p-ISSN 2407-4489
e-ISSN 2528-7303

Development of a Web-Based Application for Predicting Stroke Patient Emergency Levels Using the Naïve Bayes Algorithm

Nurul Abdillah[a], Oki Dahwanu[b], Hamzah Alghifari[c], Niko Akbar[d]

[,a,b,c,d] *Information Systems, Universitas Jambi, Jambi, Indonesia*
email: [a] nurulabdillah@unja.ac.id, [b]okidahwanu@unja.ac.id, [c]hamzah.alghifari@unja.ac.id, [d]nikoakbar@unja.ac.id

Abstract

Stroke is a serious disease that requires prompt and appropriate treatment; therefore, determining the level of patient Emenrgency Levels is critically important. This study aims to develop a web-based application for predicting the Emenrgency Levels level of stroke patients using the Naïve Bayes algorithm as a classification method. The research data were obtained from the medical records of stroke patients at RSUP Dr. M. Djamil Padang during March and April 2025, with a total of 222 data samples. The attributes used in this study include age, gender, address, length of stay, ward class, BPJS insurance membership status, and comorbidities, with Emenrgency Levels status as the class attribute classified into Emenrgency and non-Emenrgency. The application was developed as a web-based system to facilitate easy access for medical personnel in utilizing the prediction system. The experimental results indicate that the Naïve Bayes algorithm achieved an accuracy of 77.48% with an error rate of 22.52%. The findings of this study are expected to assist medical personnel in supporting faster and more objective decision-making regarding the Emenrgency Levels level of stroke patients.

Keywords: Stroke, Emenrgency Levels Prediction, Naïve Bayes, Data Mining, Web Application.

p-ISSN 2407-4489
e-ISSN 2528-7303

INTRODUCTION

Stroke is one of the non-communicable diseases that serves as a leading cause of Emenrgency and disability globally as well as in Indonesia. Based on the results of the 2023 Indonesia Health Survey (SKI), the national prevalence of stroke reached 8.3 per 1,000 population, with the highest prevalence found in the elderly group (≥75 years). This data indicates the high burden of stroke in Indonesian society during that year (Azhar et al., 2022).

In Indonesia, stroke is also categorized as one of the main causes of mortality and significantly contributes to the economic burden on the healthcare system. The healthcare cost burden due to stroke was reported to reach approximately IDR 5.2 trillion in 2023, adding complexity to the management of non-communicable diseases within the national health system (Rahayu & Yamasari, 2024).

Regionally, West Sumatra Province has a higher stroke prevalence than the national average, at 8.8 per 1,000 population according to the 2023 Indonesia Health Survey (SKI). This figure indicates a high incidence of stroke in the West Sumatra region that requires serious attention from local healthcare services (Astiah et al., 2023).

Determining the Emenrgency Levels level of stroke patients is a crucial step in the clinical process, as patients in critical condition require immediate medical intervention to prevent complications, including permanent disability or Emenrgency. However, in daily practice, this determination process is still largely performed manually and depends on the clinical experience of medical personnel, which may lead to inconsistent decision-making.

The advancement of information technology and data mining provides opportunities to optimize the use of medical record data in clinical decision-making. Classification techniques such as the Naïve Bayes algorithm can be used to predict medical conditions based on historical data patterns efficiently and objectively (Abdillah & Dewi, 2024).

The Naïve Bayes algorithm works based on probabilistic principles with an assumption of independence among attributes, making it suitable for handling clinical datasets that contain multiple predictor variables. This algorithm is also known for its low computational complexity and good classification performance in many healthcare cases (Riany et al., 2023).

This study utilizes medical record data of stroke patients at Dr. M. Djamil General Hospital, Padang, from March to April 2025, consisting of 222 samples. The attributes include age, gender, address, length of stay, treatment class, BPJS membership status, and comorbidities. The Emenrgency Levels status of patients is used as the class attribute with two categories: critical and non-critical.
Based on the description above, this study aims to develop a web-based application for predicting the Emenrgency Levels level of stroke patients using the Naïve Bayes algorithm. This application is expected to assist medical personnel in making clinical decisions quickly, objectively, and data-driven to improve the quality of healthcare services in hospitals.

Based on the description above, this study aims to develop a web-based application for predicting the emergency level of stroke patients using the Naïve Bayes algorithm. This application is expected to assist medical personnel in making clinical decisions quickly, objectively, and in a data-driven manner to improve the quality of healthcare services in hospitals. The research problems addressed in this study are how to develop a web-based application for predicting the emergency level of stroke patients using the Naïve Bayes algorithm, and how accurate the algorithm is in classifying patients into emergency and non-emergency categories based on medical record data.

p-ISSN 2407-4489
e-ISSN 2528-7303

METHOD

This study employs a quantitative approach using data mining methods to predict the Emenrgency Levels level of stroke patients. The algorithm used is the Naïve Bayes Classifier as a classification method to determine critical and non-critical categories based on patients' demographic data. The system development is carried out using the Waterfall method, which is a systematic and sequential development model starting from the requirements analysis phase to maintenance, where each stage must be completed before proceeding to the next.

2.1 Data Collection Methods, Research Instruments, and Testing Methods

The data collection method in this study uses a documentation technique by utilizing secondary data in the form of medical records of stroke patients obtained from Dr. M. Djamil General Hospital, Padang, during the period from March to April 2025, consisting of 222 samples. The use of secondary data in quantitative research enables researchers to conduct objective and efficient analysis based on historical data (Saifullah et al., 2024). The collected data include attributes such as age, gender, address, length of stay, treatment class, BPJS membership status, and comorbidities, with patient emergency status as the class attribute categorized into emergency and non-emergency. Prior to analysis, the data undergo a preprocessing stage, including data cleaning and data transformation to ensure the quality and consistency of the dataset (Abdillah & Ihksan, 2022).

In this study, the dataset consisting of 222 stroke patient records was divided into training data and testing data using an 80:20 ratio. A total of 80% of the data (178 records) were used as training data to build the Naïve Bayes classification model, while the remaining 20% (44 records) were used as testing data to evaluate the model's performance. This data partitioning approach is commonly used in classification research to ensure that the model can generalize well to unseen data.

The research instrument used is a structured dataset stored in a system database, along with a web-based application developed to implement the classification algorithm. The algorithm used is the Naïve Bayes Classifier, a probabilistic classification method based on Bayes' Theorem with the assumption of independence among attributes. This algorithm is selected due to its low computational complexity and good performance in medical data classification (Jefri, J., & Fatah, 2025). The application is developed using the Waterfall system development model, which consists of sequential stages: requirements analysis, system design, implementation, testing, and maintenance (Fachri & Rizal, 2024).

The testing method was conducted to evaluate the performance of both the classification model and the developed system. Model validation employed a confusion matrix by comparing prediction results with actual class labels. The evaluation metrics used include accuracy, precision, recall, and error rate. This validation process is important to measure the reliability and effectiveness of the Naïve Bayes algorithm in classifying stroke patient emergency levels into emergency and non-emergency categories (Nugroho et al., 2022). In addition to model evaluation, functional testing of the system was also performed to ensure that all application features operated according to the designed specifications (Srirahayu & Pribadie, 2023).

p-ISSN 2407-4489
e-ISSN 2528-7303

2.2 Research Stages

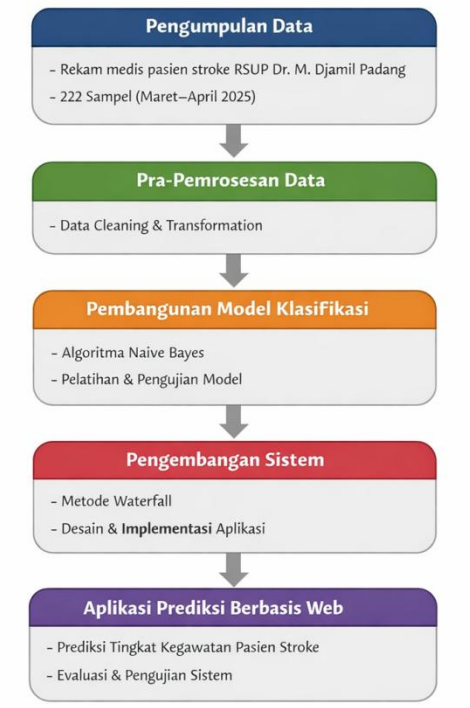


Fig 1. Research Stages

This study was conducted through Emenrgencyal systematic stages, starting from data collection to the implementation of the prediction system.

1. The first stage is data collection, where the researcher obtained secondary data in the form of medical records of stroke patients from Dr. M. Djamil General Hospital, Padang, for the period March to April 2025. A total of 222 samples were collected and had undergone selection to ensure the completeness of the required attributes, such as age, gender, address, length of stay, treatment class, BPJS membership status, comorbidities, and Emenrgency Levels status as the class attribute.
2. The second stage is data preprocessing, which includes data cleaning to remove duplicate or incomplete data, as well as data transformation to ensure compatibility with the algorithm's processing format. This process aims to improve the quality of the dataset so that the classification results become more accurate and valid (Argina, 2020).
3. The third stage is building the classification model using the Naïve Bayes Classifier algorithm. This algorithm operates based on probabilistic principles with the assumption of independence among attributes, enabling efficient prediction of patient Emenrgency Levels status. This stage includes training the model using the processed data and testing the model to evaluate its performance using a confusion matrix, resulting in accuracy and error rate values (Nugraha, F. F., & Juliane, 2022).
4. The fourth stage is the development of a web-based application system using the Waterfall method. This stage includes system requirements analysis, interface and database design, application implementation, functional testing to ensure all features operate according to specifications, and system maintenance. The final result of this stage is a stroke patient Emenrgency Levels prediction application that can be used as a clinical decision support tool for medical personnel (Permana, B. C., & Patwari, 2021).

p-ISSN 2407-4489
e-ISSN 2528-7303

RESULTS AND DISCUSSION

This study uses medical record data of stroke patients from Dr. M. Djamil General Hospital, Padang, during the period from March to April 2025, with a total sample of 222 patients. The attributes used include age, gender, address, length of stay, treatment class, BPJS membership status, and comorbidities, with patient Emenrgency Levels status as the class attribute categorized into critical and non-critical.

This study uses medical record data of stroke patients from Dr. M. Djamil General Hospital, Padang, during the period from March to April 2025, with a total sample of 222 patients. The attributes used include age, gender, address, length of stay, treatment class, BPJS membership status, and comorbidities, with patient emergency level status as the class attribute categorized into emergency and non-emergency. In this study, patients with deceased outcomes were classified into the emergency, while patients with recovered outcomes were classified into the non-emergency category. This categorization was conducted to represent the emergency level of stroke patients based on their final clinical condition.

Table 1. Attribute Data Details

NO		Jumlah Kasus		Hidup	Meninggal
	Total		222	162	60
1.	Umur	Balita	6	6	0
		Anak-Anak	2	0	2
		Remaja	8	8	0
		Dewasa	24	14	10
		Lansia	105	75	30
		Manula	77	59	18
2.	JK	L	120	93	27
		P	102	69	33
3.	Alamat	Dalam Sumbar	186	136	50
		Luar Sumbar	36	26	10
4.	Lama Rawat	<= 3 Hari	36	17	19
		> 3 Hari	186	145	41
5.	Kelas Rawat	ICU/NICU/PICU/CVCU	15	8	7
		HCU	132	83	40
		Kelas I, II, III	82	69	13
		VIP	2	2	0
6.	Status Kepersetaan	BPJS	219	160	59
		Non BPJS	3	2	1
7.	Penyakit Bawaan	Ada	208	152	56
		Tidak Ada	14	10	4

3.1 Naïve Bayes Classification

1. Calculating Class Prior Probabilities

After constructing the attribute data details table, the initial step in the Naïve Bayes calculation is to compute the class prior probabilities for Non-Emenrgency and Emenrgency.

$$P(X = \text{Sembuh}) = \frac{P(X \cap Y)}{P(Y)}$$

$$= \frac{P(Keputusan = Sembuh \cap Total\ Data)}{P(Y = Total\ data)}$$

$$= \frac{162}{222} = 0.730$$

$$P(X = \text{Meninggal}) = \frac{P(X \cap Y)}{P(Y)}$$

$$= \frac{P(Keputusan = Meninggal \cap Total\ Data)}{P(Y = Total\ data)}$$

$$= \frac{60}{222} = 0.270$$

p-ISSN 2407-4489
e-ISSN 2528-7303

2. Calculating Conditional Probabilities

In the conditional calculation, the probabilities of Non-Emenrgency and Emenrgency are computed for each variable, such as age, gender, address, length of stay, treatment class, BPJS membership status, and comorbidities.

3. Determining the Decision Class for a Case

After calculating the prior and conditional probabilities, the next step is to perform the Naïve Bayes computation to predict the decision class for a given case.

Table 2. Data Probability Calculation

Usia	Jenis kelamin	Alamat	Lama Rawat	Kelas Rawatan	Status Kepersetaan	Prediksi Keputusan
Lansia	L	Dalam Sumbar	>3 hari	ICU/NICU/PICU/CVCU	BPJS	?

Based on the data above, a case is identified with the following characteristics: elderly age, gender = male, address = within West Sumatra, length of stay > 3 days, treatment class = ICU/NICU/PICU/CVCU, and BPJS membership status. The next step is to calculate the probabilities of Non-Emenrgency and Emenrgency to determine the higher probability.

a. Probability of Recovery

The calculation of recovery probability in the Naïve Bayes algorithm aims to classify patients based on their likelihood of recovery using available medical data. By calculating this probability, the algorithm can provide predictions regarding the patient's health status.

P(H|X) x P(X) = ((P(Usia = Lansia | Sembuh) x P(Jenis Kelamin = L| Sembuh) x P(Alamat = Dalam Sumbar | Sembuh) x P(Lama Rawat = >3 Hari | Sembuh) x P(Kelas Rawatan = ICU | Sembuh) x P(Status Kepersetaan = BPJS | Sembuh) P(Penyakit Bawaan = Ada | sembuh)))))

P(H|X)x P(X) = 0,463 × 0,574 × 0,839 × 0,895 × 0,494 × 0,938 × 0,730

PP(H|X) x P(X) = 0,0067

b. Probability of Emenrgency

The calculation of Emenrgency probability in the Naïve Bayes algorithm aims to classify patients based on their likelihood of Emenrgency using available medical data. By calculating this probability, the algorithm can predict the patient's health outcome.

P(H|X) x P(X) = ((P(Usia = Lansia | Meninggal) x P(Jenis Kelamin = L| Meninggal) x P(Alamat = Dalam Sumbar | Meninggal) x P(Lama Rawat = >3 Hari | Meninggal) x P(Kelas Rawatan = ICU | Meninggal) x P(Status Kepersetaan = BPJS | Meninggal) x P(Penyakit Bawaan = Ada | meninggal))))))

P(H|X) x P(X) = 0,500 × 0,450 × 0,833 × 0,683 × 0,117 × 0,983 × 0,933 × 0,270

PP(H|X) x P(X) = 0,0037

p-ISSN 2407-4489
e-ISSN 2528-7303

Based on the calculation of the probabilities for recovery and Emenrgency, the predicted decision is Non-Emenrgency, because the probability of recovery is higher than the probability of Emenrgency.

The probability calculation shown above (Table 4.10) represents one example of the training data calculation from Table 4.8. The overall Naïve Bayes prediction results for all data are presented in Table 2.

Table 3. Probability of Non Emenrgency and Emenrgency

No	Probabilitas		Probabilitas Perhitungan Naïve Bayes
	Sembuh	Meninggal	
1	0,006669014	0,003707802	Hidup
2	0,069191019	0,021187437	Hidup
3	0,000395962	0,000857177	Meninggal
4	0,000101401	0,000166416	Meninggal
5	0,040383747	0,015537454	Hidup
6	0,002052683	0,000459061	Hidup
7	0,013227695	0,004237487	Hidup
.	.	.	.
.	.	.	.
216	0,033572031	0,005049673	Hidup
217	0,054430268	0,012712462	Hidup
218	0,033572031	0,005049673	Hidup
219	0,000230906	0	Hidup
220	0,005003429	0,003900154	Hidup
221	0,000122733	0	Hidup
222	0,045249259	0,00413155	Hidup

Based on Table 3, the Probability of Non Emenrgency and Emenrgency table is obtained from manual calculations using the Naïve Bayes algorithm formula. These calculations produce the probability values for Non Emenrgency and Emenrgency, which are then compared to determine the higher probability. The highest probability value is used as the prediction result of the Naïve Bayes algorithm in predicting the discharge condition of stroke patients.

4. Naïve Bayes Prediction Results
   After performing the Naïve Bayes probability calculations, the probabilities of recovery and Emenrgency are obtained. By comparing these probabilities, the prediction accuracy of recovery based on the training data can be determined. Based on the probabilities generated from the Naïve Bayes calculations, the prediction results for stroke patient data are presented as shown in Table 4.

p-ISSN 2407-4489
e-ISSN 2528-7303

Table 4. Naïve Bayes Prediction Results

No	Usia	Jenis kelamin	Alamat	Lama Rawat	Kelas Rawatan	Status Kepersetaan	Penyakit Bawaan	Hasil Prediksi Naïve Bayes
1	Lansia	L	Dalam Sumbar	>3 hari	ICU/NICU/PICU/CVCU	BPJS	Ada	Hidup
2	Lansia	L	Dalam Sumbar	>3 hari	HCU	BPJS	Ada	Hidup
3	Lansia	P	Dalam Sumbar	<= 3 hari	HCU	BPJS	Tidak Ada	Meninggal
4	Lansia	L	Dalam Sumbar	<= 3 hari	HCU	Non BPJS	Ada	Meninggal
5	Manula	P	Dalam Sumbar	>3 hari	HCU	BPJS	Ada	Hidup
6	Dewasa	L	Luar Sumbar	>3 hari	Kelas I, II, III	BPJS	Ada	Hidup
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
216	Manula	P	Dalam Sumbar	>3 hari	Kelas I, II, III	BPJS	Ada	Hidup
217	Manula	L	Dalam Sumbar	>3 hari	HCU	BPJS	Ada	Hidup
218	Manula	P	Dalam Sumbar	>3 hari	Kelas I, II, III	BPJS	Ada	Hidup
219	Dewasa	P	Dalam Sumbar	>3 hari	VIP	BPJS	Ada	Hidup
220	Lansia	P	Dalam Sumbar	<= 3 hari	Kelas I, II, III	BPJS	Ada	Hidup
221	Remaja	P	Luar Sumbar	<= 3 hari	HCU	BPJS	Ada	Hidup
222	Manula	L	Dalam Sumbar	>3 hari	Kelas I, II, III	BPJS	Ada	Hidup

5. Calculating Naïve Bayes Accuracy



Fig 1. Naïve Bayes Accuracy

The results of this study show an accuracy value of **77.48%**, as 172 out of 222 data instances were correctly predicted. Meanwhile, the error rate is **22.52%**, with 50 out of 222 data instances incorrectly classified.

The accuracy result of 77.48% indicates that the Naïve Bayes algorithm has a reasonably good capability in classifying the emergency level of stroke patients based on demographic and clinical attributes. The algorithm performs efficiently because it can process multiple predictor variables with relatively low computational complexity. However, the accuracy value also indicates that Emenrgencyal data instances were still misclassified, which

p-ISSN 2407-4489
e-ISSN 2528-7303

may be influenced by the limited number of samples, the imbalance of data distribution, and the assumption of attribute independence in the Naïve Bayes method.

Compared with previous studies in the healthcare domain, the obtained accuracy is considered acceptable for a probabilistic classification model using secondary medical record data. Previous research conducted by Riany et al. (2023) also showed that the Naïve Bayes algorithm achieved good performance in medical classification problems due to its simplicity and fast computation process. In this study, the developed web-based system can support medical personnel in identifying stroke patient emergency levels more quickly and objectively, thereby assisting early clinical decision-making processes in hospitals.

3.2 System Development

In the system development stage, the CodeIgniter framework is used. CodeIgniter is a PHP-based MVC framework that is lightweight, fast, and easy to use for developing web applications with a clean and well-structured code architecture (Abdillah, Nurul., Dahwanu, Oki., Alghifari, Hamzah., Akbar, 2025)

Home Page

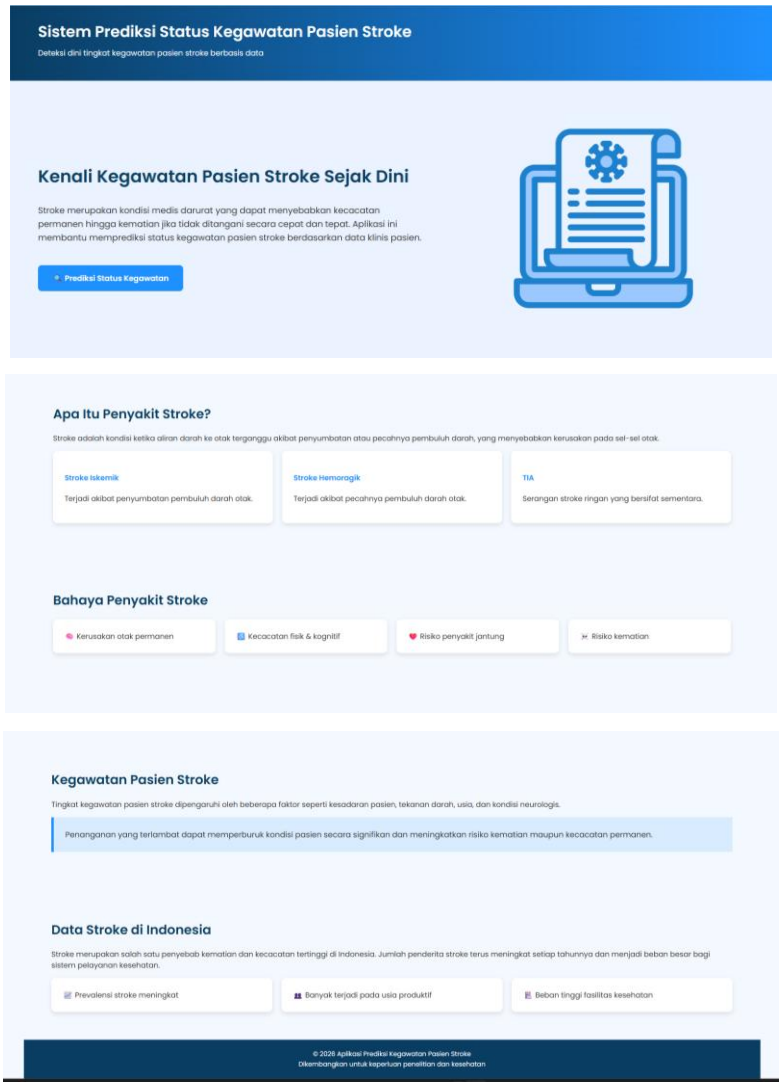


Fig 2. Home Page

p-ISSN 2407-4489
e-ISSN 2528-7303

Patient Data Input Page

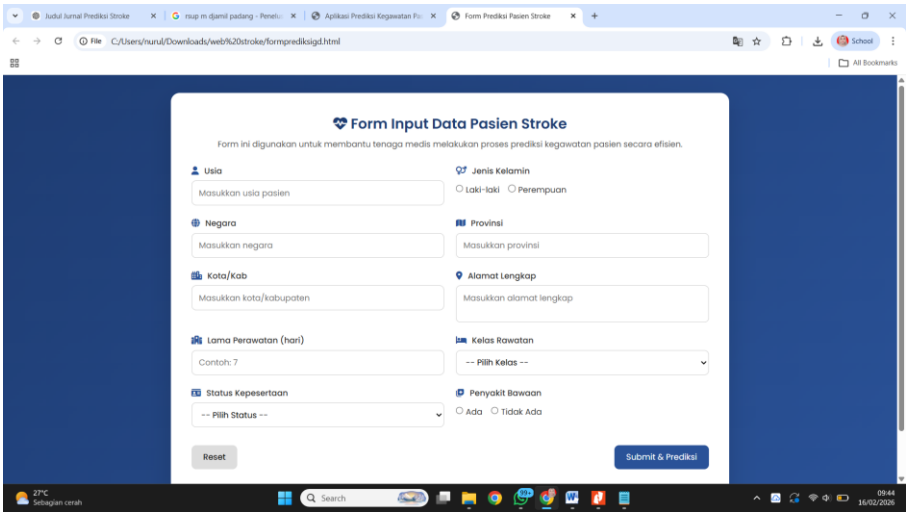


Fig 3. Input Data Page

Prediction Result Page

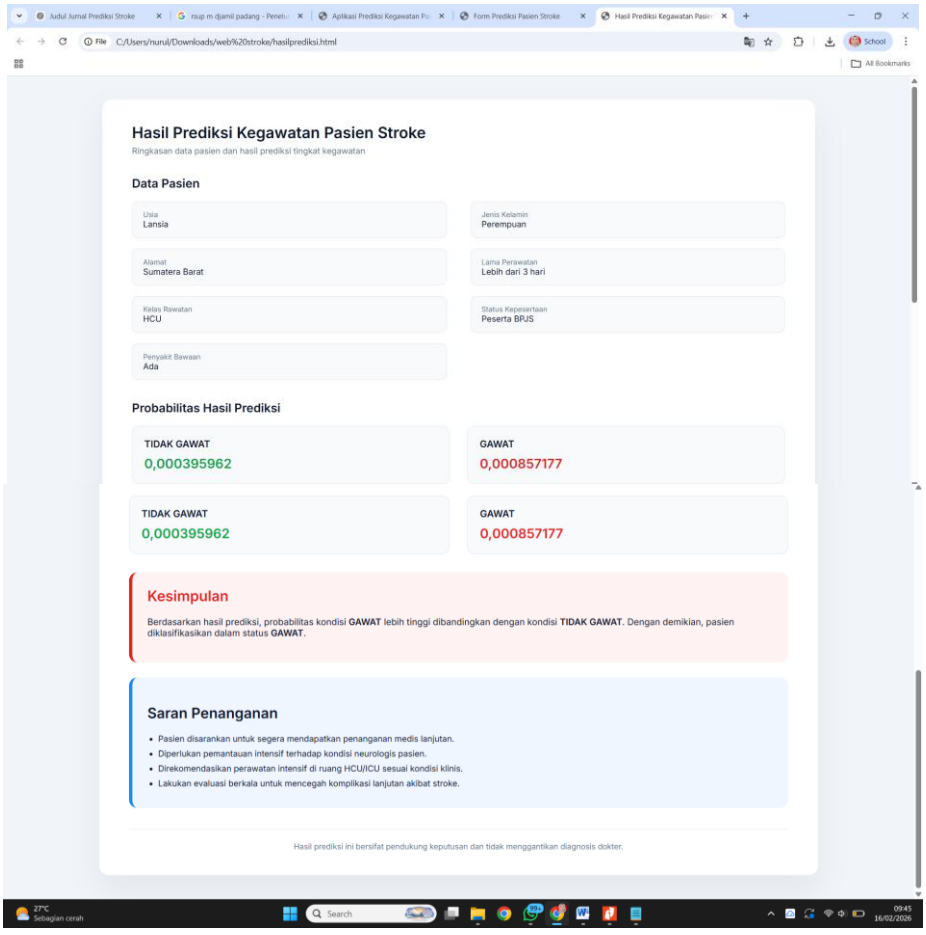


Fig 4. Prediction Result Page

CONCLUSIONS

The web-based stroke patient Emenrgency Levels prediction application developed using the Naïve Bayes algorithm and the CodeIgniter framework was able to provide fairly good prediction performance, achieving an accuracy of 77.48%. The results indicate that attributes such as age, length

of stay, and comorbidities significantly influence the classification of stroke patients into Emenrgency and non-Emenrgency categories. The developed system can assist medical personnel in making clinical decisions more quickly, objectively, and consistently, particularly in supporting the early identification of Emenrgency stroke conditions in hospital environments.

However, this study has Emenrgencyal limitations. The dataset used was limited to 222 medical record samples collected from a single hospital within a relatively short observation period, which may affect the generalizability of the model. In addition, the Naïve Bayes algorithm assumes independence among attributes, whereas clinical variables in real-world medical cases may be interrelated. Therefore, future studies are recommended to use larger and more diverse datasets, apply feature selection techniques or hybrid classification algorithms, and integrate the system with electronic medical records to improve prediction accuracy and system functionality..

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Dr. M. Djamil General Hospital, Padang, Universitas Jambi, and Universitas Trunodjoyo Madura for providing the facilities and resources for this research and the publication of this article. We also extend our appreciation to all parties who have contributed to this work, many of whom cannot be mentioned individually.

REFERENCES

Abdillah, Nurul., Dahwanu, Oki., Alghifari, Hamzah., Akbar, N. . (2025). Perancangan Sistem Informasi Promosi dan Penjualan Berbasis Web Pada Unit Produksi SMKN 1 Ampek Angkek Menggunakan Framework CodeIgniter. *Informatika*, *12*(3).

Abdillah, N., & Dewi, A. (2024). Pengembangan Aplikasi Berbasis Web Dengan Framework Codeigniter untuk Prediksi Status Kegawatan Pasien Tuberculosis Menggunakan Algoritma NBC Di Kota Padang. *The Indonesian Journal of Computer Science*, *13*(6), 10013–10023.

Abdillah, N., & Ihksan, M. (2022). Application Of The C4.5 Algorithm For Classification Of Medical Record Data At M . Djamil Hospital Based On The International Disease Code. *Mantik*, *6*(36), 576–581.

Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, *1*(2), 29–33.

Astiah, A. A., Galindra, Y., & Sananda, G. (2023). Hubungan Pengetahuan Pasien Pasca Stroke Dengan Tingkat Kepatuhan Menjalani Fisioterapi Di Rsu M Natsir Solok Sumatera Barat Periode Januari–Juni 2022. *Zona Kedokteran*, *13*(2), 416–423.

Azhar, Y., Firdausy, A. K., & Amelia, P. J. (2022). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *Sintech Journal*, *5*(2), 191–197.

Fachri, B., & Rizal, C. (2024). Penerapan Metode Waterfall Dalam Perancangan Sistem Informasi Merdeka Belajar Kampus Merdeka Berbasis Web. *JUKTISI*, *2*(3), 591–597.

Jefri, J., & Fatah, Z. (2025). Klasifikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa. *Jurnal Ilmiah Multidisiplin Ilmu*, *2*(1), 29–37.

Nugraha, F. F., & Juliane, C. (2022). Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika Dan Sistem Informasi*, *9*(4).

Nugroho, B. I., Ma'arif, Z., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Metode Klasifikasi Untuk Menganalisa Penyalahgunaan Sosial Media: Array. *Jurnal Sistem Informasi Dan Teknologi Peradaban (JSITP)*.

Permana, B. C., & Patwari, I. D. (2021). Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes. *Infotek : Jurnal Informatika Dan Teknologi*, *4*(1), 63–69.

Rahayu, S., & Yamasari, Y. (2024). Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM). *Journal of Informatics and Computer Science*, *05*, 440–446.

Riany, A. F., Testiana, G., Informasi, S. S., & Palembang, K. (2023). Penerapan Data Mining untuk

p-ISSN 2407-4489
e-ISSN 2528-7303

Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes. *Saintekom*, *9*, 42–54.

Saifullah, Y. Y., Rachman, M. E., Limoa, L. T., & Hamado, N. (2024). Literature Review: Hubungan Hipertensi dengan Kejadian Stroke Iskemik dan Stroke Hemoragik KYayan. *Fakumi Medical Journal*, *04*(10).

Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, *14*(April).