International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Implementation of Naïve Bayes Classifier to Determine Disease

Yunita Sartika Sari
Universitas Mercu Buana, Jl. Meruya Selatan No.1
Email address: yunita.sartika@mercubuana.ac.id

Abstract*—Information system is a system that provides information for management in making decisions. One of the efforts made to improve the quality of health services requires an optimal and structured system that is easy to use. In this study data collection was carried out through interviews and using the Naïve Bayes Classifier method. For the method of analysis using the SWOT method. The purpose of this study is to determine common diseases by implementing the Naïve Bayes method*.

***Keywords*—** *Information Systems, Naïve Bayes, Classifier*.

I. INTRODUCTION

Background

In the current era of globalization, technological advances are growing rapidly, including in the health sector. The availability of health information is very necessary in the implementation of effective and efficient health efforts. One of these technological advancements is Information Technology which has developed throughout all areas of human life. Medical record is a file that contains notes and documents regarding patient identity, examination, treatment and other actions for patients while receiving treatment at service providers for patients both inpatient and outpatient [1]. In 2003 RAND Health Information Technology (HIT) began conducting studies to better understand the role and importance of Electronic Medical Records (EMR) in improving health services and informing the government so they can maximize the benefits of EMR and improve health services. Electronic Medical Record (EMR) is a system that contains a patient's medical history and illness, diagnostic test results, other medical data and information on treatment costs.

The development of information technology that is currently happening allows the development of a way of storing and managing data electronically, at the Pekanbaru "X" Hospital just implementing the Electronic Medical Record (EMR)[2]. Based on the Attachment to SK PB IDI No 315/PB/A.4/88 Concerning Health Medical Records, records in written form or descriptions of service activities provided by medical/health service providers to a patient.

Based on the Decree of the Minister of Health Number: 269/Menkes/PER/III/2008 concerning medical records, it is explained that a medical record is a file that contains notes and documents regarding patient identity, examination, treatment, actions and other services that have been provided to patients. This study aims to facilitate the community's needs in knowing the disease based on the symptoms felt and make it easier to receive health information online using the Naïve Bayes Classifier Method.

Based on the background above, we identify and formulate problems to provide information regarding health easily and efficiently

The goal to be achieved in this research is to be able to easily monitor diseases that often occur in the community. provide information about the history of the disease in each patient, and provide education to the public about the disease. The expected benefits of this research are expected to provide a sense of security to the community when checking conditions and provide interesting experiences so that people are not afraid to carry out health checks.

II. LITERATURE REVIEW

Naïve Bayes Classifier

The RFM Method is used to analyze data for each customer based on segmentation using the "usage rate" attribute on the data, helping Naive Bayes, a method of classifying data, produce the result that 62% of the data is feasible and 38% is not feasible [4]. The processed data can then be used as a basic reference in decision-making. Based on the Bayes Theorem, the Naive Bayes Classifier is a straightforward probability classifier. The Bayes theorem will be paired with "Naive," which denotes the independence of each attribute and variable. In supervised learning, the Naive Bayes Classifier can be effectively trained [5].

X : Data with unknown class

$$P(H \mid X) = \frac{P(H/X) + P(X/H)}{P(X)}$$

H: Hypothesis data is a specific class
P(H/X): Probability of hypothesis H based on condition X (posteriori probability)
P(H): Probability of the hypothesis H (probability prior)
P(X/H): Probability of X based on conditions in the hypothesis
HP(X): Probability of X

Characteristics of the Naive Bayes Classifier

The Naïve Bayes method works robustly on isolated data which is usually data with different characteristics (outliners). Naïve Bayes can also handle incorrect attribute values by ignoring training data during the model building and prediction processes.

1. Tough dealing with irrelevant attributes.
2. Attributes that have a correlation can degrade classification performance

Yunita Sartika Sari, "Implementation of Naïve Bayes Classifier to Determine Disease," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 6, Issue 2, pp. 160-163, 2023.

3. Naïve Bayes because the assumption of independence of these attributes no longer exists.

The strength of Naïve Bayesian is that it is easy to implement and gives good results in most cases. While the weakness is having to assume that the features are not related (independent). In reality, the linkage exists. The linkage cannot be modeled by the Naïve Bayesian Classifier.

SWOT Analysis Method

The SWOT analysis method is a method used in evaluating strengths, weaknesses, opportunities and threats in an ongoing business process. SWOT can also be utilized in determining the goals of the business process, identifying existing factors both internal and external and helping to achieve the goals to be achieved [9]. The word SWOT consists of 4 components [10], including:

a) Strengths, namely internal conditions which are the main factors of success in achieving the goals to be achieved.
b) Weaknesses, namely internal conditions that can hinder success in achieving the goals to be achieved.
c) Opportunities, namely external conditions that can support success in achieving the goals to be achieved.
d) Threats, namely external conditions that can be a threat or obstacle to success in achieving the goals to be achieved.

III. Research Methodology

Method of collecting data

In this study the data collection method used was Primary Data and Secondary Data:

1. Primary data

Data obtained directly from the object of research through the process of observation and interviews. Observation by observing the activities of the hospital and recording important things and running system processes.

2. Data Secondary

Data obtained from the process of hospital activities, the organizational structure and documents used.

Research Stages

In this case it explains the steps taken to obtain a research methodology which is a stage that must be applied so that research can be carried out in a more directed manner and makes it easier to analyze existing problems.

The stages of the research are described as follows:

1. Identification of problems

Identifying the problem is the stage that the writer takes about the problem to be studied. The method used is the SWOT method to analyze system requirements.

2. Collecting data

Collect data by conducting interviews with stakeholders.

3. Data analysis

Analyzing and processing the data obtained from the interview process and data collection.

4. Study Literature

Study and collect references in the form of books, journals, ebooks related to theory in research.

IV. Results And Discussion

Identify the problems that exist in the disease experienced by the community based on the symptoms felt, an analysis can be carried out on Strengths, Weakness, Opportunity, Threats known as SWOT analysis.

TABLE 1. SWOT analysis

Internals / External	Strength (S)	Weaknesses (W)
	• Have a great business opportunity • The User Interface display is attractive and easily understood by users • Make it easier for patients to know the disease from the symptoms they feel	• The system cannot order drugs.
Opportunities (O) • Consultation process is easier with efficient time • Very easy to use display.	SO strategy • Give Service With Fast Response • Display easy to understand in all circles	WO strategy • Add drug ordering system
Treat(T) • Has many competitors in similar business systems • There are no security features that can protect user data	ST Strategy • Added live chat feature • Upgradesafety features	WT Strategy • Firecurrently available features

Naïve bayes classifier calculation:

$$P = (ai|vj) = \frac{nc + mp}{n + m}$$

nc = number of records in learning data v = vj and a = ai
p = 1/ number of class/disease types
m = number of parameters/symptoms
n = the number of records in the learning data
v = vj for each class

Specifies the value for each class.

Then calculate the value of P (ai |vj) and calculate the value of P (vj)

$$V_{MAP} = \text{argmax}_{vj} \in V \, P\,(V_j)\pi_i \, P\,(a_i \mid V_j)$$

$$P\,(A_i|\ (V_j\ ) = \frac{n_c + \text{m.p}}{n + m}$$

Calculates P(ai|vj) x P(vj) for each v

Determine the results of the classification, namely v which has the largest multiplication result.

In applying the Naïve Bayes Classifier Method to a diagnostic system, symptom data is required to be input into the system, and processed so that it displays the results of disease diagnoses according to the Naïve Bayes Classifier Method based on symptoms.

Disease data information is needed based on symptoms that will be entered into the system and then processed, so as to produce a disease diagnosis according to the method used. The data used in this study applied the naive Bayes classifier method with 5 types of disease and 29 symptoms. The disease data and symptoms are as follows:

P01 : Pharyngitis
P02 : Dengue Hemorrhagic Fever
P03 : Influenza
P04 : Diarrhoea
P05 : Types

From these 5 types of disease, 29 types of symptoms were obtained, while the symptoms were from typedisease as follows:

Yunita Sartika Sari, "Implementation of Naïve Bayes Classifier to Determine Disease," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 6, Issue 2, pp. 160-163, 2023.

G01	Runny nose	G16	Red Spots On The Skin
G02	Hoarseness	G17	Body Feels Hot Cold
G03	High fever	G18	Sore throat
G04	Abdominal pain	G19	Muscle ache
G05	Vomit	G20	Severe Headache
G06	Hard to breath	G21	Cough
G07	Pain When Swallowing	G22	Weaknesses and Feelings Uncomfortable
G08	Bleeding (nosebleeds, skin rashes, bruising without cause, vomiting blood, bloody bowel movements	G23	Fever
G09	Dizziness And Pain Behind The Eyes	G24	Liquid CHAPTER
G10	Nauseous	G25	Sunken eyes
G11	No appetite	G26	Dehydration
G12	Body Pain	G27	Stomacache
G13	Weak, Weary, Lethargic	G28	Elasticity Decrease
G14	Body Feels Cold	G29	The tongue is white
G15	Constipation (Difficult to Poop)		

Stages of the Naive Bayes method process:

1. Counts the number of classes / labels.
2. Calculating the Number of Cases Per Class
3. Multiply All Class Variables
4. Compare Results Per Class

Naive Bayes calculation steps:

1. Determine the nc value for each class

P01(Pharangitis)
N = 1, P = 1/5 = 0.2, M = 29, G01.nc = 1, G02.nc = 1, G03.nc = 1, G04.nc = 1
P02(Dengue Hemorrhagic Fever)
N = 1, P = 1/5 = 0.2, M = 29, G01.nc = 0, G02.nc = 0, G03.nc = 1, G04.nc = 0
P03(Influenza)
N = 1, P = 1/5 = 0.2, M = 29, G01.nc = 0, G02.nc = 0, G03.nc = 1, G04.nc = 0
P04(diarrhoea)
N = 1, P = 1/5 = 0.2, M = 29, G01.nc = 0, G02.nc = 0, G03.nc = 0, G04.nc = 0
P05(Types)
N = 1, P = 1/5 = 0.2, M = 29, G01.nc = 0, G02.nc = 0, G03.nc = 0, G04.nc = 0

2. Determine the value of P(ai|vj) and calculate the value of P(vj)

P01 (Pharangitis)
G01= 0.226667, G02= 0.226667, G03=0.226667, G04=0.226667
P02 (Dengue Hemorrhagic Fever)
G01=0.193333, G02=0.193333, G03=0.226667, G04=0.193333
P03 (Influenza)
G01=0.193333, G02=0.193333, G03=0.226667, G04=0.193333
P04 (Diarrhoea)
G01=0.193333, G02=0.193333, G03=0.193333, G04=0.193333
P05 (Types)
G01=0.193333, G02=0.193333, G03=0.193333, G04=0.193333

3. Calculates P(ai|vj) x P(vj) for each v

P01 Pharyngitis	0.000527935
P02 Dengue Hemorrhagic Fever	0.000327595
P03 Influenza	0.000327595
P04 Diarrhoea	0.000279420
P05 Types	0.000279420

4. Determine the classification results that have the largest multiplication.

The result of the highest multiplication classification is the value of 0.000527935, namely pharyngitis (P01)

V. CONCLUSIONS

Conclusion

Based on the research that has been done, some conclusions can be drawn as follows:

1. With this system, it is hoped that it will make it easier for patients to find out the disease based on the symptoms they feel.
2. Can save patient time without the need to come to the hospital directly.
3. With this system, it can avoid queues of patients who want to check their condition.

Suggestion

Here are some suggestions that can be used as input for the next application development:

1. With so much competition for similar applications, this application still has few features compared to other competitors. So to be able to compete, the authors suggest that in the development process to be able to add drug ordering features, so that the system can be better at providing services to users that will create a sense of satisfaction.
2. It is hoped that in the future, health applications can diagnose more than 20 diseases.

REFERENCES

[1.] S. Apriliyani, P. Piksi, and G. Bandung, "Use Of Electronic Medical Records For * Correspondent Author : Sinta Apriliyani," vol. 1, no. 10, p. 1399–1410, 2021.

[2.] Fitriani Astika, "Application of Electronic Medical Record (EMR) at 'X' Hospital Pekanbaru in 2019," J. Hosp. Manag. Heal. Sc., vol. 1, no. 1, pp. 43–53, 2020, [Online]. Available: http://journal.al-matani.com/index.php/jhmhs/article/view/26.

[3.] Putra, R. R., Johan, M. E., & Kaburuan, E. R. (2019). A naïve bayes sentiment analysis for fintech mobile application user review in Indonesia. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(5), 1856-1860.

[4.] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "Applement Of Data Mining For Patient Medical Record Data Grouping Based On Type Of Disease With Clustering Algorithms (Case Study: Polyclinic PT. Inecda)," J. Computer Science ., vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.

[5.] I. Ranggadara, G. Wang, and ER Kaburuan, "Applying Customer Loyalty Classification with RFM and Naïve Bayes for Better Decision

Making," Proc. - 2019 Int. Monday. appl. Technol. inf. commun. ind. 4.0 Retrosp. Prospects. Challenges, iSemantic 2019, no. september, pp. 564–568, 2019, doi: 10.1109/ISEMANTIC.2019.8884262.
[6.] ASR Sinaga and D. Simanjuntak, "Expert System for Detecting Malnutrition in Toddlers Using the Naïve Bayes Classifier Method," J. Inkofar, vol. 1, no. 2, pp. 54–60, 2020, doi: 10.46846/jurnalinkofar.v1i2.110.
[7.] A. Febrianto, S. Achmadi, and AP Sasmito, "Application of the K-Means Method for Clustering Visitors to the Itn Malang Library," J. Mhs. Tech. Inform., vol. 5, no. 1, pp. 61–70, 2021.
[8.] R. Rusliani, "Cash Receipt and Disbursement Application Program Using PHP at the Noor Hidayaturrahim Mosque, Banjarmasin. This study discusses the application of the accounting information system for cash receipts and disbursements at the Noor Hidayaturrahim Mosque in," no. 36, pp. 1–10, 2020.
[9.] F. Technology, AND Informatics, and U. Dynamics, "Website User Interface Design at Cv. Alpha Raya Technic," 2021.
[10.] MI Puspita, I. Ranggadara, and I. Prihandi, "Zachman's framework for design information system logistics management," Int. J. Recent Technol. Eng., vol. 8, no. 3, pp. 4030–4034, 2019, doi: 10.35940/ijrte.C5377.098319.
[11.] AF Sallaby and I. Kanedi, "Designing a Doctor's Schedule Information System Using the CodeIgniter Framework," J. Media Infotama, vol. 16, no. 1, pp. 48–53, 2020, doi: 10.37676/jmi.v16i1.1121.
[12.] RT Waruwu and A. Sindar, "Expert System Determines Types of Child Development Disorders Using the Certainty Factor Method," Sci. Comput. sci. Informatics J., vol. 1, no. 2, pp. 1–4, 2020.
[13.] M. Sari, S. Defit, and GW Nurcahyo, "Expert System for Disease Detection in Children Using the Forward Chaining Method," J. Inf System. and Technol., vol. 2, pp. 5–9, 2020, doi: 10.37034/jsisfotek.v2i4.114.
[14.] A. Apriliani et al., "Forecasting Restaurant Menu Sales Trends Using Sales Trend Forecast of a Restaurant Menus Using Single Moving," J. Teknol. inf. and Computing Science., vol. 7, no. 6, pp. 1161–1168, 2020, doi: 10.25126/jtiik.202072732.
[15.] RR Rizky and ZH Hakim, "Expert System for Determining Hypertension in Pregnant Women at Adjidarmo Rangkasbitung Hospital, Banten Province," J. Sisfokom (Inf. and Computer Systems), vol. 9, no. 1, pp. 30–34, 2020, doi: 10.32736/sisfokom.v9i1.781.
[16.] Dwiasnati, S., & Devianto, Y. (2018). Naïve Bayes Optimization Based On Particle Swarm Optimization to Predict the Decision of Insurance Customer Candidate. *Int. J. Comput. Tech*, *5*(5), 8-14.
[17.] Priyono and N. Nuris, "The TOPSIS Method Decision Support System for Diagnosing Dengue Fever," Inti Nusa Mandiri Journal, vol. 15, no. 1, pp. 51–58, 2020.
[18.] EA Saputra and S. Nurmiati, "Decision Support System for Determining Major Illness Symptoms at Web-Based Salak Bogor Hospital," J. Computer Science., pp. 1–4, 2019, [Online]. Available: https://jurnal.pranataindonesia.ac.id/index.php/jik/article/view/58.
[19.] Y. Yuvidarmayunata, "Web-Based Expert System Using Backward Chaining Methods To Determine Proper Nutrition For Pregnant Women," INTECOMS J. Inf. Technol. Comput. Sc., vol. 1, no. 2, pp. 231–239, 2018, doi: 10.31539/intecoms.v1i2.302.
[20.] AW Ganda Anggara, Gede Pramayu, "Building an expert system using Bayes' theorem to diagnose lung disease," Semin. Nas. Technol. inf. and Multimed. 2016, pp. 79–84, 2016.
[21.] TF Ramadhani, I. Fitri, and ETE Handayani, "Web-Based ISPA Diagnostic System Using Forward Chaining Method," JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 5, no. 2, p. 81, 2020 , doi: 10.31328/jointecs.v5i2.1243.
[22.] A. Gunawan, S. Defit, and S. Sumijan, "Expert System in Identifying Gynecological Diseases Using the Android-Based Forward Chaining Method," J. Inf System. and Technol., vol. 2, no. 1, pp. 15–23, 2020, doi: 10.37034/jsisfotek.v2i1.30.