Fraktal: Jurnal Matematika dan Pendidikan Matematika
Volume 5, No. 2, November 2024, Hal. 65-74 (e-ISSN 2776-0073)
Available online at https://ejurnal.undana.ac.id/fraktal

SISTEM VISUALISASI DATA SERTIFIKASI KARYAWAN BERBASIS WEBSITE MENGGUNAKAN METODE NAIVE BAYES

Djeliyana Talaba[1], I Wayan Sudarsana[2], Iman Al Fajri[3]

[1,2,3] Program Studi Matematika Jurusan Matematika FMIPA Universitas Tadulako
Email: djeliyanat@gmail.com[1], sudarsanaiwayan@yahoo.co.id[2], iman.alfajri@gmail.com[3]

Abstrak

Sertifikasi karyawan dalam sebuah perusahaan merupakan salah satu aspek penting dalam dunia kerja yang akan menunjukkan kompetensi dan kualitas seorang karyawan. Namun, untuk mendapatkan hasil yang sesuai, perusahaan menggunakan sistem pendukung dari analis data. Penelitian ini bertujuan untuk membuat sistem visualisasi data berbasis website yang dapat menyajikan informasi sertifikasi karyawan secara jelas, mendapatkan hasil analisis data karyawan untuk pengambilan keputusan sertifikasi dengan menggunakan metode Naive Bayes. Penelitian ini dilakukan dalam 2 tahap, yaitu tahap pelatihan dan pengujian dengan menggunakan 5 atribut. Pengolahan data dilakukan dengan menggunakan 59 data training dan 59 data testing. Penelitian ini menghasilkan sebuah sistem yang dapat membantu dalam pengambilan keputusan kelayakan karyawan yang ingin mengikuti sertifikasi menggunakan metode naive bayes dan memvisualisasikannya, hasil yang didapat dari sistem tersebut adalah 10 data yang tidak layak mengikuti sertifikasi dengan tingkat keakuratan sebesar 96%.
Kata kunci: Algoritma Naive Bayes, Sertifikasi, Website

Abstract

Employee certification in a company is one of the important aspects in the world of work that will show the competence and quality of an employee. However, to get the appropriate results, the company uses a support system from data analysts. This research aims to create a website-based data visualization system that can present employee certification information clearly, get the results of employee data analysis for certification decision making using the Naive Bayes method. This research was conducted in 2 stages, namely the training and testing stages using 5 attributes. Data processing was carried out using 59 training data and 59 testing data. This research produces a system that can assist in making decisions on the eligibility of employees who want to take certification using the naive bayes method and visualize it, the results obtained from the system are 10 data that are not eligible for certification with an accuracy of 96%.
Keywords: Naive Bayes Algorithm, Certification, Website

PENDAHULUAN

Employee certification is one of the important aspects in the world of work that shows a person's competence and quality in a certain field In a company, information about employee certification can be an important reference in making decisions related to project assignments, promotions, or career development

PT United Tractors Tbk (UT) is a subsidiary of PT Astra Internasional Tbk (Astra) which operates in the fields of construction machinery, mining contractors, coal mining, gold mining, energy, and construction industries in Indonesia (TRACTORS, 2024) Because it operates in the construction sector, this company requires competency certification to show the quality of its employees before they work in the field As one of the large companies, PT United Tractors Tbk (UT) has a significant number of employees This encourages companies to require data visualisation

related to the certification

In the context of employee certification data, there are still not many studies that combine data visualisation with the right analysis methods In this study, the researcher wanted to create a website that can provide certification recommendations that are in accordance with employee profiles, facilitate data analysis, and make it easier for management to make decisions related to employee eligibility and can visualise it using the naïve Bayes method

Based on the explanation above, the author is interested in conducting research using the naive bayes method with the Decision Support System (DSS) system because there is no system that focuses on employee certification The use of this method is expected to help better decision-making in terms of project assignments, promotions, or career development of employees In addition, the website contains the first feature that can test the feasibility of employee data that wants to take part in certification, the second feature that can appear a certification recommendation that is in accordance with the employee's division and the third feature can visualize employee certification data With this system, it is hoped that it can help manage employee certification data more efficiently and effectively.

METHOD

The location of data collection was carried out at the office of PT United Tractors Tbk (UT) which is located in Cakung, East Jakarta And the place of research was carried out at the ROPD Laboratory of Tadulako University in Tondo, East Palu. The tools and materials used in this study include Lenovo Laptop, Microsoft Excel, Xampp, Php My Admin, Visual Studio Code, MySql, and boostrap. The data source used in this study is in the form of secondary data obtained from PT United Tractors Tbk (UT) located in Cakung, East Jakarta in the form of a database Where the data is employee data who want to take part in certification.

The procedure in this research are

a. Reviewing the literature related to problems in decision-making on the eligibility of employees who want to take part in certification and problems in delivering employee certification information.
b. Formulate problems about decision-making on the eligibility of employees who want to take part in certification and problems in the delivery of employee certification information, as well as apply mathematics as a solution based on literature studies related to previous research
c. Collect data on employees who want to take part in certification consisting of 5 attributes, namely name, nrp, division, position and length of service.
d. Data is processed using a method, namely the Naive Bayes method to classify employee eligibility
e. Getting the results of the research and concluding it.

RESULTS AND DISCUSSION

The result of this study is a website-based employee eligibility decision support system and visualization of the data The classification process is carried out using the Gaussian Naive Bayes Algorithm.

Data Collection

The data used is data taken directly from PT. United Tractors. The data taken is in the form of data on employees who want to take part in certification from January to December 2023. In the data processing process, the data will be divided into 2 parts, namely training data and data testing in order to produce the most optimal performance, so data is divided with a proportion of 50% data for training and 50% data for testing using the Niaive Bayes method (Pahlevi et al., 2023).

Flowchart Naive Bayes

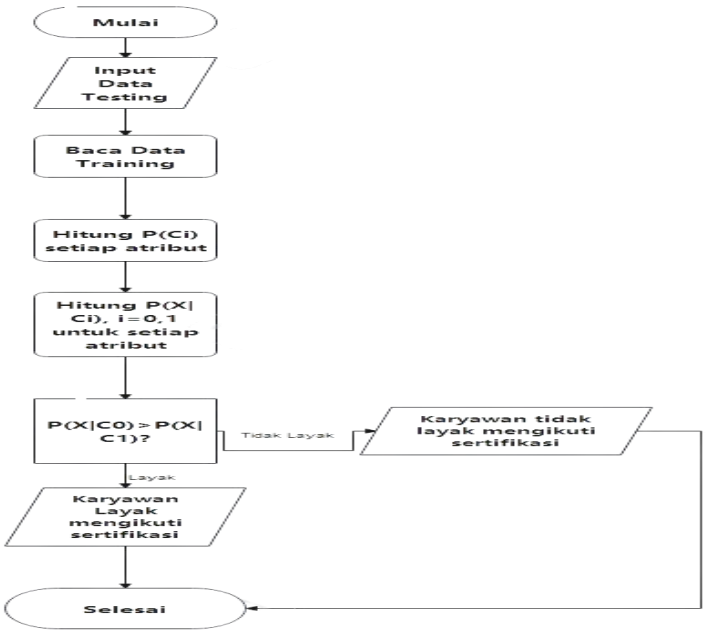


Figure 1. *Flowchart* Naive bayes

Figure 1 illustrates the flowchart of the Naive Bayes method The process starts with data input, followed by a reading of the attribute data Next, calculate P(C_i) for each attribute in the case of the test data, which includes the attributes of division, job title, and tenure with two classes, ie worthy and not worthy Then, calculate $P(X \mid C\_i)$ for each class or attribute, where i = feasible or unfeasible After that, compare the results; if $P(X \mid C\_0) > P(X \mid C\_1)$, then the conclusion is that C_1 is a class with a division category, position, and service period = feasible.

Data Processing Using the Naive Bayes Method

Data Training

This training data is used to build and train the data mining system that has been created, where the expected target value has been set in this data training. The amount of employee data used is 59 data with 5 attributes for the training process itself. Of the 59 data used, it is known that the number of

C_0 classes (Not Feasible) is 10 data and the number of C_(1) (Feasible) classes is 49 data. Then what is calculated is the number of frequencies of the class that appears divided by the total dataset:

P(C_0)= 10/59=0.1695 Unworthy class P(C_1)=49/59=0.8305 decent class

From the calculation above, the probability of the unworthy class is 0.1695 and the decent class is 0.8305. After calculating the class probability, the next step is to determine the conditional probability value for each class P(X | Ci), where i = 0.1, and for each attribute of the input data sample. The following two attributes, namely division and position, are included in the categorical type probability, while the other attribute, namely the length of service, is included in the numerical type probability.

To calculate the probability of each attribute can be done in the following way.

$P(Chcu|C_0) = \frac{0}{10} = 0{,}000$ $P(AST|C_0) = \frac{3}{10} = 0{,}300$

$\mathrm{P(Chcu|C_1)} = \frac{7}{49} = 0{,}143$ $P(AST|C_1) = \frac{10}{49} = 0{,}204$

The results of the calculation of the probability of division and position attributes are presented in table 1 and table 2.

Table 1. Division Attribute Probability

Divided	Class		Probability	
	Not Eligible	Proper	Not Eligible	Proper
Chcu	0	7	0,000	0,143
Cau	0	3	0,000	0,061
Cfa	0	9	0,000	0,184
⋮	⋮	⋮	⋮	⋮
Cgs	6	15	0,600	0,306

Table 2. Probability of Position Attributes

Position	Class		Probability	
	Proper	Not Eligble	Proper	Not Eligble
AST	10	3	0,204	0,300
DH	3	0	0,061	0,000
⋮	⋮	⋮	⋮	⋮
TM	21	4	0,429	0,400

in the probability table for division and position attributes, it can be seen that there are many 0 values, this is an anomaly because in the Naive Bayes method, the value of 0 indicates probability does not occur so there should be no 0 value, so it is overcome with the Laplacian Correction technique, to ensure that the probability value is not 0 Therefore, in correcting the class with a value of 0 in table 1 and table 2 above, calculations are carried out using the Laplacian Correction technique in accordance with the following equation formula:

$$P(c,t) = \frac{1 + N(ct,D)}{N(c,D) + |V|}$$

Table 3. Results of Division Attribute Calculation using laplacian correction technique

Attribute	P (Not Eligible)	P(Proper)
CHCU	0,053	0,138
CAU	0,053	0,069
⋮	⋮	⋮
CGS	0,368	0,276

Table 4. The results of the calculation of Position Attributes using the laplacian correction technique

Atribute	P(Proper)	P(Not Eligible)
AST	0,2	0,25
DH	0,073	0,063
⋮	⋮	⋮
TM	0,400	0,313

As explained earlier, the attribute of service life is a numerical type of probability Where the probability with this type is only looking for the mean value and standard deviation Therefore, the resulting mean and standard deviation values can be seen in the following tables 5 and 6:

Table 5. Mean Value

Class	Working Period
Proper	13,04
Not Eligible	2,90

Table 6. Value of Deviation

Class	Working Period
Propor	6,34
Not Eligible	4,98

Testing data is data used for testing on mining data In this test data, the output or target results are predetermined The amount of data used is 59 data with 5 attributes Table 7 is an example of data that becomes testing data

Table 7. Example of Data Testing (test)

Employee Name	NRP	Divided	Position	Working Time	Class	Result
FSB	0	CHCU	AST	1	Not Eligible	?

As explained in the data training process, the calculation of division and position attributes is different from the attributes of service period For the calculation process of testing data on division and position attributes, the results of training data are used, whilst for the calculation process of testing data for service period attributes using the Gaussian naive bayes formula (Hasanah et al, 2022) by taking mean values and standard deviations from the previous data training calculations.

Calculating the likelihood of including an Unworthy class (C_0)

$$P(C_0) = \frac{10}{59} = 0{,}1695 \text{ Class Not Worthy}$$

$P(Division = \ Chcu \mid C_0)$ P (Position $=$ AST $\mid C_0$)

$= \frac{1+0}{10+9} = \frac{1}{19}$ $= \frac{1+3}{10+6} = \frac{4}{16}$

$= 0{,}053$ $= 0{,}25$

To find the probability of the working period using the Gaussian formula.

$$P(X_i) = x_i|Y = y_j = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{\frac{(x_k-\mu_{ik})^2}{2\sigma_{ij}^2}}$$

$$P\ (Working\ Time = 1|C_0) = \frac{1}{\sqrt{2(3{,}14)(4{,}98)}} e^{\frac{(7-2{,}90)^2}{2(4{,}98)^2}} = 0{,}166303662$$

The next step is to calculate likelihood with the equation:

P (X |Credentials $=$ C_0)

$= P(Division = Chcu|Credentials = C_0)$

$\times P(Position = AST\ |Credentials\ C_0) \times P(Working\ Time$

$= 1|C_0) \times P(C_0)$

$= 0{,}053 \times 0{,}25 \times (1{,}66304E-01) \times 0{,}1695 = 0{,}000370882$

Calculating the likelihood of including an Eligible class (C_1)

$P(C_1) \quad = \frac{49}{59} = 0{,}8305$ Decent Class

P(Division = Chcu | C_1) P (Jabatan = AST | C_1)

$= \frac{1+7}{49+9} = \frac{8}{58}$ $= \frac{1+10}{49+6} = \frac{11}{55}$

= 0,138 = 0,2

To find the probability of the working period using the Gaussian formula

$$P(X_i) = x_i|Y = y_j = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{\frac{(x_k-\mu_{ik})^2}{2\sigma_{ij}^2}}$$

$$P\ (Working\ Time = 1|\ Proper) = \frac{1}{\sqrt{2(3{,}14)(6{,}34)}} e^{\frac{(7-13{,}04)^2}{2(6{,}34)^2}}$$

$$= 0{,}0261596442$$

The next step is to calculate likelihood with the equation:

P (X |Credentials $=$ C_1)

$= P(Division = Chcu|Credentials = C_1)$

$\times P(position = Asspciate\ |Credentials\ C_1)$

$\times P(Working\ Time = 1|C_1) \times P(C_1)$

$= 0{,}138 \times 0{,}2 \times\ 0{,}026159642 \times 0{,}8305$ = 0,00059933

From the results of the calculation of testing data (test) with the name FSB got a greater value for the Decent class 〖(C〗_1) with a value of 000059933, it can be concluded that the employee named FSB is included in the Decent class category.

Database *Design*

In designing a database for a website portal system, several user data tables, personal data tables, and other tables are needed, namely:

- User Table
- Atribut Table
- Table Data Hitung
- Table Data Latih
- Table Data Peserta
- Table Data Sertifikasi
- Table Data Uji
- Table Gausian
- Table Hasil Hitung
- Table Hitung Uji

From the results of these tables, a web-based system will be built to test the feasibility of employees using the Naive Bayes method The results of these classifications and predictions will be displayed in visual form on the website Here is the Use Case of the built system.

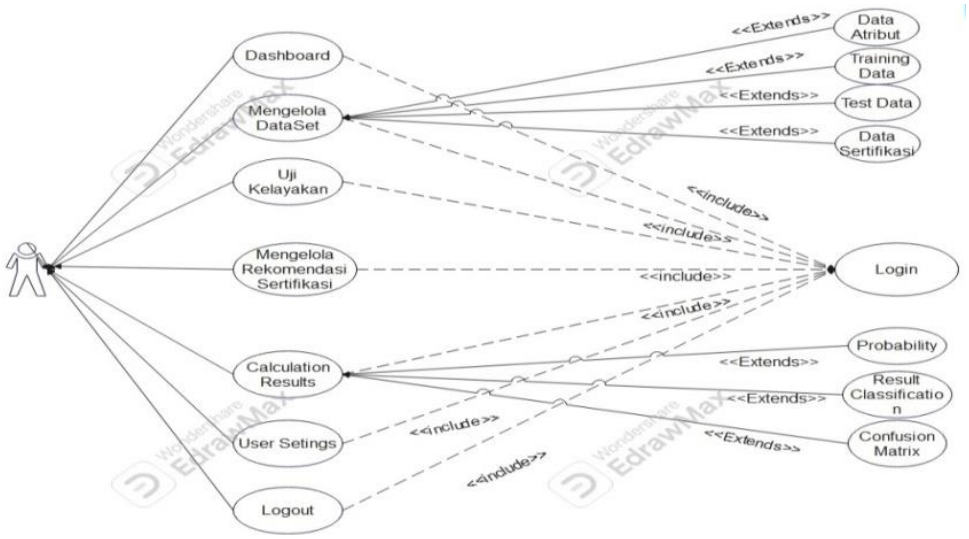


Figure 2. ***Use Case System***

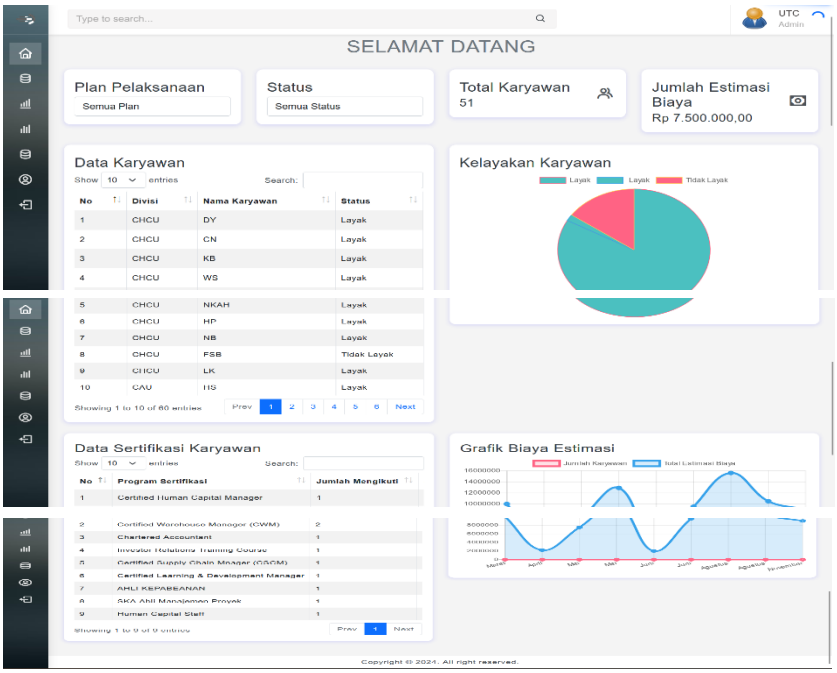
The training data that has been processed is then evaluated to determine whether the results are in line with expectations The following table 8 shows a comparison between the original data results and the results from the system.

Table 8. Original Data Results and Systems

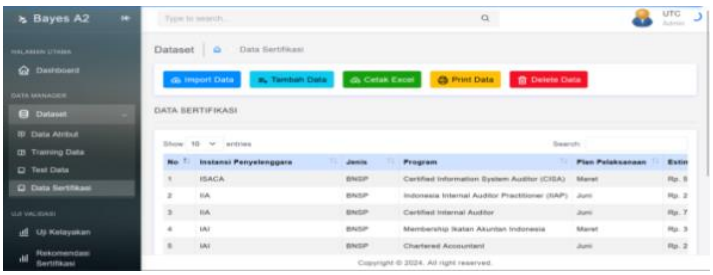
No	Name	NRP	Division	Position	Working Period	Class	Result	Description
1	DY	0	CHCU	AST	7	Proper	Proper	Eligible
2	CN	0	CHCU	AST	6	Proper	Proper	Eligible
3	KB	0	CHCU	AST	5	Proper	Proper	Eligible
4	WS	0	CHCU	AST	19	Proper	Proper	Eligible
5	NKAH	0	CHCU	AST	6	Proper	Proper	Eligible

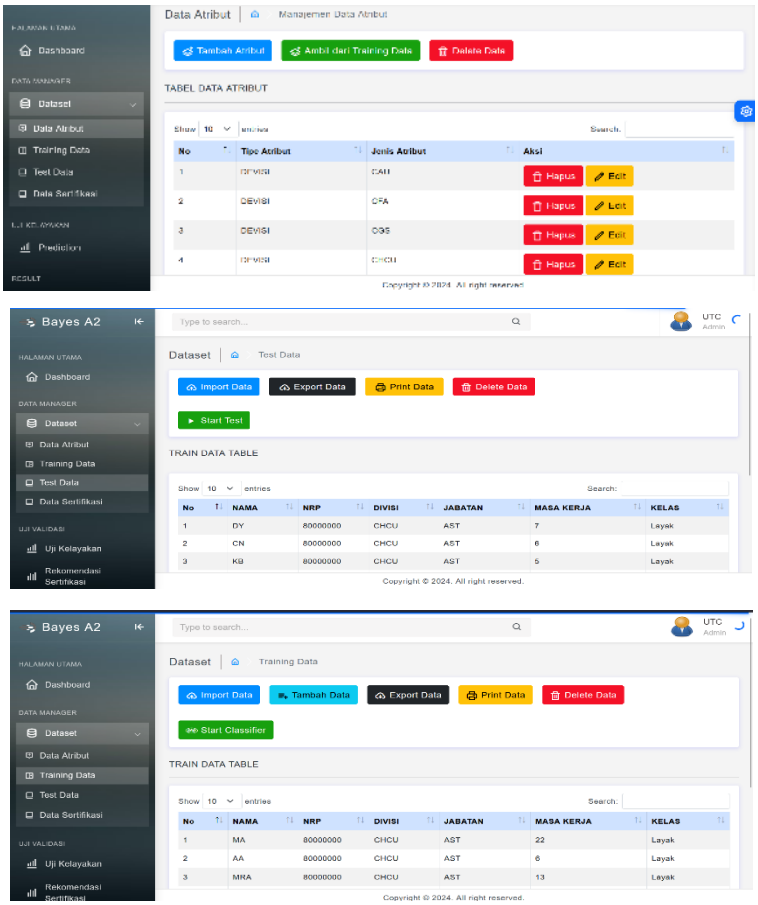
No	Name	NRP	Division	Position	Working Period	Class	Result	Description
6	HP	0	CHCU	AST	13	Proper	Proper	Eligible
7	NB	0	CHCU	AST	14	Proper	Proper	Eligible
8	FSB	0	CHCU	AST	1	Proper	Proper	Not Eligible
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
59	APO	0	CGS	SPS	14	Proper	Proper	Eligible

After the results of the original data test, an evaluation of the system test is carried out where this section will discuss the evaluation of the system testing that has been developed. Where from the results of the calculation and testing of the system, 49 decent class data were obtained that were predicted to be correct, 2 decent class data that were predicted to be unfeasible, 8 decent class data that were predicted to be not worthy of being predicted correctly, and 1 unworthy class data that was predicted to be feasible. With an accuracy value of 96%. The results of this study show that this prediction system is very accurate in predicting the feasibility of employee certification. The results of the calculation will be visualized in the website-based system that has been created.

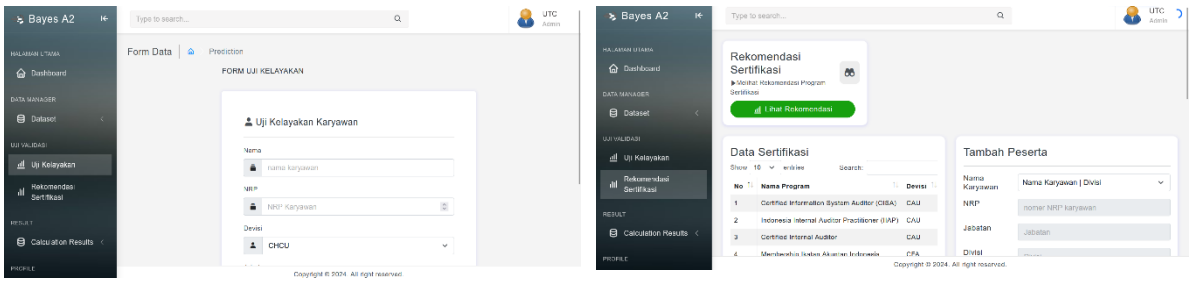


The display above shows the main feature of the website, namely the dashboard. This dashboard feature displays a visualization of the results of employee due diligence using the naïve bayes method as well as overall employee certification data.

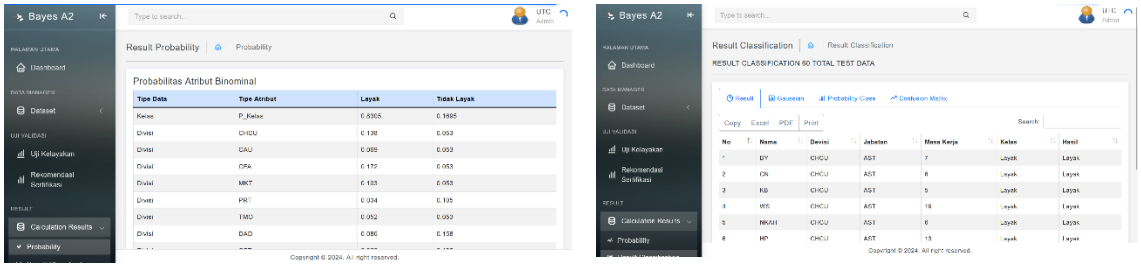




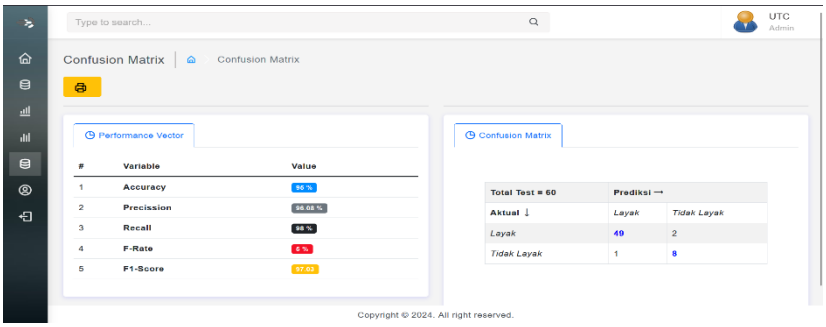
The displasy above is a dataset feature in the dataset feature there are attribute data features, training data, test data, certification data funds.



The display above is the feasibility test and certification recommendation feature in the dataset feature there are attribute data features, training data, test data, certification data funds.



The above display is the display of the probability and Result Classification features located on the calculation results page. This feature will Display the probability of the attributes we entered and the results of the classification that we have done using the Naive Bayes method.



The display above is the display of the Confusion Matrix feature located on the calculation results page. In this feature, the accuracy level of the calculation results that we have done using the naïve bayes method and also the amount of data whose prediction results are correct and incorrect. Where the accuracy level obtained is 96%.

CONCLUTION

Based on the results of the research that has been conducted, it can be concluded that this study shows that a web-based employee certification data visualization system using the naïve bayes method can provide accurate prediction results and help companies manage the employee certification process more efficiently. This system can be integrated with a human resource management system (HRMS) to provide automatic and real-time prediction of certification eligibility.

REFERENCES

Hasanah, Q., Oktavianto, H., & Rahayu, Y. D. (2022). Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung Gaussian Naive Bayes Algorithm Analysis Of Data Classification Of Heart Failure Patiens. Journal of Smart Technology, 3(4), 2774–1702. http://jurnal.unmuhjember.ac.id/index.php/JST

Pahlevi, R., Negara, E. S., Sutabri, T., & Herdiansyah, M. I. (2023). Application of the Naive Bayes Method to Determine the Classification of Eligibility for Receiving Rehabilitation and School Construction Assistance at the Banyuasin Regency Education and Culture Office. Journal of Information and Computer Technology, 9(2), 1176–1188. https://doi.org/10.37012/jtik.v9i2.1790

Tractors, U. (2024). United Tractors. https://www.unitedtractors.com/