

Prediction of the Number of Patient Visits Using Naive Bayes Algorithm

Zidhane Syahputra^{1)*}, Rakhmat Kurniawan²⁾

^{1,2)}Computer Science, State Islamic University of North Sumatra, Indonesia

¹⁾zidhansyahputra@gmail.com, ²⁾rakhmat.kr@uinsu.ac.id

ABSTRACT

This study was conducted to predict the number of patient visits at Prof. Dr. M. Ildrem Mental Hospital using the Naive Bayes algorithm, which is relevant given the increasing need for global mental health care. The main problem of this study is the difficulty in managing hospital resources efficiently due to unpredictable fluctuations in the number of patient visits. The research aims to apply the Naive Bayes algorithm to predict the number of patient visits and evaluate their performance. The method used is a naïve Bayes algorithm with systematic steps including historical data collection, data preprocessing using LabelEncoder, and dividing the dataset into training data and test data (80:20) where the training data totals 1331 data and the test data has 333 data. The Naive Bayes model is built and tested with metrics such as accuracy, precision, recall, and F1-score. The results of the study based on confusion matrix analysis, the model achieved an accuracy of 0.8108108108109 or 81%, a precision of 0.8206686930091185 or 82.07%, a recall value of 0.9926470588235294 or 99.26%, and an F1-score of 0.90 or 90%, which shows that this model is quite effective in predicting service units with the dominance of adolescent category patient data where it is concluded that this prediction model is able to provide accurate estimates of patient visits, supporting the management of hospital resources, and improving the operational efficiency of mental health services. This research is expected to help hospitals in planning facilities and workforce more effectively.

Keywords: Naive Bayes Algorithm, Mental Hospital, Data Mining, Patient Visit Prediction, Machine Learning

INTRODUCTION

WHO data in 2023 shows that more than 20% of the world's population has experienced mental disorders throughout their lives. This shows that mental and emotional disorders are one of the growing global health problems. Therefore, the need for advanced data-driven prediction technology is urgent to help mental hospitals plan services and manage resources effectively.

Mental hospitals are one of the health services provided by the government to provide medical care for individuals who experience serious mental and emotional disorders. In contrast to general hospitals, psychiatric hospitals have a very specific focus on the diagnosis, treatment, and rehabilitation of patients with a variety of mental conditions, ranging from depressive and anxiety disorders to psychiatric disorders (Ambarita 2020). Thus, mental hospitals are not only a place for physical treatment, but also a place of healing and recovery for those struggling with mental health challenges.

In the process of carrying out their roles, the hospital faces a number of unique challenges that affect their operations and services. The demand for mental health services can continue to increase over time, which can put additional pressure on mental hospitals to provide adequate and timely services to all patients. Mental hospitals often face challenges in managing resources such as beds, therapy facilities, and medical

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

personnel who have specialized expertise in mental health care.

Therefore, to overcome the possibility of unexpected fluctuations in patient needs, mental hospitals need an accurate prediction system to determine the estimated number of patient visits at a time in order to provide more effective, efficient and quality services. Accurate prediction of the number of patient visits can help hospital management to plan for more efficient use of facilities and manpower, reduce patient waiting times, and maximize the availability of medical services.

In this case, data mining is the right scientific solution to be able to provide the results of analysis to predict the number of visits to mental hospitals. Data mining can be used to dig up information from big data, including historical data on visits so that later information can be obtained that can be used in predicting patient visits (Tanjung 2021).

Naive Bayes algorithm has been shown to have good predictive performance in various previous studies. For example, research conducted by Dzulkarnain et al. that utilized this algorithm for classifying student graduation rates managed to achieve significant prediction accuracy, demonstrating Naive Bayes' potential in producing a strong and relevant prediction model. Thus, the application of the Naive Bayes algorithm in the context of predicting patient visits in mental hospitals is expected to be able to provide useful results for hospital management in optimal resource management (Dzulkarnain et al. 2023).

Furthermore, Tugiman et al conducted research using the naïve bayes algorithm where data on drug sales for five years and drugs used by patients participating in the National Health Insurance (JKN) were obtained. This study aims to develop a drug use prediction system for patients participating in the National Health Insurance (JKN) using the Naïve Bayes Classifier algorithm. Based on the results of software testing regarding Drug Use Prediction for Patients Participating in National Health Insurance (JKN) using the Naïve Bayes Classifier algorithm conducted by the researcher, the hypothesis and the results of the system test reached the expectation, which was 78.64% (good) (Tugiman et al. 2022).

Based on this background, this study utilizes the Naive Bayes algorithm to predict patient visits, which is an innovative step that offers a technology-based solution to overcome the problem. This study aims to assist mental hospitals in estimating the number of patient visits at a certain time, so that they are able to provide more efficient and responsive mental health services. Accurate predictions are expected to support hospital management in better financial and administrative planning, minimize unnecessary operational costs, and ensure that every patient receives adequate services without delay.

LITERATURE REVIEW

Prof. Dr. Muhammad Ildrem Psychiatric Hospital, North Sumatra Province, which is located on Jalan Letjend. Jamin Ginting Km.10/Jl. Tali Air number 21, Medan, is the only government psychiatric hospital in North Sumatra Province that has the ability to provide services classified as Type "A" with its specificity (Siallagan et al. 2023). This hospital is not only the leading mental health service center in Sumatra, but also serves as a referral mental hospital for other health facilities in the province as well as for public hospitals on the island of Sumatra.

Data mining is a concept that is applied in finding more value stored in a large database (Ulfa, Handayani, and Urva 2023). Data mining can also be interpreted as extracting new information taken from chunks of big data that helps in decision-making (Wibowo, 2020). From some of the explanations above, it can be concluded that data mining is a data processing technique that is sourced from large data that can produce useful information and can be used for an agency that uses data mining techniques in making better decisions (Mustika et al. 2021).

The Naïve Bayes algorithm is one of the algorithms in data mining where in its easy use and shorter process, and can be said to have a fairly simple structure with a high level of effectiveness (Fitri 2020). The Naive Bayes Classifier Theorem is a theorem with two different interpretations. This theorem states how

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

far the degree of subjective belief should change rationally when given a new clue. This theorem comes from the application of probability theory, which is how to know the probability of two different interpretations (M. Afriansyah et al. 2024). In the context of this study, Naive Bayes will be used to estimate the number of patient visits based on historical variables, such as the day of the visit, the time of year, and other factors that may affect the rate of visits (Attamami et al. 2023). Mathematically, Naive Bayes' Algorithm calculates the probability of a class based on a combination of existing features. The basic formula used in Naive Bayes is as follows equation (1):

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \quad (1)$$

Where:

$P(C|X)$: the posterior probability of class C (number of patient visits) given feature X (attributes such as days or the time of year).

$P(X|C)$: probability of the X feature given class C.

$P(C)$: the initial probability of class C.

$P(X)$: probability of feature X.

This research will apply a type of Naive Bayes known as Multinomial Naive Bayes. Naive Bayes' multinomials are specifically designed to handle data that can be represented as the frequency or occurrence count of features within each target category. For example, in the case of predicted patient visits, data such as days of the week, months, or other factors that can affect the number of visits can be interpreted as the frequency of occurrence in various time conditions (Dewi et al. 2023). One of the main advantages of Naive Bayes' algorithms, specifically Multinomial Naive Bayes, is its ability to handle big data efficiently and accurately (Mahendra et al. 2024).

Prediction is a process of systematically estimating what is most likely to happen in the future based on the past and present information that we have, so that the error (the difference between something that happened and the estimated result) can be minimized. Prediction does not have to give an exact answer to the event that will occur, but rather tries to find the answer as close as possible to what will happen (Baihaqi et al. 2023).

This study uses an evaluation or testing model in the form of a Confusion Matrix model, which is an evaluation carried out on Text Classification, including accuracy, precision, recall, and F1-Score. The following is the formula of precision, recall and accuracy in equations (2), (3), (4) and (5):

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

$$F1 - Score = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$

Where:

TP = The total of the positive data classified as a positive value by the system.

FP = The total of the negative data classified as positive values by the system.

FN = The total of the positive data classified as positive values by the system.

MR = The total of the negative data classified as negative values by the system.

Python is a high-level programming language created by Guido Van Rossum and released in 1991. In

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

addition, python is also a multi-functional programming language, one of which is in the fields of Machine Learning and Deep Learning (Alfarizi et al. 2023). In this study, the Python programming language was used with Google Colabs tools. Some of the libraries used in this study are numpy, pandas, and others. And for the data collection in this study, it was obtained from the Prof.Dr.M.Ildrem Mental Hospital.

METHOD

A research framework is a conceptual description or structure that illustrates the process of conducting research from beginning to end. The goal is to provide a clear understanding and guidance for researchers in conducting research in a systematic and structured manner.

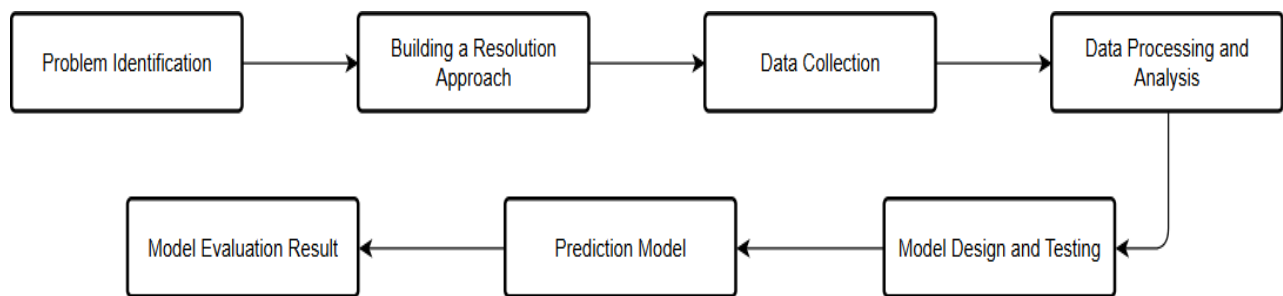


Fig 1. Research Outline

Problem identification, this stage is the first step in research, where the researcher defines and understands the existing problem. In the context of this study, the main problem is how to predict the number of patient visits at the Prof. Dr. M. Ildrem Mental Hospital so that the hospital can manage resources efficiently. Problem identification involves analyzing hospital needs in predicting patient visits based on existing historical data.

Data collection is an activity to collect data or information that is in accordance with the scope of the research. Data collection for this study was conducted by conducting observations and interviews at the Prof.Dr.M.Ildrem Mental Hospital located at Jl. Tali Air No.21, Mangga, Medan Tuntungan District, Medan City, North Sumatra 20141.

At this stage, data relevant to the number of patient visits in the hospital is collected through two main methods: interviews and direct data requests to the hospital. Interviews are conducted with hospital staff or management to understand factors that may affect the number of patient visits, such as certain times or events that have the potential to increase the number of visits. In addition, the researchers also asked for historical data on patient visits from hospitals to obtain numerical data that would be used in the training process and testing the model. Data obtained from these two sources must be verified and processed to ensure their quality and accuracy in supporting the research. The following is the flowchart of the Naive Bayes Algorithm:

* Corresponding author



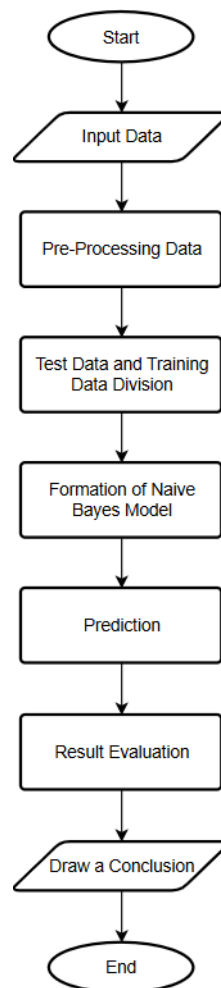


Fig 2. Naive Bayes Algorithm Flowchart

The prediction results, then the next stage is to obtain the prediction results from the model that has been tested. The results of this prediction show the estimated number of patient visits in the future, which can be used by hospital management for resource planning, workforce allocation, and provision of adequate facilities. The results of this prediction are expected to help hospitals provide more optimal services and respond better to patient needs.

The results of the model evaluation, at this stage, the evaluation performance process is carried out by mapping the prediction results to the confusion matrix table, with the existence of confusion matrix makes it easier to calculate precision, recall, fl-score, and accuracy. The next stage is to calculate precision, which is the comparison of the number of correct predictions with the total predictions. Then calculate the recall, which is the comparison of the correct prediction with the total correct prediction, when the precision and recall values are already there, the fl-score value can be calculated. Then it ends by calculating the accuracy of the model that has been made. The evaluation stage ends by analyzing the results of the evaluation performance.

RESULT

In this study, the author conducted a data collection process through observation and interviews at the

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Prof. Dr. M. Ildrem Mental Hospital, located on Jl. Tali Air No. 21, Mango, Medan Tuntungan District, Medan City, North Sumatra 20141. The data collected includes 1663 patient visit data with various attributes, such as age, gender, patient status, payment method, and service unit.

Input Data

Using the Naive Bayes algorithm, this data is used to predict the service unit to which the patient is going. Before use, pre-processing processes are performed, such as using LabelEncoder to convert non-numeric attributes to numeric and add new features, such as age categories, number of patient visits per month, and number of patient visits per day.

The pre-processing data results are stored in a csv file, and then divided into training data and test data with an 80:20 ratio. Next, the model is tested to calculate performance metrics such as accuracy, precision, recall, and F1 score. The goal is to evaluate the accuracy of the model in predicting service units. In addition, data visualization is used to support the analysis and interpretation of research findings. The following dataset is used.

Table 1. Raw Data

No.	No.R	Patient Name	JK	Age	Stat	Tgl	How to	Service
	M				us	Reg	Pay	Unit
1.	49290	AGUS MILA	Man	27	New	3/6/20	No	Drug
		SEMBIRING				24	Insura	Clinics
							nance/Ge	
							neral	
2.	40770	IRIANTO	Man	61	New	4/6/20	No	Psychoger
						24	Insura	iatric
							nance/Ge	Clinic
							neral	
...	...	...	...	...	...	...	...	...
5.	47516	SURYA	Man	10	New	4/6/20	BPJS /	Child and
		NATANAEL				24	JKN	Adolescen
		SIMORANG						t
		KIR						Psychiatri
								c Clinic

The next step is to categorize the data by age range to make it easier to analyze and model predictions. The age categories are grouped into three, namely Children (≤ 12 years old), Teenagers (13–25 years old), and Adults (> 25 years old). The grouping process is done by adding a new column named Age Category to the dataset.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Table 2. Additional Data on Age Categorization

No.	No. RM	Patient Name	JK	Age	Status	... Kategori_U
						mur
1.	49290	AGUS MILA	Man	27	New	... Adult
		SEMBIRING				
2.	40770	IRIANTO	Man	61	New	.. Adult
...	...	...	...	...	...	...
166	41761	MASITAH	Woman	69	Old	... Adult
3						

Furthermore, the "No.RM" and "Patient Name" columns were removed because they did not contribute directly to the prediction or data analysis process, so they were removed. In order for the data to be processed correctly, then change the format of the "Reg Date" column to the datetime format. Previously, the data in those columns did not have a standard format (mm/dd/yyyy), so the `pd.to_datetime()` function had to be used to change the available date format. Once the conversion process is complete, two new columns are added: "Month" and "Day". The "Month" column shows the month taken from the registration date, while the "Day" column shows the days of the week by date, with the number 0 for Monday and 6 for Sunday. In addition, the number of patient visits is calculated based on a specific month and day. To do this, the `groupby()` and `transform()` functions are used to count the number of patients enrolled each month and day.

Preprocessing Data

Table 3. Data Before Pre-Processing

No.	JK	U	S	TR	CB	UP	I	Mont	Da	JPT	JPT
								h	y	B	H
1.	Man	2	Ne	2024	No	Drug Clinics	Adul	6	0	422	372
		7	w	-06-	Insurance/Gene		t				
				03	ral						
2.	Man	6	Ne	2024	No	Psychogeriat	Adul	6	1	422	283
		1	w	-06-	Insurance/Gene	ric Clinic	t				
				04	ral						
...	...	...	...	...	...	...	...	...	...	...	...
166	Woma	6	Old	2024	No	Adult	Adul	9	4	409	327
3	n	9		-09-	Insurance/Gene	Psychiatric	t				
				06	ral	Clinic					

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Information:

JK :Gender

U :Age

S :Status

TR : Reg Times

CB : How to Pay

UP : Service Unit

I : Age Category

Month :Moon

JPTB : Number of Patients

Last Month

JPTH : Number of Patients

Last Day

Then the pre-processing process, after adding new columns, the next step is to pre-process the data to make the dataset ready to be used in the prediction model. LabelEncoding is one of the pre-processing methods used. The LabelEncoding technique converts categorical data, which consists of text or labels, into numerical data so that machine learning models can process the data in numerical format. It converts each category present in the column into a unique numeric value. For example, if there are two categories such as "Male" and "Female" in the "Gender" column, LabelEncoding will convert them to the numbers 0 and 1, for example, for "Male" and "Female", respectively.

Table 4. Data Pre-Processing Results

No.	JK	Age	Status	Tgl Reg	How to Pay	Service Unit	Age Category	Moon	Day	JPTB	JPTH
1.	0	27	0	2024- 06-03	3	4	1	6	0	422	372
2.	0	61	0	2024- 06-04	3	6	1	6	1	422	283
...	...	...	...	...	...	...	...	...	...	...	...
1663	1	69	1	2024- 09-06	3	2	1	9	4	409	327

Test data and Training Data Division

One of the important steps in the process of creating a machine learning model is the division of data into training data and test data. The goal of this division is to avoid overfitting and ensure that the built model can function properly with data that has never been seen before. The total of 1664 data in the research dataset is divided into two parts, namely 80% for exercise data and 20% for test data.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

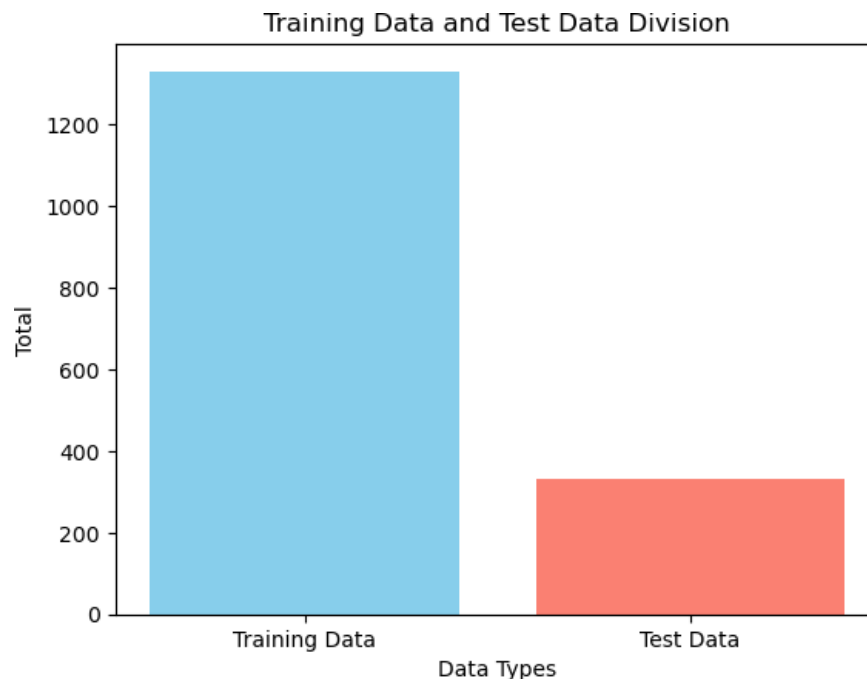


Fig 3. Distribution of the Number of Test Data and Training Data

From the figure above, as many as 1331 data are used as training data to train the Naive Bayes model. This section is very important because it helps the model learn the relationship between the feature and the target variable it wants to predict. Meanwhile, 333 pieces of data were separated into test data to test the ability of trained models to accurately predict data that had never been seen before.

4.1 Formation of Naive Bayes Model

Furthermore, the implementation of the Naive Bayes algorithm model, where in this study, the model used is Naive Bayes with the MultinomialNB approach. Naive Bayes' algorithm works based on Bayes' theorem which states that the probability of a class (label) being given certain features. The following is the manual calculation of the naïve bayes model, namely there are 10 training data and 5 test data as an example of how the naïve bayes algorithm works:

Table 5.
Sample Training Data

J K	Stat us	How to Pay	Age Category	Number of Patients Last Month	Number of Patients Last Day	Mo on	Da y	Service Unit
0	1	1	1	413	250	7	2	2
0	1	1	1	422	283	6	1	2
...	...	...	...	...	...	...	...	...
0	0	3	1	422	372	6	0	2

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Table 6. Sample Test Data

J K	Stat us	How to Pay	Age Category	Number of Patients Last Month	Number of Patients Last Day	Mo on	Da y	Service Unit
1	1	3	1	409	250	9	2	6
0	0	1	1	422	372	6	0	2
..	...	...	...	...	...	...	...	...
0	0	3	1	422	250	6	2	7

Calculation of the prior probability value ($P(Class)$) :

Using equation (1), we will obtain the prior probabilities for the first 5 classes:

$$P(\text{Service Units} = 2) = \frac{1038}{1331} = 0.779865$$

$$P(\text{Service Units} = 6) = \frac{106}{1331} = 0.079639$$

$$P(\text{Service Units} = 7) = \frac{82}{1331} = 0.061608$$

$$P(\text{Service Units} = 9) = \frac{29}{1331} = 0.021788$$

$$P(\text{Service Units} = 4) = \frac{25}{1331} = 0.018783$$

Furthermore, the calculation of the value of the conditional probability: ($P(Fitur|Kelas)$)

JK feature, if "JK" = 1 (Female):

$$P(\text{JK} = 1 | \text{Service Unit} = 2) = \frac{375}{1038} = 0.361272 \text{ (5 times calculation to service unit = 4)}$$

Feature "Status" = 1 (Old):

$$P(\text{Status} = 1 | \text{Service Unit} = 2) = \frac{703}{1038} = 0.677264 \text{ (5 times calculation to service unit = 4)}$$

"How to Pay" Feature = 3 (Uninsured/General):

$$P(\text{How to Pay} = 3 | \text{Service Unit} = 2) = \frac{220}{1038} = 0.211974 \text{ (5 times the calculation to the service unit = 4)}$$

Feature "Age Category" = 1 (Teen):

$$P(\text{Age Category} = 1 | \text{Service Unit} = 2) = \frac{415}{1038} = 0.399227 \text{ (5 times the calculation to the service unit = 4)}$$

Calculation of posterior probability values: ($P(Class | Data)$)

For Service Unit = 2 :

$$\begin{aligned} P(\text{Unit Pelayanan} = 2 | \text{Data}) &= 0.779865 * 0.361272 * 0.677264 * 0.211974 * 0.399227 \\ &= 0.03108 \end{aligned}$$

... (up to service unit = 4)

$$P(\text{Unit Pelayanan} = 4 | \text{Data}) = 0.018783 * 0.04 * 0.8 * 0.2 * 0.2 = 0.00001$$

The prediction model concludes from the calculation results, the class with the greatest probability is Service Unit = 2. So, the prediction for the first test data is 2.

Prediction

Predict the number of patients where the prediction will be made by creating new data in the form of a Python dictionary that contains features for prediction, such as gender (JK), status, payment method (How to Pay), age category (Kategori_Umur), the last number of patients per month and day, and time information (month and day). Each key in the dictionary has a value in the form of a list, such as [0] for a category number or example data. The new data used is where the patient is 'Female' [0] with the status of 'new' [0], the payment method uses 'BPJS' [1] and is categorized as 'Adult' [2] where the number of patient data this

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

month is [200] and the total today is [30] and he came in 'May' [5] and on 'Wednesday' [3]. The prediction results are as follows:

```
31]: # Prediksi menggunakan model
prediksi_kunjungan = model_nb.predict(df_baru)

# Menampilkan hasil prediksi
predicted_label = label_encoders['Unit Pelayanan'].inverse_transform(prediksi_kunjungan)
print("\nPrediksi Unit Pelayanan:", predicted_label[0])
```

Prediksi Unit Pelayanan: Klinik Kulit dan Kelamin

Fig 4. Prediction of the intended service unit

From the picture above, it shows that the new patient data will go to the "Skin and Venereal Clinic" service unit. Then make a prediction model to determine the total number of patients.

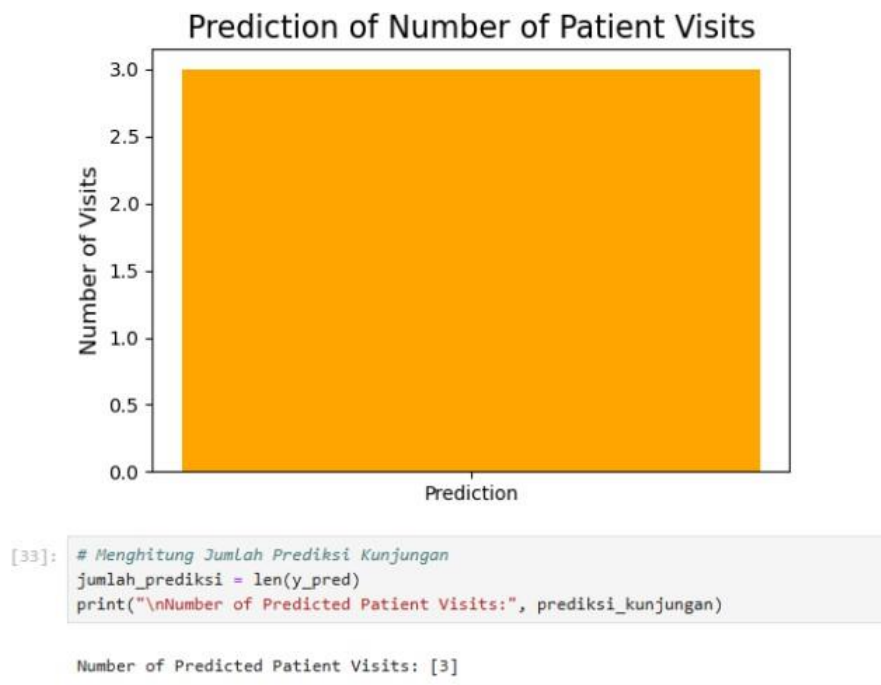


Fig 5. Prediction of Number of Patients

From the picture above, it can be concluded that the number of patients who will go to the 'Skin and Venereology Clinic' service unit is 3 patients.

Result Evaluation

Then model evaluation is needed where the classification results to look for accuration, precision, recall and F1-score follow the following formula, for example in the "Adult Psychiatric Clinic" element:

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

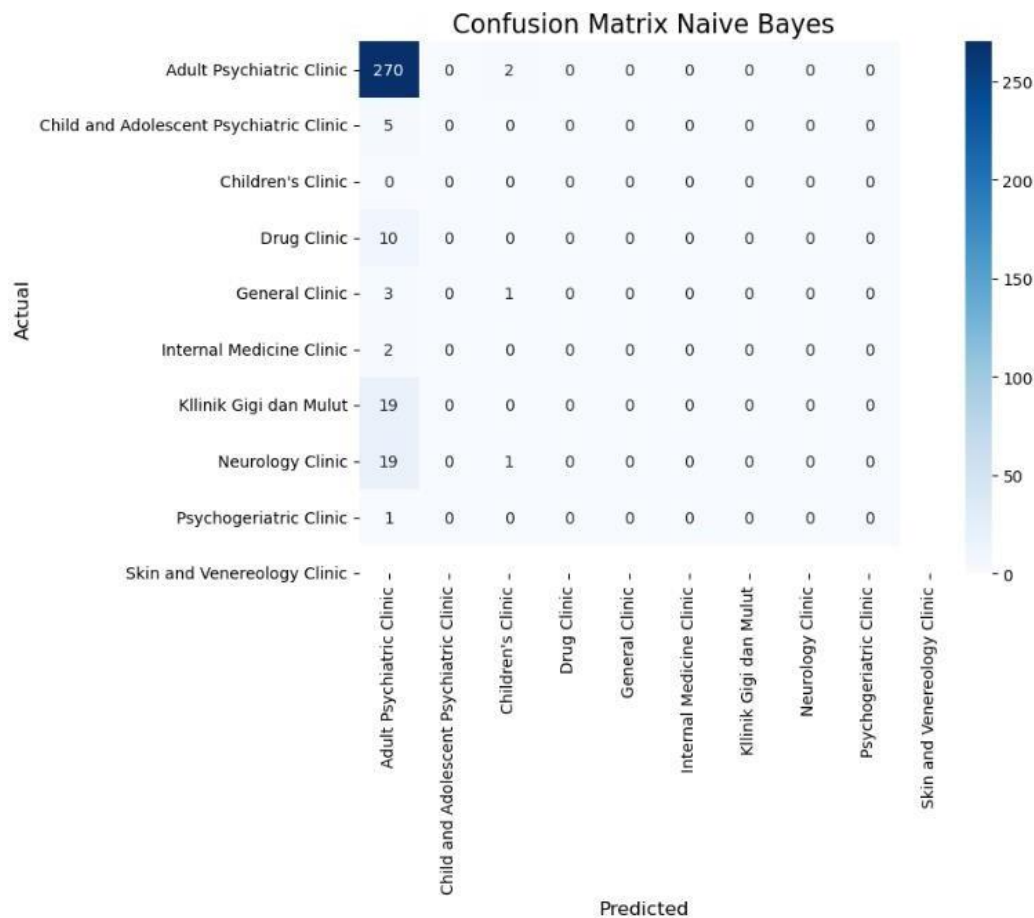


Fig 6. Confusion Matrix

$$\begin{aligned}
 accuracy &= \frac{TP}{TP + TN + FP + FN} = \frac{270}{270 + 3 + 58 + 2} = \frac{270}{333} \\
 &= 0.8108108108108109 \\
 precision &= \frac{TP}{TP + FP} = \frac{270}{270 + 58} = \frac{270}{329} = 0.8206686930091185 \\
 recall &= \frac{TP}{TP + FN} = \frac{270}{270 + 2} = \frac{270}{272} = 0.9926470588235294 \\
 F1 - Score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
 &= 2 \times \frac{0.8206686930091185 \times 0.9926470588235294}{0.8206686930091185 + 0.9926470588235294} \\
 &= 2 \times 0.449251247920133 = 0.90
 \end{aligned}$$

Draw a Conclusion

Based on the conclusion, the evaluation results showed an accuracy of 0.8108108108108109 or 81%, followed by a % precision of 0.8206686930091185 or 82.07%, recall had a value of 0.9926470588235294 or 99.26%, and an F1-score of 0.90 or 90%.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Classification Report:

precision recall f1-score support

0	0.82	0.99	0.90	272
1	0.00	0.00	0.00	5
2	0.00	0.00	0.00	0
3	0.00	0.00	0.00	10
4	0.00	0.00	0.00	4
6	0.00	0.00	0.00	2
7	0.00	0.00	0.00	19
8	0.00	0.00	0.00	20
9	0.00	0.00	0.00	1

accuracy			0.81	333	
macro avg	0.09		0.11	0.10	333
weighted avg	0.67		0.81	0.73	333

Akurasi Model: 0.8108108108108109

DISCUSSIONS

The results of this study demonstrate the effectiveness of the Naïve Bayes algorithm, specifically its Multinomial approach, in predicting patient visits with high reliability, achieving an accuracy of 81%, precision of 82.07%, recall of 99.26%, and an F1-score of 90%. These findings align with prior studies, such as Dzulkarnain et al. (2023), which also highlight the robustness of Naïve Bayes for classification tasks in healthcare contexts. The application of this model provides significant benefits for hospital management, including optimizing resource allocation, reducing patient wait times, and improving service quality. However, the study is limited by its dataset, which is dominated by adolescent patient data, potentially impacting the generalizability of the findings. Future research should involve more diverse datasets and consider other algorithms, such as Random Forest or Support Vector Machines, to enhance the predictive model's accuracy and versatility.

CONCLUSION

This study uses the Naïve Bayes algorithm to predict the service unit at the Prof. Dr. M. Ildrem Psychiatric Hospital. The data collection process included observations and interviews, resulting in 1,662 patient visit data with attributes such as age, gender, patient status, payment method, and service unit. This data is processed by pre-processing using LabelEncoder and adding features such as age categories and the number of visits per day and month. The results of the data division showed an 80:20 ratio for training data and test data, where training data amounted to 1331 data and test data had 333 data. Naïve Bayes' Multinomial algorithm was chosen for its ability to handle categorical data. Based on confusion matrix analysis, the model achieved an accuracy of 0.8108108108108109 or 81%, a precision of 0.8206686930091185 or 82.07%, a recall value of 0.9926470588235294 or 99.26%, and an F1-score of 0.90 or 90%, which shows that this model is quite effective in predicting service units with a dominance of adolescent patient data.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

REFERENCES

- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). "The Use of Python as a Programming Language for Machine Learning and Deep Learning." *Scientific Work of Tawhid Students (KARIMAH TAUHID)* 2(1):1–6.
- Ambarita, Togi Fitri A. (2020). "Correlation of Psychological Wellbeing with Burnout in Nurses of the Mental Hospital Prof. Ildrem Medan Provincial Government." *Journal of Psychology of HKBP University Nommensen* 6(2):78–91. doi: 10.36655/psychology.v6i2.138.
- Attamami, Nurfazriah, Agung Triayudi, and Rima Tamara Aldisa. (2023). "Performance Analysis of Naive Bayes and C4.5 Classification Algorithms for Prediction of Health Insurance Assistance Recipients." *Journal of JTIK (Journal of Information and Communication Technology)* 7(2):262–69. doi: 10.35870/jtik.v7i2.756.
- Dewi, Christine, R. C. Chen, Christanto, & Cauteruccio. (2023). "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database." *Vietnam Journal of Computer Science* 10(4):485–98. doi: 10.1142/S2196888823500100.
- Dicky, Rahmat, & Mumpuni. (2024). "Implementation of Naive Bayes' Multinomial Method in Clickbait News Title Classification." *Neptune: Journal of Computer Science and Information Technology* 2(3):303–16. doi: 10.61132/neptune.v2i3.249.
- Fitri, Evita. (2020). "Sentiment Analysis of Ruangguru Application Using Naive Bayes, Random Forest and Support Vector Machine Algorithms." *Journal of Transformatika* 18(1):71. doi: 10.26623/transformatika.v18i1.2317.
- Iqbal, Maulana, Syaripudin, & Nugroho. (2023). "Implementation of Random Forest Algorithm on Stock Price Prediction Based on Historical Data." *Jubitek: JOURNAL OF BIG DATA AND INFORMATION TECHNOLOGY* 1:42–51.
- M. Afriansyah, Saputra, Ardhana, & Sa'adati. (2024). "Efficient Naive Bayes Algorithm for Classification of Plantains Based on Color Features." *Journal of Information Systems Management and Digital Business* 1(2):236–48. doi: 10.59407/jismdb.v1i2.438.
- Mustika, Ardila, Manuhutu, Ahmad, Hasbi, Guntoro, Manuhutu, Ridwan, Hozairi, Wardhani, Alim, Romli, Religia, Octafian, Sufandi, & Ernawati. (2021). *Data Mining and Its Applications*.
- Siallagan, Ance, Girsang, & M. R. Ompusunggu. (2023). "The Relationship between Family Support and the Rate of Schizophrenia Relapse at the Prof. DR. M. Ildrem Medan Psychiatric Hospital in 2023." *Tambusai Education Journal* 7(3):21747–58.
- Tanjung, Yani. (2021). "Optimization of C4.5 Algorithm for Prediction of ATM Machine Failure." *Infosys (Information System) Journal* 6(1):12. doi: 10.22303/infosys.6.1.2021.12-21.
- Tsania, Ratnawati, Rahayudi. (2023). "The Use of the Naïve Bayes Classifier Method in Sentiment Analysis of the Community's Assessment of Hospital Services in." 10(7):1453–60. doi: 10.25126/jtiik.2024117979.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

- Tugiman, Damayanti, Gunawan, & Elkana. (2022). "Prediction of Drug Use of National Health Insurance Participants Using the Naïve Bayes Classifier Algorithm." *Journal of Applied Computer Science and Technology* 3(1):144–50. doi: 10.52158/jacost.v3i1.295.
- Ulfa, Novia, Handayani, & Gellysa Urva. (2023). "Implementation of the Frequent Pattern Growth Algorithm to Analyze Consumer Shopping Patterns at Pharmacy Hopefully Jaya." *JUTEKINF (Journal of Computer and Information Technology)* 11(2):154–65. doi: 10.52072/jutekinf.v11i2.663.
- Wibowo, Radityo, & Jananto. (2020). "Implementation of FP-Growth Algorithm Association Method Data Mining in Retail Companies." *Inspiration: Journal of Information and Communication Technology* 10(2):200. doi: 10.35585/inspir.v10i2.2585.

* Corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).