Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Naïve bayes algorithm for early diagnosis of non-communicable diseases

Nuzul Ramadhan[1]*, Rakhmat Kurniawan Ramli[2]

[1,2]*Department of Computer Science, Universitas Islam Negeri Sumatera Utara, Indonesia*

Article Info

Article history:

Received Sep 27, 2025
Revised Sep 2, 2025
Accepted Sep 13, 2025

Keywords:

Community Health;
Early Diagnosis;
Expert System;
Naïve Bayes Algorithm;
Non-Communicable Diseases.

ABSTRACT

Non-communicable diseases such as heart disease, diabetes mellitus, hypertension, stroke, asthma, rheumatism, and GRED are still the main causes of illness and death in Indonesia. This problem is more serious in rural areas with limited health services, such as Lubuk Palas Village, Asahan Regency, which faces obstacles in distance, road infrastructure, and a limited number of medical personnel, so early diagnosis is often neglected. This research aims to apply the Naïve Bayes method in a non-communicable disease diagnosis expert system and develop web and mobile-based applications to support the community and medical personnel in early detection. The research method combines primary data from observations and interviews with health workers and secondary data from medical literature. Each symptom is given a probability weight of 0.00–1.00 according to medical consultation, then processed using the Naïve Bayes algorithm with two approaches, namely direct calculation and gradual filtering. The results show that the system produces a posterior probability of 99.32% in the heart disease scenario with typical symptoms and 90.00% in the stroke scenario with partial symptoms. The findings of this research are that the application of two Naïve Bayes inference pathways is proven effective in producing an initial diagnosis that is adaptive to variations in symptoms, relevant for rural conditions with limited health services, and capable of providing fast, practical, and widely accessible medical decision support.

Corresponding Author:

Nuzul Ramadhan
Department of Computer Science
Universitas Islam Negeri Sumatera Utara,
Jl. William Iskandar Ps. V, Medan Estate, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara, 20371, Indonesia
Email: nuzulramadhan53@gmail.com

1. Introduction

Health is a vital foundation for the progress of a nation, encompassing physical, mental, and social well-being that enables every individual to live a productive life, as mandated in Health Law No. 36 of 2009 (P. Syahputra & Kurniawan, 2024). Elkomy & Jackson (2024) also defined health comprehensively, beyond the mere absence of disease. Nationally, the Healthy Indonesia program with a family approach has prioritized the prevention and control of non-communicable diseases, including hypertension and anti-smoking campaigns (Fiqriansyah et al., 2022; Wahidin et al., 2023). However, the reality in Lubuk Palas Village, Silau Laut Sub-district, Asahan District, North Sumatra, shows significant challenges in accessing health services. Poor road infrastructure, especially during the rainy season, makes traveling to the nearest puskesmas difficult. This condition encourages the community to rely on on-call medical personnel or "Mantri". However, Mantri services are often constrained by long travel times (15-40 minutes), long examination durations (10-15 minutes), and limited mobility and medicine supplies (Asyim & Yulianto, 2022). Consequently, disease management is often delayed, which in turn increases the risk of complications Chairina & Candrasa (2022) and Tugiman et al. (2022). This situation is exacerbated by low public awareness; data from the Ministry of Health shows that 7 out of 10 people with non-communicable diseases are unaware of their condition until complications worsen, and of those who are detected, only one-third seek regular treatment (Ehatisham-ul-Haq et al., 2021). Therefore, this crucial condition underscores the urgent need for a system that can help identify diseases early, enabling Mantri to more quickly prepare medicines and determine the appropriate course of action, so that treatment can be expedited and the risk of acute or chronic illness can be minimized (Tugiman et al., 2022).

Based on the problems described, this research focuses on two main aspects. The first aspect is the application of the Naïve Bayes method in an expert system for the diagnosis of non-communicable diseases in the community in Lubuk Palas Village (Riany & Testiana, 2023; Setiawan et al., 2024; S. Syahputra & Hasibuan, 2024). The second aspect is measuring the accuracy of the Naïve Bayes method in predicting the type of non-communicable disease based on available symptom data (Fadhillah et al., 2023; Islam et al., 2022). In line with this focus, this research aims to apply the Naïve Bayes method in an expert system for diagnosing non-communicable diseases in the Lubuk Palas Village community, measuring the accuracy of the predictions produced, and developing an artificial intelligence-based diagnosis system that can help medical personnel and the community in early detection of disease (Dwiramadhan et al., 2022; Pohan & Chairunisah, 2024; Yuliza, 2022).

This research implements the Naïve Bayes Algorithm, a statistical and probability-based classification method based on Bayes' Theorem (Aprilia, 2024; Suparyanto & Rosad, 2022; Yuliza, 2022). This algorithm is known for its simplicity, speed, and high accuracy in predicting probabilities based on previous case data. In fact, some studies show that Naïve Bayes can achieve an accuracy of up to 99.51% and has a relatively higher prediction value compared to other algorithms (Aprilia, 2024). In the realm of diagnosis of non-communicable diseases, a number of studies have evaluated alternative methods. Research by Saptawan et al. (2024) applied Random Forest for the prediction of non-communicable diseases at the Sekura Health Center, but the system only produced a total accuracy of 27.64% so it is not yet feasible for use directly in the field. Research by Pratama et al. (2022) used K-Nearest Neighbor (KNN) for the classification of heart failure disease with the best accuracy of 70.65% at k=9, accompanied by a precision of 75.00% and a recall of 70.73%. Meanwhile, Pinem & Putra (2025) applied a linear kernel Support Vector Machine (SVM) in diabetes prediction, resulting in an accuracy of 70%, but the performance in the positive class was still low (precision 0.57; recall 0.63; F1 0.60). These findings suggest that although various algorithms have been tried, the results are often limited in terms of both accuracy and disease coverage. These approaches are integrated into an expert system framework, which is a technological product that models expert thinking to solve problems without direct interaction (Pohan & Chairunisah, 2024). The expert system will involve a knowledge base (symptoms and weights), an inference engine (Naïve Bayes algorithm), and a user interface. Several previous studies have shown the application of Naïve Bayes in the medical field. For example, Oskar et al. (2024) used Naïve Bayes to manage heart disease data and obtained an accuracy of 79.92%, although they noted a weakness in calculating classification accuracy internally. Meanwhile, Jefi et al. (2022) applied Naïve Bayes to predict gastric diseases and achieved 75% accuracy with an Area Under the Curve (AUC) value of 0.852, showing good classification quality, but still limited to one type of disease. These results suggest that Naïve Bayes has the potential to be used more widely, but previous studies have generally focused on a single disease with structured data, rather than on multi-disease and symptom-based scenarios in the field.

Different from previous studies, this study is designed to produce an early diagnosis of non-communicable diseases in a multi-disease manner at the community level, focusing on rural conditions that have limited access to health services. Symptom data was obtained through direct consultation with medical personnel in Lubuk Palas Village and community interviews, then converted into a probabilistic weight of 0.00–1.00 which represents the degree of association of symptoms with disease (Armansyah & Ramli, 2022). Furthermore, two Naïve Bayes inference approaches were used, namely direct calculation and gradual screening, to make the diagnosis more accurate even if the patient reported only partial symptoms. In this way, the research offers different contributions by building a Naïve Bayes-based expert system that is relevant to the context of rural communities, presenting symptom-based multi-disease diagnosis mechanisms rather than relying solely on medical record data, combining local medical knowledge with artificial intelligence, and providing easily accessible web and mobile applications. The purpose of this research is to provide a more practical, fast, and adaptive diagnostic system, so that it can improve the early detection of non-communicable diseases, strengthen the readiness of medical personnel, and at the same time expand the academic literature on the application of probabilistic-based expert systems in the field of public health.

2. Method

This research addresses the challenge of limited healthcare access in Lubuk Palas Village, Silau Laut Subdistrict, Asahan Regency, North Sumatra, where early diagnosis of non-communicable diseases (NCDs) is hindered by poor infrastructure and scarce medical resources. A technology-based system is proposed to enable early symptom recognition and support medical personnel in delivering healthcare services.

1. Data Collection
   a. Primary Data, primary data were collected through direct observations and interviews in Lubuk Palas Village. Observations identified prevalent symptoms of NCDs among the community. Interviews with two qualified medical personnel, Surya Darma, S.Kep., Ns, and Sigit Sapmono, S.Kep., CWCCA, verified symptom relevance and assigned probability weights for seven NCDs: heart disease, diabetes mellitus, hypertension, stroke, asthma, rheumatism, and gastroesophageal reflux disease (GERD). Community interviews further elucidated specific complaints related to these diseases.
   b. Secondary Data, were sourced from scientific literature, including journals, health reports, medical books, and open datasets from the World Health Organization and the Indonesian Ministry of Health. These data enriched the knowledge base and were validated by local medical personnel to ensure contextual relevance.
2. Data Analysis, data were organized by grouping symptoms and their relationships with the seven NCDs. Preprocessing involved cleaning data (removing duplicates, addressing missing values) and formatting it into a dataset with cases as rows and symptoms as columns. Symptom weights, ranging from 0.00 (no association) to near 1.00 (strong association, avoiding 1.00 to prevent false positives), were assigned based on medical consultations. Symptom combination analysis enhanced diagnostic accuracy by evaluating patterns among symptoms, resulting in a dataset ready for modeling.
3. Implementation of Naïve Bayes Method.

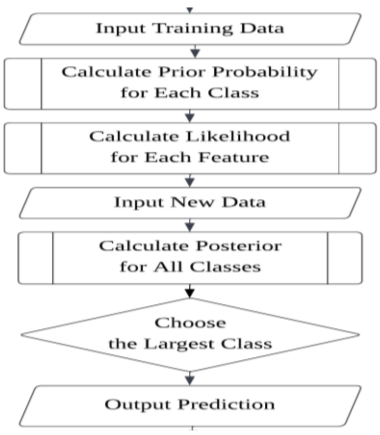

Figure 1. Flowchart of Naïve Bayes Algorithm

Naïve Bayes algorithm is used as a classification method because it is simple, fast, and has high accuracy in diagnosing non-communicable diseases (Sano et al., 2023). The basic principle refers to

Bayes' Theorem with the formula:

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

These stages include input of training data, calculation of prior probability for each disease, calculation of likelihood for each symptom, input of new data, posterior calculation, selection of the class with the highest probability, and prediction output. The probability calculation in this algorithm uses three main formulas, namely:

a. Direct Approach, computes probabilities for all seven diseases simultaneously based on input symptoms, selecting the disease with the highest posterior probability as the primary diagnosis. This approach is efficient for cases with distinct, disease-specific symptoms.

$$P(D_i) = \prod_{j=1}^{n} w_{ij} \times P(C_i) \quad (2)$$

b. Filtering Approach, calculates initial probabilities using basic symptoms, identifies the three diseases with the highest probabilities, prompts the user for additional (advanced) symptoms, and recalculates probabilities to refine the diagnosis. This method enhances accuracy when symptoms are partial or overlapping, adapting to real-world clinical variability.

$$P(D_k^{final}) = \prod_{j=1}^{n} w_{kj}^{(basic+advanced)} \times P(C_k) \quad (3)$$

The process involves inputting training data, calculating prior probabilities and likelihoods, processing new symptom data, computing posterior probabilities, and selecting the most likely disease

4. System Planning

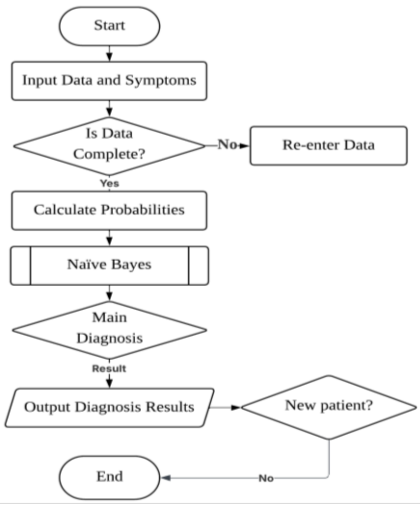

Figure 2. Flowchart of Non-Communicable Disease Diagnosis System

The NCD diagnosis system was designed as web and mobile applications, prioritizing mobile accessibility for rural users (Elkomy & Jackson, 2024; Ether & Saif-Ur-Rahman, 2021). The workflow includes inputting personal data and symptoms, verifying data completeness, applying the Naïve Bayes algorithm, determining the primary diagnosis, and displaying results. Users input data to receive initial diagnoses, while admins (medical personnel) manage disease data, symptoms, weights, and patient histories. The database, built with PHP and MySQL, includes tables for diseases, symptoms, weights, diagnoses, and users. The interface, developed using HTML and CSS, is simple and mobile-friendly (Dwiramadhan et al., 2022; Mitiche et al., 2021; Prahasti, 2022).

5. System Testing, ensured alignment with research objectives and usability in Lubuk Palas Village. It focused on: (1) verifying accurate diagnosis based on user-input symptoms, (2) confirming functionality, ease of use, and mobile accessibility for community and medical personnel, and (3) identifying strengths (e.g., accessibility, early diagnosis support) and areas for improvement (Fourniawan et al., 2024; Krosman et al., 2021).
6. System Implementati, the system was implemented as a web and mobile application, integrating the Naïve Bayes algorithm, database, and user interface, with a focus on mobile usability to address rural

healthcare challenges.

3. Results and Discussions

The dataset used in this study consists of 30 symptoms that are divided into two categories, namely basic symptoms and advanced symptoms. Each symptom is assigned a probability weight that reflects the degree of association with seven types of non-communicable diseases, namely heart disease, diabetes mellitus, hypertension, stroke, asthma, rheumatism, and GRED. The weight values are in the range of 0.00-1.00, where the higher the weight value indicates the stronger the association between the symptom and a particular disease.

Table 1. Dataset and Symptoms

Symptom ID	Symptom Name	Symptom Type	Heart Disease Weight	Diabetes Weight	Hypertension Weight	Stroke Weight	Asthma Weight	Rheumatism Weight	GRED Weight
G1	Chest pain	Basic	0.60	0.10	0.30	0.10	0.20	0.05	0.50
G2	Shortness of breath	Basic	0.60	0.05	0.15	0.10	0.65	0.05	0.10
G3	Frequent thirst	Advanced	0.01	0.85	0.01	0.01	0.00	0.02	0.01
G4	Frequent urination at night	Basic	0.02	0.98	0.01	0.03	0.01	0.01	0.01
G5	Fatigue	Basic	0.10	0.10	0.10	0.10	0.10	0.10	0.10
G6	Nausea	Advanced	0.30	0.20	0.10	0.10	0.05	0.01	0.80
G7	Burning sensation in the chest	Basic	0.35	0.05	0.10	0.05	0.20	0.01	0.60
G8	Palpitations	Basic	0.70	0.15	0.40	0.05	0.20	0.01	0.40
G9	Dizziness	Advanced	0.40	0.10	0.80	0.80	0.10	0.05	0.30
G10	Blurred vision	Basic	0.10	0.80	0.40	0.50	0.10	0.02	0.10
G11	Slurred speech	Advanced	0.02	0.03	0.15	0.90	0.04	0.04	0.04
G12	Facial drooping	Basic	0.01	0.15	0.15	0.90	0.05	0.05	0.05
G13	Difficulty lifting arm	Basic	0.10	0.15	0.10	0.70	0.05	0.60	0.05
G14	Joint stiffness	Basic	0.01	0.10	0.05	0.05	0.01	0.80	0.05
G15	Swollen joints	Basic	0.02	0.10	0.05	0.04	0.03	0.90	0.01
G16	Wheezing	Basic	0.25	0.05	0.04	0.06	0.70	0.03	0.20
G17	Dry cough	Basic	0.50	0.15	0.40	0.05	0.80	0.04	0.60
G18	Noisy breathing	Basic	0.40	0.10	0.05	0.02	0.80	0.01	0.15
G19	Stomach pain	Basic	0.15	0.20	0.05	0.01	0.02	0.03	0.75
G20	Frequent belching	Advanced	0.15	0.20	0.05	0.01	0.02	0.03	0.85
G21	Neck heaviness/tightness	Basic	0.10	0.10	0.70	0.05	0.01	0.01	0.01
G22	Pain radiating to left arm or jaw	Basic	0.70	0.05	0.10	0.02	0.01	0.01	0.05
G23	Tingling/numbness in hands or feet	Advanced	0.30	0.80	0.10	0.35	0.01	0.20	0.01
G24	Chest tightness/pressure	Basic	0.40	0.05	0.10	0.03	0.85	0.01	0.10
G25	Redness & warmth in joints	Advanced	0.01	0.05	0.01	0.01	0.01	0.95	0.01
G26	Sour taste in mouth	Advanced	0.05	0.05	0.03	0.02	0.05	0.01	0.90
G27	Cold sweat	Advanced	0.90	0.15	0.10	0.05	0.01	0.02	0.05
G28	Ringing in ears (tinnitus)	Advanced	0.05	0.05	0.85	0.60	0.01	0.01	0.02
G29	Easily fatigued	Basic	0.60	0.30	0.40	0.10	0.10	0.05	0.01
G30	Rapid heartbeat at rest	Advanced	0.60	0.10	0.90	0.10	0.15	0.01	0.01

This dataset became the basis for posterior calculations with Bernoulli Naïve Bayes. Manual calculations are displayed to verify the consistency of the weight and output of the system. Direct example: patients report Fatigue (G5), Dizziness (G9), and Tinnitus (G28). Based on the formula (likelihood × prior, with smoothing), the posterior of each class is calculated and then normalized. Detailed results are shown in Table 2.

Table 2. Calculating Weight

Disease	Calculation	Result
Heart	0.10× 0.40× 0.05× 0.1428	0.00028
Diabetes	0.10× 0.10× 0.05× 0.1428	0.00007
Hypertension	0.10× 0.80× 0.85× 0.1428	0.00971
Stroke	0.10× 0.80× 0.60× 0.1428	0.00685
Asthma	0.10× 0.10× 0.01× 0.1428	0.00001
Rheumatism	0.10× 0.05× 0.01× 0.1428	0.00001
GERD	0.10× 0.30× 0.02× 0.1428	0.00008
	Total	0.01701

Each value is then normalized using $\text{Disease Percentage} = \frac{\text{P (Disease)}}{\text{Total of All Values}} \times 100\%$, so that the percentage probability for each disease is obtained as follows:

Table 3. Normalization of Disease Probabilities

Disease	Formula	Result (%)
Heart Disease	0.00028 ÷ 0.01701 × 100%	1.65%
Diabetes	0.00007 ÷ 0.01701 × 100%	0.41%
Hypertension	0.00971 ÷ 0.01701 × 100%	57.08%
Stroke	0.00685 ÷ 0.01701 × 100%	40.27%
Asthma	0.00001 ÷ 0.01701 × 100%	0.06%
Rheumatism	0.00001 ÷ 0.01701 × 100%	0.06%
GERD	0.00008 ÷ 0.01701 × 100%	0.47%

Based on the normalization results, the system selects the disease with the highest percentage. In this example, hypertension reached 57.08% followed by stroke at 40.27%. Thus, the patient is indicated to have hypertension. This result can be used as a basis for decision-making, although it is not an absolute medical diagnosis. Furthermore, if the patient enters four symptoms, such as "chest pain" (G1), "palpitations" (G8), "pain radiating to the left arm or jaw" (G22), and "cold sweat" (G27), the calculation is done in the same way. normalization results show that heart disease dominates with a value of 99.32%. This diagnosis is more convincing because the number of relevant symptoms is increasing. Therefore, the system recommends that patients enter more than two symptoms so that the accuracy of diagnosis increases and the results obtained are more reliable as decision support.

In the filtering approach, for example, the symptoms selected are "Fatigue (G5), Blurred Vision (G10), Difficulty lifting arm (G13), and Easily fatigued (G29)." Based on the weight of each symptom, an initial calculation is carried out using $\text{P(Disease)} = (\text{Symptom Weight 1}) \times (\text{Symptom Weight 2}) \times (\text{Symptom Weight 3}) \times \ldots \times (\text{Prior Probability})$
, the results are as follows:

Table 4. Symptom Weight Filtering Approach

Disease	Calculation	Result
Heart	0.10× 0.10× 0.10× 0.60× 0.1428	0.00001
Diabetes	0.10× 0.80× 0.15× 0.30× 0.1428	0.00051
Hypertension	0.10× 0.40× 0.10× 0.40× 0.1428	0.00023
Stroke	0.10× 0.50× 0.70× 0.10× 0.1428	0.00050
Asthma	0.10× 0.10× 0.05× 0.10× 0.1428	0.00001
Rheumatism	0.10× 0.02× 0.60× 0.05× 0.1428	0.00001
GERD	0.10× 0.10× 0.05× 0.01× 0.1428	0.00001
	Total	0.00128

The first stage of normalization resulted in three diseases with the highest values, namely Diabetes 39.84%, Stroke 39.06%, and Hypertension 17.97%. In the second stage, the patient added further symptoms such as “Dizziness (G9) and Slurred speech (G11).” The weights of these additional symptoms are combined with the basic symptoms that have been selected, hence the results:

Table 5. Advanced Symptoms

Symptoms	Diabetes	Stroke	Hypertension
G5	0.10	0.10	0.10
G10	0.80	0.50	0.40
G13	0.15	0.70	0.10
G29	0.30	0.10	0.40
G9	0.10	0.80	0.80
G11	0.03	0.90	0.15

The second stage of calculation is done by multiplying all symptom weights (basic + advanced) for each disease. The calculation results are as follows:

Table 6. Stage 2 Calculation

Disease	Calculation	Value
Diabetes	0.10 × 0.80 × 0.15 × 0.30 × 0.10 × 0.03 × 0.1428	0.00001
Stroke	0.10 × 0.50 × 0.70 × 0.10 × 0.80 × 0.90 × 0.1428	0.00036
Hypertension	0.10 × 0.40 × 0.10 × 0.40 × 0.80 × 0.15 × 0.1428	0.00003
	Total	0.00040

The second stage of normalization results in the following percentage probabilities: Diabetes 2.50%, Hypertension 7.50%, and Stroke 90.00%. Thus, the system diagnosed the patient as having the highest probability of suffering from stroke with a probability level of 90.00%. This result shows that the more relevant symptoms are entered, the higher the confidence level of the diagnosis generated by the system.

The implementation results show that the non-communicable disease diagnosis system has been successfully realized in the form of web-based and mobile applications with a simple interface designed to facilitate use by ordinary people. The initial display of the system facilitates filling in the patient's personal data, such as name, gender, and age. After that, the patient can choose the basic symptoms and advanced symptoms they experience through a structured input page. The diagnosis process produces an output in the form of a percentage probability for each detected disease. These results are displayed clearly on the screen, along with recommendations for initial actions to guide the user before obtaining further medical examination (Figure 3 & Figure 4). The consistency between the system's diagnosis results and manual calculations has also been verified, shown through the correspondence of the resulting percentage values (Figure 5).

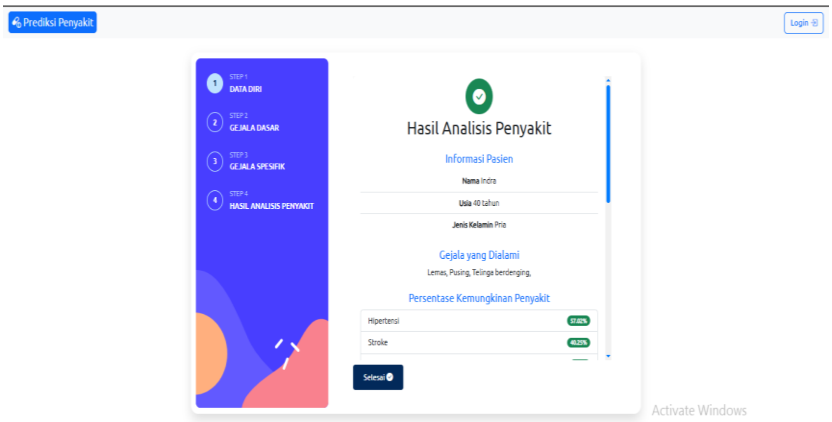

Figure 3. Diagnosis Result (1)

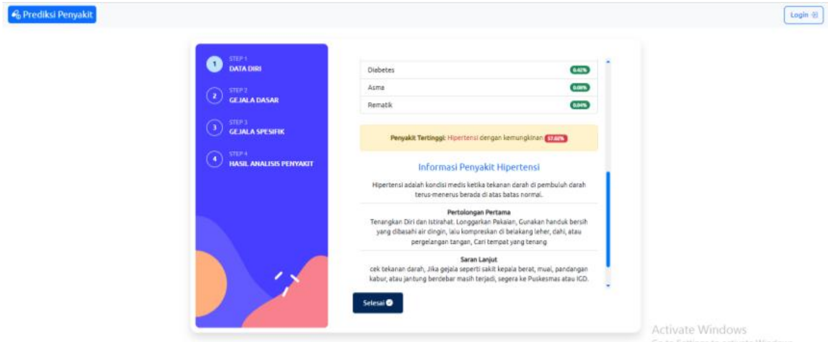

Figure 4. Diagnosis Result (2)

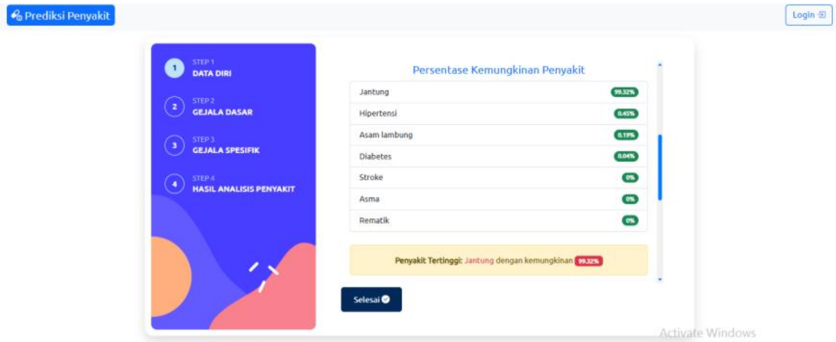

Figure 5. Manual Calculation Result

The results of this study show that the Naïve Bayes algorithm is able to produce an early diagnosis of non-communicable diseases with a high posterior probability. In scenarios with typical cardiac symptoms, such as chest pain, palpitations, pain radiating to the arm or jaw, as well as cold sweats, the system produces a posterior probability of 99.32% for heart disease, which illustrates a very strong level of system confidence. Technically, the dominance of this value is triggered by the weight of specific symptoms and repeatedly appears in heart cases, so that the result of probability multiplication immediately suppresses the probabilities of other classes. Clinically, this combination of symptoms is indeed a classic indicator of acute coronary syndrome that is rarely found together in other diseases, making the diagnosis more certain. In contrast, in stroke scenarios with a filtering approach, the system produces a posterior probability of 90.00% for stroke, as early symptoms such as dizziness, blurred vision, and arm weakness also often appear in hypertension and diabetes. The probability of a new stroke increases after advanced symptoms such as slurred speech and facial deviation are added, but still leaves the contribution of another class. This suggests that Naïve Bayes is able to provide a high level of certainty when symptoms are very typical, but remains realistic when symptoms are partial or overlapping.

Compared to the studies by Oskar et al. (2024) and Jefi et al. (2022) this study is superior because it does not only test on a single disease, but designs a multi-disease diagnosis mechanism with a separation of basic and advanced symptoms. This study introduces two inference strategies, namely direct and filtering, which provide flexibility according to the condition of the symptoms. When compared to studies Pratama et al. (2022) and Pinem & Putra (2025), the advantage of this study is not only in the high posterior probability, but in its ability to maintain the reliability of diagnosis when the initial symptoms are still partial through a filtering approach. This gradual strategy has never been used in previous studies, and even compared to studies Saptawan et al. (2024), the system is more stable in rural areas due to the combination of a fast direct approach to typical symptoms and an adaptive filtering approach to overlapping symptoms.

These findings show that the design of the two inference paths in Naïve Bayes can provide outputs that correspond to variations in field conditions. The direct approach is very effective for typical symptoms because it speeds up the certainty of diagnosis, while the filtering approach allows the system to remain reliable when patient information is still limited. In addition, the implementation of the system in the form of web and mobile applications makes this diagnosis mechanism widely accessible to the community and village medical personnel, thus not only strengthening the speed of early detection of diseases, but also increasing the availability of practical health services in areas with limited infrastructure. This study shows the development of two Naïve Bayes inference paths that not only increase the posterior probability of diagnosis, but also adapt the diagnosis flow to clinical realities in the field. Thus, the results of this study not only confirm the algorithm's ability to generate high posterior probability, but also show how technical strategies and forms of implementation can support medical decisions while expanding the reach of public health services.

4. Conclusions

This research proves that the implementation of the Naïve Bayes algorithm can produce a non-communicable disease diagnosis expert system that is relevant to the limited health service conditions in rural areas. The system, which was designed to efficiently process symptom data and provide an initial diagnosis based on posterior probability, supports early detection and strengthens the role of medical personnel in decision-making. The main contribution of this research lies in the development of two inference pathways: direct for typical symptoms and filtering for partial symptoms, which allows for a more adaptive diagnosis to variations in clinical conditions. Implementation in the form of a web and mobile application also expands public access to initial diagnostic services, while potentially increasing health awareness and reducing the risk of disease complications. However, this research is still limited to seven types of non-communicable diseases. For further development, it is suggested that the scope of the dataset and symptom variations be expanded, and that comparative experiments be carried out with more sophisticated approaches, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or other deep learning models, to assess whether these methods can provide improvements in the representation of more complex symptom features. Integration with structured medical record data, provision of personalized healthy lifestyle recommendation features, and connectivity with telemedicine services can also strengthen the usefulness of this system and extend its benefits in the field of public health.

References

Aprilia, T. (2024). Klasifikasi Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor dan Metode Naive Bayes. *SATESI: Jurnal Sains Teknologi Dan Sistem Informasi*, *4*(2), 156–163. https://doi.org/10.54259/satesi.v4i2.3167

Armansyah, & Ramli, R. K. (2022). Model Prediksi Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *6*(1), 1–10. https://doi.org/10.29408/edumatic.v6i1.4789

Asyim, RB., & Yulianto. (2022). Perilaku Konsumsi Obat Tradisional dalam Upaya Menjaga Kesehatan Masyarakat Bangsawan Sumenep. *Jurnal Keperawatan*, *Vol. 15*(No. 2), 2.

Chairina, & Candrasa, L. (2022). Peran Manajemen Arsip dalam Pengamanan Data Base. *All Fields of Science Journal Liaison Academia and Sosiety*, *2*(4), 29–35. https://doi.org/10.58939/afosj-las.v2i4.471

Dwiramadhan, F., Wahyuddin, M. I., & Hidayatullah, D. (2022). Sistem Pakar Diagnosa Penyakit Kulit Kucing Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *6*(3), 429–437. https://doi.org/10.35870/jtik.v6i3.466

Ehatisham-ul-Haq, M., Arsalan, A., Raheel, A., & Anwar, S. M. (2021). Expert-novice classification of mobile game player using smartphone inertial sensors. *Expert Systems with Applications*, *174*, 114700. https://doi.org/https://doi.org/10.1016/j.eswa.2021.114700

Elkomy, S., & Jackson, T. (2024). WHO non-communicable diseases Global Monitoring Framework: Pandemic resilience in sub-Saharan Africa and Low-income Countries. *Socio-Economic Planning Sciences*, *95*, 102043. https://doi.org/https://doi.org/10.1016/j.seps.2024.102043

Ether, S., & Saif-Ur-Rahman, K. M. (2021). A systematic rapid review on quality of care among non-communicable diseases (NCDs) service delivery in South Asia. *Public Health in Practice*, *2*, 100180. https://doi.org/https://doi.org/10.1016/j.puhip.2021.100180

Fadhillah, M., Sari, H. L., & Elfianty, L. (2023). Analisis dan Penerapan Algoritma Naïve Bayes untuk Klasifikasi Penyakit Gingivitis. *MEANS (Media Informasi Analisa Dan Sistem)*, *8*(2), 222–226. https://doi.org/10.54367/means.v8i2.3264

Fiqriansyah, R., Akbar, F., Andiko, V. C., Ahmad, K. G., Rasywir, E., Meisak, D., Pratama, Y., & Feranika, A. (2022). Penerapan Algoritma Naïve Bayes Untuk Mengetahui Pasien Penyakit Gagal Jantung. *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, *2*(2), 263–266. https://doi.org/10.33998/jakakom.2022.2.2.117

Fourniawan, R. H. A., Eviyanti, A., & Busono, S. (2024). Pengembangan Artificial Intelligence Berupa Virtual Assistant Mobile Menggunakan Framework Flutter. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, *7*(1), 53. https://doi.org/10.37600/tekinkom.v7i1.1169

Islam, Md. M., Rahman, Md. J., Islam, M. M., Roy, D. C., Ahmed, N. A. M. F., Hussain, S., Amanullah, M., Abedin, Md. M., & Maniruzzaman, Md. (2022). Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh. *International Journal of Cognitive Computing in Engineering*, *3*, 46–57. https://doi.org/https://doi.org/10.1016/j.ijcce.2022.02.002

Jefi, Hendri, Afni, N., & Salim, A. (2022). Penerapan Algoritma Naive Bayes Dalam Memprediksi Penjualan Material. *Jurnal Universitas Muhammadiyah Jakarta*, *12*(2), 10–15.

Krosman, K., Sosnowski, J., & Gawkowski, P. (2021). Object oriented time series exploration: Applied to power consumption analysis of embedded systems. *Expert Systems with Applications*, *184*, 115531. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115531

Mitiche, I., Jenkins, M. D., Boreham, P., Nesbitt, A., & Morison, G. (2021). An expert system for EMI data classification based on complex Bispectrum representation and deep learning methods. *Expert Systems with Applications*, *171*, 114568. https://doi.org/https://doi.org/10.1016/j.eswa.2021.114568

Oskar, Fransiska Ria, Rupina, & P, N. (2024). Pengelolaan Data Penyakit Jantung Dengan Menggunakan Metode Naive Bayes. *Jurnal Sains Dan Komputer*, *8*(02), 49–54. https://doi.org/10.61179/jurnalinfact.v8i02.531

Pinem, T. H., & Putra, Z. P. (2025). Evaluasi Kinerja Algoritma Klasifikasi Deep Learning dalam Prediksi Diabetes. *Jurnal Ilmiah FIFO*, *17*(1), 17. https://doi.org/10.22441/fifo.2025.v17i1.003

Pohan, K. P., & Chairunisah, C. (2024). Sistem Pakar Mendiagnosa Penyakit Kulit Pada Manusia Menggunakan Metode Naive Bayes Classifier Berbasis Web. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, *23*(1), 204. https://doi.org/10.53513/jis.v23i1.9521

Prahasti. (2022). Aplikasi Pelayanan Antrian Pasien Menggunakan Metode FCFS Menggunakan PHP dan MySQL. *Jurnal Media Infotama*, *18*(1), 341139.

Pratama, Y., Prayitno, A., Azrian, D., Aini, N., Rizki, Y., & Rasywir, E. (2022). Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor. *Bulletin of Computer Science Research*, *3*(1), 52–56. https://doi.org/10.47065/bulletincsr.v3i1.203

Riany, A. F., & Testiana, G. (2023). Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes. *Jurnal SAINTEKOM*, *13*(1), 42–54. https://doi.org/10.33020/saintekom.v13i1.352

Sano, A. V. D., Stefanus, A. A., Madyatmadja, E. D., Nindito, H., Purnomo, A., & Sianipar, C. P. M. (2023). Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier. *Procedia Computer Science*, *227*, 482–489. https://doi.org/10.1016/j.procs.2023.10.549

Saptawan, F., David, Wijaya, T., Kosasi, S., & Margaretha Kuway, S. (2024). Prediksi Epidemiologi Penyakit Tidak Menular Menggunakan Algoritma Random Forest Pada Puskesmas. *Jurnal TIMES*, *13*(2), 192–201. https://doi.org/10.51351/jtm.13.2.2024788

Setiawan, I. B., Maulindar, J., & Nurchim. (2024). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Pada Aplikasi Kesehatan Digital. *G-Tech : Jurnal Teknologi Terapan*, *8*(1), 186–195.

Suparyanto, & Rosad. (2022). Klasifikasi Polri Tentang Larangan Merokok. In *Suparyanto dan Rosad (2015* (Vol. 5, Issue 3).

Syahputra, P., & Kurniawan, R. (2024). Analisis Sentimen Terhadap Kesehatan Mental Remaja Menggunakan Metode Naive Bayes. *Journal of Information System Research (JOSH)*, *5*(4), 1216–1224. https://doi.org/10.47065/josh.v5i4.5644

Syahputra, S., & Hasibuan, M. S. (2024). Analisis Sentimen Jamaah Umrah Di Media Sosial X Menggunakan Algoritma Naive Bayes. *Jurnal Teknika*, *19*(1), 107–116.

Tugiman, Lily Damayanti, Alexius Hendra Gunawan, & Samuel Ryon Elkana. (2022). Prediksi Penggunaan Obat Peserta Jaminan Kesehatan Nasional Menggunakan Algoritma Naïve Bayes Classifier. *Journal of Applied Computer Science and Technology*, *3*(1), 144–150. https://doi.org/10.52158/jacost.v3i1.295

Wahidin, M., Agustiya, R. I., & Putro, G. (2023). Beban Penyakit dan Program Pencegahan dan Pengendalian Penyakit Tidak Menular di Indonesia. *Jurnal Epidemiologi Kesehatan Indonesia*, *6*(2). https://doi.org/10.7454/epidkes.v6i2.6253

Yuliza, R. (2022). Sistem Pakar Akurasi dalam Mengidentifikasi Penyakit Gingivitis pada Gigi Manusia dengan Metode Naive Bayes. *Jurnal Sistim Informasi Dan Teknologi*, *5*, 27–32. https://doi.org/10.37034/jsisfotek.v5i1.157