2	82.69	84.28	82.64	84.23	86.31	85.82	82.82	85.31	82.05
C3	80	82.25	72.5	80	82.5	69.5	78.25	68.25	73.75
C4	68.38	67.00	67.13	74.38	74.63	69.50	66.75	70.88	67.00
C5	50	62	50	67	60	50	68	50	68

Then recalculate the ratio value by comparing the BCV (*Between Cluster Variation*) and WCV (*Within Cluster Variation) values* . The following is the calculation process for BCV (*Between Cluster Variation*) and WCV (*Within Cluster Variation)* .

- Calculate the BCV value with the formula adding up the results of the distance between each centroid.
  BCV = 9.36 + 23.31 + 48.32 + 78.95 + 101.06
  = 261.00
- Calculating the WCV value is by squaring the distance calculation results and adding up all the exponents of the distance calculation results.
  WCV = 60.07 + 73.47 + 49.55 + 249.83 + 87.55 +……+ 0
  = 11141.96
- Calculate the ratio value by dividing the BCV value by WCV

BCV/WCV = 261/11141.96
= 0.023425

Then the next step to enter iteration 3 is to create a new centroid center. Like the table below :

Table 6. New Centroid

CENTROID	K1	K2	K3	K4	K5	K6	K7	K8	K9
C1	91	91	90	91	91	91	91	91	91
C2	82.26	84.00	82.26	83.71	85.83	85.83	82.51	85.57	82.37
C3	80	81.8	75	80	83	68.6	78.6	68.6	73
C4	68.38	67.00	67.13	74.38	74.63	69.50	66.75	70.88	67.00
C5	50	62	50	67	60	50	68	50	68

Then recalculate the ratio value by comparing the BCV (*Between Cluster Variation*) and WCV (*Within Cluster Variation) values* . The following is the calculation process for BCV (*Between Cluster Variation*) and WCV (*Within Cluster Variation)* .

- Calculate the BCV value with the formula adding up the results of the distance between each centroid.
  BCV = 10.16 + 22.44 + 47.44 + 78.18 + 100.34
  = 258.56
- Calculating the WCV value is by squaring the distance calculation results and adding up all the exponents of the distance calculation results.
  WCV = 66.29 + 71.08 + 51.93 + 258.72 + 103.29 +……+ 0
  = 11010.11

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13890 p-ISSN : 2541-044X

- Calculate the ratio value by dividing the BCV value by WCV
  BCV/WCV = 258.56/11010.11
  = 0.023484

The ratio value increases from 0.023425 to 0.023484 and *the cluster* does not change from literacy 3 then *clustering* is complete. The following is a table for grouping the final results of the evaluation of the performance of educational staff as follows.

Table 7. Grouping of Final Results of Performance Assessment Evaluation

Clusters	**Number**	**Number of *Clusters***	**Performance Achievement of Education Personnel**
C1	{P11,P14,P15,P20,P23,P24,P25,P29,P31,P32,P34,P35,P37,P44,P48,P49,P53,P54,P57,P58,P59,P60,P61,P63,P66 ,P7,P47,P55}	28	Very Good Education Staff
C2	{P1,P2,P4,P5,P9,P10,P12,P13,P16,P17,P21,P22,P26,P27,P36,P38 ,P39,P40,P41,P42,P43,P45,P51,P56,P62 ,P64,P65,P67,P68,P69,P71,P72,P73,P75,P77}	35	Good Education Staff
C3	{ P46, P6, P8, P28, P33 }	5	Educational Staff Are Quite Good
C4	{P18,P19,P30,P50,P52,P70,P74,P76}	8	Educational Staff Are Not Good
C5	{P3}	1	Bad Education Staff

Then, after the final results of the performance assessment evaluation have been grouped, the *K-Means* algorithm is based on calculation results with an accuracy of 100%, which means that all model predictions match the actual performance labels without any errors. Precision for each category was also 100%, indicating that each performance category was predicted correctly without any false positives. In addition, recall for all categories reached 100%, indicating that the model was able to identify all educational staff in each appropriate performance category. These results show that the *K-Means algorithm*, in this case, succeeded in grouping educational staff into their performance categories very accurately and effectively. This evaluation proves that the K-Means model used has provided optimal results in predicting the performance of educational staff with no errors in classification.

Naïve Bayes Algorithm Calculation Process Results

The stages carried out in the *Naïve Bayes algorithm calculation process* for evaluating personnel performance based on the internal quality assurance system are as follows.

Table 8 . Aspects of Performance Evaluation Criteria

CODE	**CRITERIA**
K1	Personal Work Results
K2	Job Knowledge
K3	Initiative
K4	Mental Dexterity
K5	Attitude
K6	Work Discipline
K7	Teamwork Results
K8	Instruction Compliance
K9	Loyalty

Table 9. Educational Performance Assessment Categories

NO	**RANGE OF VALUES**	**ACHIEVEMENTS OF EDUCATIONAL PERFORMANCE PERFORMANCE**
1.	91-100	Very Good Education Staff
2.	81-90	Good Education Staff
3.	71-80	Educational Staff Are Quite Good
4.	61-70	Educational Staff Are Not Good

*name of corresponding author

5.	0-60	Bad Education Staff

Table 10. Evaluation Data for Education Personnel Performance Assessment

Education Personnel Data	K1	K2	K3	K4	K5	K6	K7	K8	K9	Total	Performance Achievement of Education Personnel
P1	85	75	80	85	80	80	82	80	91	82	Good Education Staff
P2	88	89	85	85	85	75	85	76	75	83	Good Education Staff
P3	50	62	50	67	60	50	68	50	68	58	Bad Education Staff
P4	80	85	80	80	83	86	82	80	86	82	Good Education Staff
P5	88	89	85	85	90	87	85	89	85	87	Good Education Staff
P6	75	77	65	80	80	85	80	65	81	76	Educational Staff Are Quite Good
P7	89	86	85	90	90	90	89	90	88	89	Good Education Staff
P8	80	85	70	80	85	65	75	65	70	75	Educational Staff Are Quite Good
…	…	…	…	…	…	…	…	…	…	…	…
P77	85	86	90	90	90	86	80	88	84	87	Good Education Staff

The following are the steps for completing the *Naïve Bayes algorithm calculation process* for evaluating energy performance based on an internal quality assurance system where these calculations are taken from random data samples, namely:

1. Calculate the number of classes for each label
   Calculating training data against test data taken from training data
   - Calculating the Probability of the Class " Very Good Performance of Educational Personnel "
     = Total data on Educational Personnel with Very Good Performance / Total Data
     = 28 / 76
     = 0.3684
   - Calculating the Probability of the Class "Educational Personnel Performing Well"
     = Total data on Good Performance Education Personnel / Amount of All Data
     = 34 / 76
     = 0.4474
   - Calculating the Probability of the Class " Educational Personnel Performing Quite Well"
     = Total data on educational staff performing quite well / Amount of all data
     = 6 / 76
     = 0.0789
   - Calculating the Probability of the Class "Educational Personnel Performing Poorly"
     = Total data on educational staff with poor performance / Amount of all data
     = 8 / 76
     = 0.1053
   - Calculating the Probability of the Class "Poor Performance of Educational Personnel"
     = Total data on educational staff with poor performance / Amount of all data
     = 0 / 76
     = 0

2. Calculate the mean value of each class for each criterion
   Calculating training data against test data taken from training data
   - Calculating the mean value of the "Very Good Performance of Educational Personnel" class against the "Personal Work Results" criteria
     = Amount of Performance Evaluation Data for Education Personnel on Work Results Criteria Personal with the performance achievement of "Very Good Performance of Educational Personnel" / Total Average Data for Educational Staff Performance is Very Good
     = 2532 / 28
     = 90.43

Table 11. Results of Mean Values for Test Data Taken from Training Data

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13890
e-ISSN : 2541-2019
p-ISSN : 2541-044X

MEAN VALUE									
CLASS	K1	K2	K3	K4	K5	K6	K7	K8	K9
Very Good Education Staff	90.43	91.25	90.07	91.14	90.54	91.21	89.93	91.25	90.14
Good Education Staff	82.62	84.21	82.32	83.82	85.85	85.21	82.88	84.88	82.12
Educational Staff Are Quite Good	72.50	74.67	70.83	77.00	80.83	77.50	73.83	74.67	72.67
Educational Staff Are Not Good	67.75	68.25	66.50	73.75	72.75	63.25	68.38	66.50	68.00
Bad Education Staff	0	0	0	0	0	0	0	0	0

3. Calculate the standard deviation value of each class based on each criterion
   a. Calculating training data against test data taken from training data
      - Calculate the standard deviation value of the "Very Good Performance of Educational Personnel" class against the "Personal Work Results" criteria
      -

$$=\frac{(88-90.43)^2+(90-90.43)^2+(90-90.43)^2+\ldots\ldots\ldots\ldots+(90-90.43)^2}{28-1}$$

= 1.81

Table 12. Results of Standard Deviation Values

STANDARD DEVIATION VALUE									
CLASS	K1	K2	K3	K4	K5	K6	K7	K8	K9
Very Good Performance of Educational Staff	1.81	2.50	2.89	1.96	3.01	2.36	3.95	2.69	4.76
Good Performance Education Staff	5.69	3.70	5.36	2.88	3.55	6.07	2.85	5.71	5.60
Educational Staff Performance is Quite Good	5.24	7.63	3.76	4.69	2.04	9.87	5.85	8.16	7.79
Educational Staff Performance is Poor	8.68	5.70	10.4	6.45	8.65	5.52	6.39	7.11	2.07
Educational Staff Poor performance	0	0	0	0	0	0	0	0	0

4. Calculating Gaussian Distribution Values from Test Data
   Calculating training data against test data taken from training data
   - Calculating the Gaussian distribution value of the " Very Good Performance Education Personnel" class against the "Personal Work Results" criteria

$$=\frac{1}{\sqrt{2*3{,}14*1.81}}-exp^{\frac{(92-90.43)^2}{2*1.81^2}}$$

= 0.25

Table 13. Results of Gaussian distribution values

GAUSSIAN DISTRIBUTION VALUE										
CLASS	K1	K2	K3	K4	K5	K6	K7	K8	K9	FINAL
Training Test Data	92	91	90	90	90	91	92	90	90	
Very Good Performance of Educational Staff	0.25	0.25	0.23	0.26	0.23	0.26	0.19	0.23	0.18	2.08
Good Performance Education Staff	0.08	0.09	0.10	0.07	0.15	0.13	0.02	0.14	0.10	0.89
Educational Staff Performance is Quite Good	0.01	0.05	0.00	0.03	0.00	0.08	0.01	0.06	0.04	0.27
Educational Staff Performance is Poor	0.02	0.00	0.03	0.03	0.05	0.00	0.01	0.01	0.00	0.15
Educational Staff Poor performance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

naïve Bayes calculations, there is a confusion matrix analysis of the classification model used to assess the performance of educational staff, an overall accuracy result of 91% was obtained. This model shows excellent performance in identifying educational personnel with good and fairly good performance, with 100% precision in the categories "Educational Personnel with Good Performance" and "Educational Personnel with Fair Performance", as well as 100% recall for the category "Educational Personnel with Fair Performance Good" and "Educational Personnel Performing Poorly".

DISCUSSIONS

In this study, there are comparison results of the confusion matrix results which have differences in the two methods, namely as follows:

Table 14. Analysis of *K-Mean* and *Naïve Bayes* Algorithms Comparison of Confusion Matrix

COMPARATIVE ANALYSIS	Confusion Matrix				
	Very Good Education Staff	Good Education Staff	Educational Staff Are Quite Good	Educational Staff Are Not Good	Educational Staff Poor performance
K-Means algorithm	28	35	5	8	1
Accuracy Value	100%				
Recal Value	100%	100%	100%	100%	100%
Precision Value	100%	100%	100%	100%	100%
Naïve Bayes algorithm	29	34	6	8	0
Accuracy Value	91%				
Recal Value	100%	85%	100%	100%	0%
Precision Value	79%	100%	100%	87%	0%

CONCLUSION

This research compares two algorithms , *K-Means* and *Naive Bayes*, to evaluate the performance of educational staff based on SPMI Standards at STMIK Triguna Dharma. The results show that K-Means has 100% accuracy in grouping the performance of educational staff, while Naive Bayes achieves 91% accuracy. This evaluation shows the high effectiveness of both algorithms in assessing performance, with K-Means showing slightly superior results.

REFERENCES

Adi, MFH, & Susetyo, YA (2020). Mapping Potential Disasters in Central Java Using Google Maps API and KML with the K-Means Method. *Indonesian Journal of Computing and Modeling* , *3* (1), 43–51.

Al-Muhtadi, D., & Sumiati. (2023). The Influence of Workload and Job Satisfaction on Employee Performance Using the Structural Equation Modeling (SEM) Method at PT. Indonesia Tri Sembilan. *Journal of Engineering Science Research (JUPRIT)* , *2* (3), 1–21. https://doi.org/10.55606/juprit.v2i2.1977

Andini, F., Zilfitri, D., Filki, Y., & Ridho, M. (2022). K-Means Clustering Algorithm in Optimizing the Composition of Laying Chicken Animal Feed. *Journal of Information Systems and Technology* , *5* , 44–48. https://doi.org/10.37034/jsisfotek.v5i2.168

Arifudin, O. (2019). Management of the Internal Quality Assurance System (SPMI) as an effort to improve the quality of higher education. *MEA Scientific Journal* , *3* (1), 161–169. https://doi.org/10.31955/mea.vol3.iss1.pp161-169

Dr. Suyanto. (2019). *Data Mining for Classification and Clustering of Title Data* (Dr. Suyanto (ed.); First). Bandung Informatics Publisher.

Duei Putri, D., Nama, GF, & Sulistiono, WE (2022). Sentiment Analysis of the Performance of the House of Representatives (DPR) on Twitter Using the Naive Bayes Classifier Method. *Journal of Informatics and Applied Electrical Engineering* , *10* (1), 34–40. https://doi.org/10.23960/jitet.v10i1.2262

Fadhli, M. (2020). Internal and External Quality Assurance Systems in Higher Education Institutions. *AL-TANZIM: Journal of Islamic Education Management* , *4* (2), 53–65. https://doi.org/10.33650/al-tanzim.v4i2.1148

Harahap, ASH (2019). Analysis of strategies for improving the performance of the secretariat at the Padang City Education, Youth and Sports Department. *Journal of Applied Management Science* , *1* (1), 69–75. https://doi.org/10.31933/JIMT

Muhammad Arhami, MN (2020). *Data Mining Algorithms and Implementation* (Ratih Indah Utami (ed.)). CV ANDI OFFSET.

Ningsi, LN, Poningsih, P., & Tambunan, HS (2021). Implementation of Cluster Data Mining in Households that Have Access to Adequate Housing Based on Province. *Scientific Studies in Informatics and Computers* , *1* (5), 228–

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

234. http://djournals.com/klik/article/view/182

Risqi Ananda, M., Sandra, N., Fadhila, E., Rahma, A., & Nurbaiti, N. (2023). Data Mining in the PT Indofood Lubuk Pakam Company. *Comit: Communication, Information and Technology Journal* , *2* (1), 108–119. https://doi.org/10.47467/comit.v2i1.124

Saiful Anwar Matondang. (2023). *Technical Guide for Filling Out the SPMI Kemendikbudristek Application* . LLDikti Region I North Sumatra.

Samudin, S. (2023). *The urgency of accelerating higher education quality assurance through PT accreditation* . *March* . https://doi.org/10.13140/RG.2.2.24170.70087

Saputra, D., Sucipto, A., Al Kahfi Masputra, M., & Rizky ASiregar, H. (2022). Implementation of the BARS (Behaviorally Anchored Rating Scale) Method in the Employee Performance Appraisal System (Case Study: PT. Laksana Aneka Sarana). *Telefortech* , *3* (1), 2774–5384.

Setianingsih, I., & Ali, I. (2023). Clustering of Cash Expenditures in Pamengkang Village Using the K-Means Method. *JATI (Informatics Engineering Student Journal)* , *7* (1), 784–787. https://doi.org/10.36040/jati.v7i1.6488

Sharif, A. (2019). Data Mining to Predict Goods Sales Promotion Itemsets Using the Market Basket Analysis (MBA) Method (Case Study: Cell Phone Center Store). *Https://E-Jurnal.Pelitanusantara.Ac.Id/Index.Php/Mantik/Article/View/678* , *3* (2), 117–123.

Sidik, R., Suarna, N., & Rinaldi Dikananda, A. (2023). Data Analysis of Student Specialization Sets Using the K-Means Algorithm to Optimize Parameters in Vocational High Schools. *JATI (Informatics Engineering Student Journal)* , *7* (2), 1197–1203. https://doi.org/10.36040/jati.v7i2.6335

So seeni, TI, Hidayanti, N., & Kania, R. (2022). Implementation of Data Mining to Evaluate Research Performance of Banten Jaya University Students Using the Naive Bayes Classifier Algorithm. *THEOREMS (The Original Research of Mathematics) Journal* , *6* (2), 135. https://doi.org/10.31949/th.v6i2.3430

Supriyadi, A. (2023). Comparison of Naive Bayes and Decision Tree (C4.5) Algorithms in Classifying Outstanding Lecturers. *Generations Journal* , *7* (1), 39–49. https://doi.org/10.29407/gj.v7i1.19797

*name of corresponding author