Journal of Medical Records and Information Technology (JOMRIT)

Vol.2, No. 2, December 2024,
ISSN: XXX-XXXX,

PERFORMANCE ANALYSIS OF NAÏVE BAYES ALGORITHM IN PREDICTING THE EXIT CONDITION OF STROKE PATIENTS AT DR.M. DJAMIL PADANG HOSPITAL

Rafifah Trinanda[1] , Nurul Abdillah[2] , Hasrinal
[123] Universitas of Syedza Saintika, Padang, West Sumatra, Indonesia

Info article

Article history:

Received October 08, 2024
Revision October 12, 2024
Accepted December 28, 2024

Keywords:

Data Mining
Classification
Naive Bayes

ABSTRACT

Medical record files are often just a pile of meaningless files and are not carried out further investigation to produce better services in the future. The number of strokes in Indonesia is the most common disease and ranks first in Asia, in West Sumatra the highest rate of stroke is found in Dr. M. Djamil Padang Hospital, therefore the author is interested in taking the title Performance Analysis of Naive Bayes Algorithm in Predicting the Exit Conditions of Stroke Patients at Dr. M. Djamil Padang Hospital.

The research method used is quantitative research using data mining classification, conducted from April – May 2024 with data sources derived from medical record data at Dr.M. Djamil Hospital in January, February, and March 2023. The number of population after passing the data selection process is 500 and a sample of 222 data was obtained. The sampling technique is a random sampling technique with the slovin formula. The attributes used are Age, Gender, Address, Length of Care, Care Class, Occupational Status, and Congenital Diseases.

The results of this study obtained an accuracy value of 77.48% because it produced 172 data that were correctly predicted from 222 data, while the data error value obtained was 22.52% because it produced 50 data that were incorrectly classified from 222 data.

Based on the results of the research obtained where the level of accuracy has a higher value than the error value, and is in the category of Good Clasification, it can be concluded that the performance of the Naive Algorithm in predicting the discharge condition of stroke patients at Dr. M. Djamil Padang Hospital is very good and can be implemented in hospitals and further research is carried out on other diseases.

Correspondence Author:

Rafifah Trinanda
Health Information Management Study Program, University of Syedza Saintika
Hamka No. 228 East Freshwater, Padang, West Sumatra, Indonesia
Email: rafifahrinanda@gmail.com

1. INTRODUCTION

Stroke is one of the dangerous diseases with a high mortality rate in Indonesia. At this time, stroke is increasingly becoming a serious problem that must be faced almost all over the world. This is because a sudden stroke can result in death, physical disability, and mental disability both at productive and elderly age (Pambudi et al., 2022). These

disorders suddenly cause symptoms including paralysis of the face or limbs, slurred speech, changes in consciousness, visual impairment and others. Factors that cause stroke include; health factors (hypertension, cholesterol, obesity and heart), lifestyle factors (smoking, lack of exercise, consumption of illegal drugs and alcohol addiction), hereditary factors, and increasing age (Haryadi et al., 2021).

The stroke rate in Indonesia is the most common disease and ranks first in Asia. East Kalimantan is the region with the highest number of stroke patients with (14.7%), followed by Yogyakarta (14.3%), Bangka Belitung and DKI Jakarta each (11.4%). Meanwhile, West Sumatra is in the 12th position with the prevalence (12.2%) of sufferers. Based on the annual report of the Padang City Health Office in 2018, data was obtained that there were (10.8%) patients who experienced stroke. In accordance with the initial survey, the researcher obtained data on the most stroke patients at the Rasidin Padang Hospital with 41 patients every 3 months. And based on the results of medical records at the TK Hospital. III Dr. Reksodiwiryo Padang, the prevalence of stroke incidence in 2018 was 416 people, in 2019 it increased by 456 people and decreased in 2020 by 391 people.

Data mining involves the use of statistical techniques, mathematics, artificial intelligence, and machine learning in order to uncover valuable information related to large data sets. The main focus of data mining is to detect, uncover, or explore knowledge that can be extracted from available data or information (D. F. Pasaribu., 2021). The purpose of this classification method is to classify the value of an unknown variable from other predetermined variables (Nugraha et al., 2022).

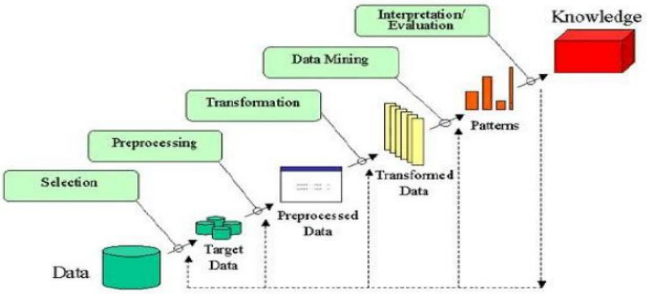
There are several classifications in data mining, one of which is *Naïve Bayes*.*Naïve Bayes* has the advantage of using this method in that it only requires small training data in determining the expected parameters in the classification (Watratan et al., 2020).

According to the research (Agus Fajar Riany, Gusmelia, 2023) whose research aims to classify data related to stroke disease using *the Naïve Bayes* algorithm to find out whether the patient has stroke or not, from this test it produces an accuracy level of 92.48% which is in the *Good Classification category*.

Based on the above problems, the author will discuss how to apply the Naive Bayes algorithm in solving the problem where medical record files whose contents are historical records of patient documents during treatment at the hospital are often only a pile of meaningless files and are not carried out further investigation so that they can be evaluated for the hospital and can produce better services in the future. West Sumatra has the highest stroke rate at Dr.M.Djamil Padang Hospital, the author obtained data in the last 3 months in 2023, namely there are 900 patients with cases of stroke disease. Therefore, the author is interested in taking the title "Performance Analysis of *Naive Bayes Algorithm* in Predicting the Discharge Conditions of Stroke Patients at Dr. M. Djamil Padang Hospital"

2. METHOD

This research was conducted at Dr. M. Djamil Padang Hospital, conducted from April – May 2024 with data sources derived from medical record data at Dr. M. Djamil Hospital in January, February, March 2024. This type of research is a quantitative research using the data mining classification method with the Naïve Bayes algorithm to calculate the level of accuracy and error of data. Data collection by: Observation, Interview, Literature Study, Browsing. The population used in this study is 500 data on patients diagnosed with stroke from January to March 2023. The sample obtained from January to March 2023 is 222 data obtained using the slovin formula.



Picture 1 Knowledge *Discovery Process in Database*

2.1 Selection

The raw data that has been obtained is selected first, for the selection of what data is in accordance with the characteristics of the author's research data. The form of raw data obtained by the bidder researcher is seen in the table below:

Table 1 Medical Record Data Format After Variable Selection

NO	*Age*	*Gender*	*Address*	*Long Maintained*	*How to Pay*	*Treatment Classes*	*Diagnosis Seconds*	*How to Exit*
1	59	Man	PADANG CITY	9	JKN MANDIRI	ICU/NICU/PICU/CVCU	18.9 - Pneumonia	Die
2	50	Man	BUKITTINGGI	7	JKN CIVIL SERVANTS CENTRAL	HCU	18.9 - Pneumonia	Die
3	58	Woman	JORONG VILLAGE	1	JKN PBI (APBN)	HCU	18.9 - Pneumonia	Die
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
220	56	Woman	FULL RIVER	3	JKN REGIONAL CIVIL SERVANTS	Class I		Recover
221	25	Woman	JAMBI	1	JKN PRIVATE EMPLOYEES	HCU	J18.9 - Pneumonia, unspecified	Die
222	80	Man	FIELD	19	JKN TNI PENSION RECIPIENTS	Class II	J18.9 - Pneumonia, unspecified	Recover

2.2 Prepocesing

Pre-processing or data pre-processing is an important stage in data processing before applying Naive Bayesian algorithms or other machine learning algorithms. The goal is to clean, prepare, and change the data to make it more suitable for analysis and provide more accurate results. Naive Bayes, like other algorithms, works best with structured, clean data.

Table 2 format after passing through pre-processing

It	***Age***	***Gender***	***Address***	***Long Treatment***	***Treatment Classes***	***Citizenship Status***	***Congenital Diseases***	***How to Exit***
1	Elderly	L	In West Sumatra	>3 days	ICU/NICU/PICU/CVCU	BPJS	Exist	Die
2	Elderly	L	In West Sumatra	>3 days	HCU	BPJS	Ada	Die
3	Elderly	P	In West Sumatra	<= 3 days	HCU	BPJS	None	Die
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
220	Elderly	P	In West Sumatra	<= 3 days	Class I, II, III	BPJS	Ada	Live
221	Adolescent	P	Outside West Sumatra	<= 3 days	HCU	BPJS	Ada	Live
222	Seniors	L	In West Sumatra	>3 days	Class I, II, III	BPJS	Ada	Die

2.3 Transformation

At this stage, the transformation process is carried out on the initial dataset. Transformations are performed to convert the values of some attributes that are numerical, into nominal. The following are some of the attributes that have numerical values that are converted into nominal, these attributes include: Age, Gender, Address, Length of Treatment, How to Pay, Treatment Class, and Secondary Daignosa, and how to exit. Which is classified as the table below:

Table 3 Attribute Transformation

NO		**NUMBER OF CASES**	**LIVE**	**DIE**
	Total	222	162	60
1.	Age	Toddler	6	6	0
		Children	2	0	2
		Adolescent	8	8	0
		Adult	24	14	10
		Elderly	105	75	30
		Seniors	77	59	18
2.	JK	L	120	93	27

NO		NUMBER OF CASES	LIVE	DIE
	Total	222	162	60
		P	102	69	33
3.	Address	In West Sumatra	186	136	50
		Outside West Sumatra	36	26	10
4.	Long Treatment	<= 3 Days	36	17	19
		> 3 Days	186	145	41
5.	Nursing Classes	ICU	15	8	7
		HCU	132	83	40
		Class I, II, III	82	69	13
		VIP	2	2	0
6.	Citizenship Status	BPJS	219	160	59
		Non BPJS	3	2	1
7.	Congenital Diseases	Ada	208	152	56
		None	14	10	4

2.4 Data Mining

This stage is the implementation stage of the selected technique and algorithm. The tools used to perform manual calculations use Microsoft Excel, while for the test they also use Rapid Miner Software. In this study, classification will be applied using the algorithm that has been selected, namely Naïve Bayes. The stages of data modeling using Naïve Bayes' algorithm are as follows:

2.4.1. Calculating the Probability of Each Class

After making a table of data analysis, the initial stage in Naïve Bayes' calculation is to calculate the Class Prior Recovered and Died.

Table 4 Decision class probability

CLASS	CLASS PROBABILITY
LIVE	0,730
DIE	0,270

2.4.2 Calculating the probability of each event per class

To calculate the probability value of each attribute by calculating the number of events or attributes in a class divided by the existing class.

Table 5 Probability of each event per class

NO	ATTRIBUTE	SUB ATTRIBUTES	LIVE	DIE
1.	Age	Toddler	0,037	0,000
		Children	0,000	0,033
		Adolescent	0,049	0,000
		Adult	0,086	0,167
		Elderly	0,462	0,050
		Seniors	0,364	0,300
2.	JK	L	0,574	0,450
		P	0,426	0.500
3.	Address	In West Sumatra	0,840	0,833
		Outside West Sumatra	0,160	0,167
4.	Long Treatment	<= 3 Days	0,080	0,317
		> 3 Days	0,895	0,683
5.	Nursing Classes	ICU/NICU/PICU/CVCU	0,493	0,117
		HCU	0,512	0,645
		Class I, II, III	0,426	0,217
		VIP	0,015	0,000

NO	ATTRIBUTE	SUB ATTRIBUTES	LIVE	DIE
6.	Citizenship Status	BPJS	0,012	0,017
		Non BPJS	0,988	0,983
7.	Congenital Diseases	Ada	0,062	0,067
		None	0,938	0,933

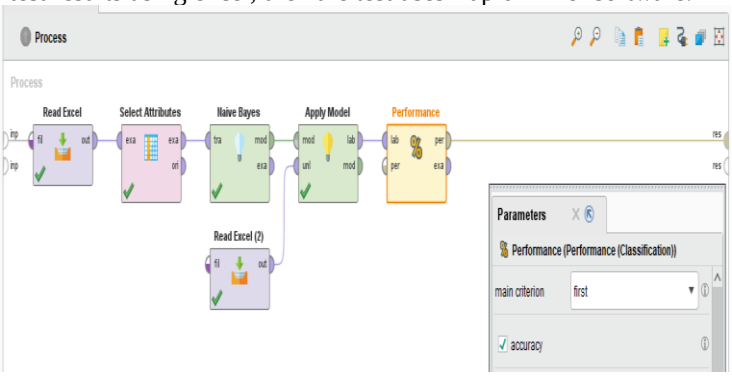
2.5 Interpretation

From the probability values that have been obtained, it can be used to predict or determine the test data label by multiplying each class attribute. then manual calculations are carried out using Microsoft Excel, this is where the application of the naïve bayes algorithm is carried out, namely by performing manual calculations using Microsoft Excel. Based on the results of the manual calculation of each class, then compare which is the largest probability value of the classification results. Like the table below:

Table 6 Results of manual calculation of naïve bayes

NO	PROBABILITY		PROBABILITY OF NAÏVE BAYES' CALCULATION
	Recover	Die	
1	0,006669014	0,003707802	Live
2	0,069191019	0,021187437	Live
3	0,000395962	0,000857177	Die
.	.	.	.
.	.	.	.
220	0,005003429	0,003900154	Live
221	0,000122733	0	Live
222	0,045249259	0,00413155	Live

After obtaining the test results using excel, then the test uses Rapid Miner software.



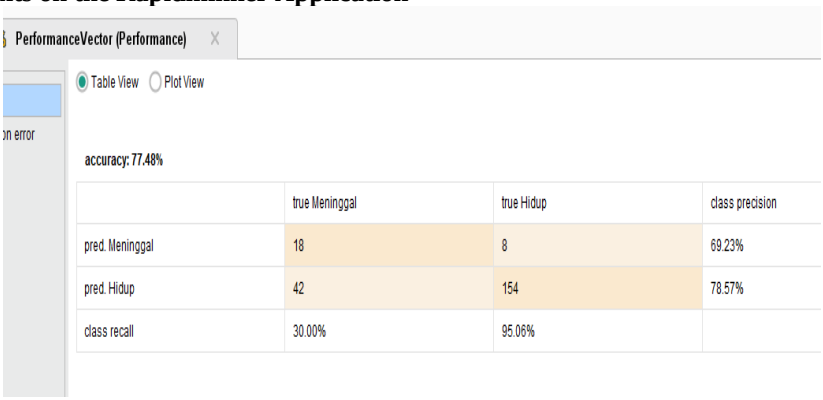
Picture 2 Modeling on Rapidminner

The image above is a model formed on the Rapid Miner tools used to test the test data.

3. RESULTS AND DISCUSSION

Data mining processing is the main result in this study whose findings will be discussed in this results and discussion.This study uses the naïve Bayes algorithm in predicting the discharge condition of stroke patients alive or dead by using the application of the data mining method, using the following stages:

3.1. Accuracy Results on the Rapidminner Application

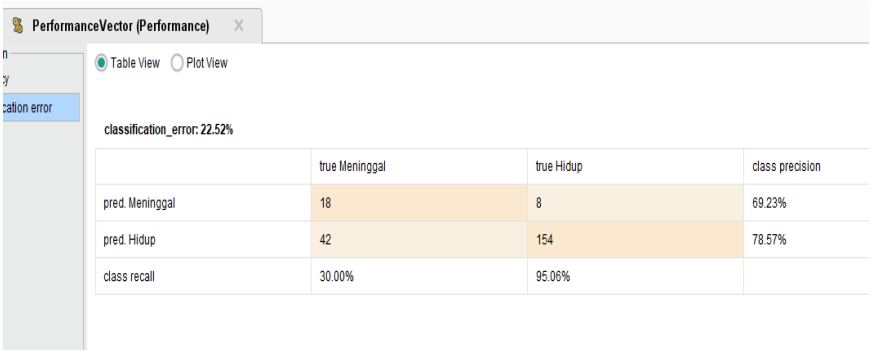


	true Meninggal	true Hidup	class precision
pred. Meninggal	18	8	69.23%
pred. Hidup	42	154	78.57%
class recall	30.00%	95.06%	

Picture 3 Accuracy on Rapidminner

Based on the figure above, it can be seen that the accuracy value obtained in the rapidminner application on stroke disease data at Dr.M.Djamil Padang Hospital was obtained with an accuracy value of 77.48%. Calculating the accuracy value is to see how well the naïve bayes algorithm classifies the data correctly, and in the figure above it is 77.48% which produces 172 data correctly out of 222 stroke disease data which is located in the Good Classification category.

3.2. Error Results in the Rapidminner Application



	true Meninggal	true Hidup	class precision
pred. Meninggal	18	8	69.23%
pred. Hidup	42	154	78.57%
class recall	30.00%	95.06%	

Picture 4 Error Value on Rapidminner

Calculating the error value is to see how the naïve bayes algorithm in classifying data as incorrect using the Rapidmiinner application, based on the figure above it can be seen that the error value obtained is 22.52% where it classifies 50 data incorrectly from 222 actual data.

4. CONCLUSION

Based on the results of a series of research stages that have been carried out, it can be concluded that the data mining process with classification techniques using the Naïve Bayes algorithm can be applied in the Grouping of Exit Conditions of Stroke Patients Living or Dead. Where based on the testing process that has been carried out manually using Microsoft Excel and proof on the Rapidminner Application, an accuracy value of 77.48% and an error value of 22.52% were obtained, where the accuracy value was higher than the error value, which can be concluded that the performance of the Naïve Bayes algorithm in predicting the discharge of stroke patients at Dr.M.Djamil Padang Hospital is effective and is in the *Good classification category.*

REFERENCE

[1] Airi, F. A. H., Suprapti, T., & Bahtiar, A. (2023). Comparison of Data Mining Classification Methods for Stroke Disease Prediction. *E-Link: Journal of Electrical Engineering and Informatics*, *18*(1), 73. https://doi.org/10.30587/e-link.v18i1.5271

[2] Fadrial, Y. E. (2021). Algoritma *Naive Bayes* Untuk Mencari Perkiraan Waktu Studi Mahasiswa *Naive Bayes* Algorithm for Finding Student Estimated Time Students. *Journal of Information Technology and Computer Science (INTECOMS)*, *4*(1), 20–29.

[3] Handayani, T., & Feoh, G. (2019). Design of a Web-Based Medical Record Information System (Case Study at Sriati Maternity Clinic Kota Sungai TFull – Jambi). *Journal of Information and Computer Technology*, *2*(2), 226–236. https://doi.org/10.36002/jutik.v2i2.148

[4] Islam, U., & Syarif, N. (2023). *Overview of Service Quality in Outpatient Installation at the Palangka Raya City Regional General Hospital in 2022*. xviii + 136 pages + 18 tables + 25 appendices. https://repository.uinjkt.ac.id/dspace/bitstream/123456789/67607/1/Imam Syafe%27i - FIKES.pdf

[5] Qualification, M., & Home, K. (2020). *1 ,) , 2*. *11*(2), 101–108.

[6] Mauliadi, R. (2022). *Journal of Informatics Business Economics Data Mining Using the K-Means Clustering Algorithm in the Analysis of Discount Rate on the Selling Price of Honda Motorcycles*. *4*, 7–9. https://doi.org/10.37034/infeb.v4i4.156

[7] Murni Wijaya, W. S., & Mutaqin, Z. (2019). Application of Database Applications in School Management Activities. *Journal of Educational Management*, *1*(2), 157–166.

[8] Prasetyo, A., & Azis, M. S. (2021). Design of a Web-Based Medical Record Information System at the Jomin Health Center. *Intercom Journal: Journal of Scientific Publications in the Field of Information and Communication Technology*, *13*(2), 31–38. https://doi.org/10.35969/interkom.v13i2.47

[9] Riany, A. F., & Testiana, G. (2023). Application of Data Mining for Stroke Disease Classification Using Naïve Bayes Algorithm. *Journal of SAINTEKOM*, *13*(1), 42–54. https://doi.org/10.33020/saintekom.v13i1.352

[10] Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, *6*(1), 1. https://doi.org/10.24076/citec.2019v6i1.178

[11] Sidik, A. D. W. M., Himawan Kusumah, I., Suryana, A., Edwinanto, Artiyasa, M., & Pradiftha Junfithrana, A. (2020). Overview of Data Mining Classification Methods. *FIDELITY : Journal of Electrical Engineering*, *2*(2), 34–38. https://doi.org/10.52005/fidelity.v2i2.111

[12] Syakura, A. (2021). Overview of the risk of decubitus in stroke patients in the working area of Dr.H.Slamet Martodirdjo Pamekasan Hospital. *Journal of Health Science and Technology*, *2*(1), 13–18. https://doi.org/10.52234/jstk.v1i1.106

[13] Jung, A. M., Irwan, M., & Nasution, P. (2023). The importance of a Database Security System to protect personal data. *JISKA: Journal of Information Systems and Informatics*, *1*(2), 44. http://jurnal.unidha.ac.id/index.php/jteksis

[14] Utama, T. P., & Haibuan, M. S. (2023). Application of Naïve Bayes and Forward Selection algorithms for stroke disease prediction. *Journal of Teknoinfo*, *17*(2), 351–357.

[15] Wijaya, H. D., & Dwiasnati, S. (2020). Implementation of Data Mining with Naïve Bayes' Algorithm on Drug Sales. *Journal of Informatics*, *7*(1), 1–7. https://doi.org/10.31311/ji.v7i1.6203