Classification of Hypertension Using Naïve Beyes Method with Risk Factor Data Discretization Approach

Yazid Munali
Computer Science Department
Faculty of Science and Technology
Universitas Islam Negeri Sumatera Utara
Medan, Indonesia
yazid.munali@uinsu.ac.id

Armansyah
Computer Science Department
Faculty of Science and Technology
Universitas Islam Negeri Sumatera Utara
Medan, Indonesia
armansyah@uinsu.ac.id

Abstract— Generally, patients are unaware of their hypertension condition before having their blood pressure checked. One out of three Indonesians suffers from hypertension, and this figure continues to rise annually. Hypertension is often referred to as the silent killer because individuals with high blood pressure do not exhibit symptoms. This study aims to classify hypertensive patients in an effort to reduce the prevalence of hypertension in Indonesia by aiding in early detection of the disease and increasing awareness of hypertension among the Indonesian population. By using the Naïve Bayes method and implementing data discretization of risk factors, the dataset used comprises 11,627 health examination records of 4,434 participants from the Framingham Heart Study (FHS) organized by the National Institutes of Health. The classification method utilizes the Naïve Bayes Algorithm, and data discretization is performed using the CART (Classification and Regression Trees) method. The system provides an estimation of the probability of hypertension occurrence based on input factors/symptoms, where Naive Bayes achieves an accuracy rate 84.28%.

Keywords— Hypertension, Classification, Data Discretization, Naive Bayes.

I. INTRODUCTION

Hypertensive disease is a fairly common health problem in society. Hypertension can lead to several complications, such as stroke, heart disease, and kidney disease[2][3][4][13][17], Hypertension is also the third largest risk factor for premature death [18]. High blood pressure is called hypertension where the blood pressure against the artery wall is quite high. Hypertension is often referred to as a silent killer, because it is a deadly disease without being accompanied by symptoms as a warning to its victims[1] [3] [4].

Although some symptoms may occur at the same time and are believed to be associated with high blood pressure (hypertension), they may not be symptoms of hypertension. In this case the symptoms include headache, bleeding from the nose (nosebleeds), dizziness, facial flushing and fatigue [1] [4].

Common risk factors include blood pressure, body mass index (BMI), heart rate, physical activity, age, diabetes, genetics/family history. In addition, other studies have included additional factors such as cholesterol levels, smoking, unhealthy diet, obesity, and gender in determining hypertension [5] .

Implementing early intervention and effective control measures can substantially decrease both the occurrence and fatality rates associated with hypertension and its complications. Therefore, establishing a highly efficient and precise risk prediction system for managing hypertension holds significant importance[17].

The learning process in machine learning utilizes specialized algorithms commonly referred to as machine learning algorithms (ML). There are two types of ML: unsupervised learning and supervised learning[6]. Supervised methods have been observed to be more accurate and consistent compared to unsupervised techniques [15].

The naive Bayes classifier algorithm is a supervised ML algorithm that employs probabilistic computations for classification. This algorithm, grounded in Bayes' theorem, assumes that each feature independently and equally influences the target class, without interactions among features [14], such that each feature independently and equally contributes to the probability of a sample to belong to a specific class because all the features independently contribute in making a decision hence it is called as “Naive” 15]. However, this approach's reliance on treating predictors as independent variables can be seen as a disadvantage of the method, since in most real fault diagnosis cases, the symptoms can be dependent on each other [16]. In this context, the use of the naïve Bayes classifier allows for the utilization of continuous features, but this necessitates a strong assumption about the distribution of their values. Discretization removes the need for this assumption by providing a direct evaluation of the conditional probability of categorical values based on counts within the dataset [16].

Previous research by Changpetch, Pitpeng, Hiriote, & Yuangyai (2021) found that the most effective discretization method was a classification tree using weight to calculate measures such as the proportion of data in each class, the proportion of data in left and right child nodes, the Gini impurity index in each node, and the reduction in impurity of split [8]. The use of pre-discretized variables and interactions was deemed crucial in improving the classification accuracy of the naïve Bayes classifier [8]. As a discretization method, they developed and described a classification tree with weighting as the most effective way to partition quantitative predictors into levels [8]. On the other hand, widely used tree-based classifiers—including classification and regression trees (CART) and random forests—can be made more efficient through discretization, eliminating the need to sort continuous feature values during tree induction [16]. Supervised methods using pre-discretized data were found to be more accurate and consistent than unsupervised methods but tended to produce larger decision trees [16].

By utilizing the Naive Bayes prediction model, the system can provide an estimate of the probability of hypertension based on the factors/symptoms inputted. These results will make it easier for users to gain an understanding of the level of risk of hypertension that they may experience. Based on the above problems, it is necessary to apply a classification system to recognize hypertension, so that with this system it is hoped that people with hypertension factors can take early prevention, and the prevalence of hypertension can be reduced.

Previous studies show various applications of the naive beyes algorithm in classifying and optimizing it, including research by Nugroho, et al.[7]optimizing Naïve Bayes and Decision Tree methods with Unsupervised Discretization Approach to determine study programs for prospective new students. The results show an increase in accuracy of 0.98% from the results before optimization of 96.68%.

Other research by Izzuddin, et al. [5] provided a review of hypertension diagnosis using expert systems and wearable devices. They focused on hypertension factors, expert system techniques, and the types of sensors in such devices. The commonly mentioned expert system techniques involve machine learning, neural networks, and fuzzy logic, they also concluded the hypertension factors in the diagnosis system that often appear involve blood pressure, heart rate, physical activity, body mass index, age, body temperature, diabetes, genetics/family history, skin response, sodium intake, cholesterol level, smoking, unhealthy diet, obesity, gender, body fat level, and race.

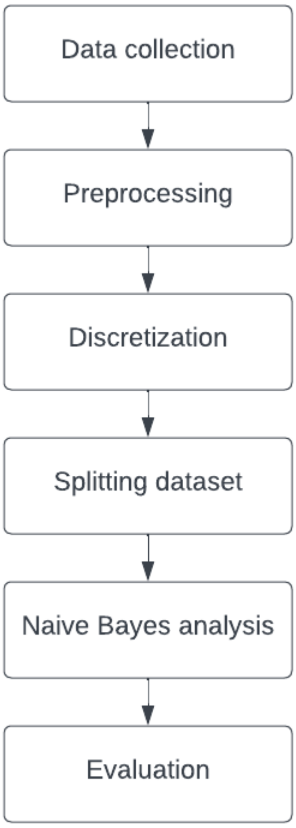
In addition, Changpetch, et al. [8]integrated data mining techniques to improve the performance of Naïve Bayes classifiers, especially in applications on medical datasets such as thyroid, diabetes, and appendicitis. The researchers combined three techniques - classification trees, analysis of association rules (ASA), and naïve Bayes classifiers - to improve the performance of naïve Bayes classifiers. The methodology excelled for all datasets, including the thyroid (LOOCV: 99.53%), diabetes (LOOCV: 81.25%), and appendicitis (LOOCV: 95.28%) datasets. The results show that combining these techniques can improve the performance of the naïve Bayes classifier on various types of datasets.

Various other studies using the naïve beyes method such as those conducted by Yanti Apriyani, et al. [9]for the diagnosis of pulmonary tuberculosis, Nurlelah, et al. [10] for dengue hemorrhagic fever, Surorejo, et al [4] for hypertension disease diagnosis, and Rizky, et al.[2] for hypertension disease detection, Other studies, such as Haffandi, et al.[11] explored the use of Naïve Bayes method in lung disease classification.

These studies reflect various approaches in the development of software for the diagnosis of hypertension, with a focus on methods such as Naïve Bayes, Decision Tree, and data mining techniques have been researched previously. This time the author applies the naïve Bayes method for hypertension classification using discrete data on hypertension risk factors.

II. RESEARCH METHODS

This research focuses on the application of the naive beyes algorithm in classifying hypertension patients based on website-based risk factors, for the sequence of steps that will be made in this study can be seen in the following figure.



Gambar 1. Research Methods

A. Data Collection

In this study, the authors used secondary data obtained from the Framingham Heart Study (FHS) organized by the National Institutes of Health. The study started in 1948 and there were 5,209 subjects initially enrolled in the study. The participants were examined every two years for cardiovascular disease. The clinical examination data included cardiovascular disease risk factors and disease markers such as blood pressure, blood chemistry, lung function, smoking history, health behaviors, ECG tracking, Echocardiography, and medication use. This resulted in 11,627 observations on 4,434 participants and consisted of 39 attributes. The data used in this study can be accessed by data request through the official website of the National Institutes of Health at : https://biolincc.nhlbi.nih.gov/studies/fhs/

B. Preprocessing

The purpose of preprocessing is to prepare text into data that can then be processed by the system The stages of preprocessing in this study include:

1) *Data cleaning*

Data cleaning is the process of removing noise or null data, removing duplicate data, checking for inconsistent data, and correcting errors in the data. From a total of 11,627 total records after deleting the data that has missing values, 11,099 data records remain

2) *Data selection*

The features used in this study refer to common factors in the diagnosis of hypertension written by Muhammad Izzuddin, et al "Review of hypertension diagnosis using expert systems and wearable devices", that there are 17 general criteria, namely: Blood Pressure, Heart Rate/Heart Rate, Steps/Physical Activity, Body Mass Index, Age, Body Temperature, Diabetes, Genetics/Family History, Skin Response, Sodium Intake, Cholesterol Level, Smoking, Unhealthy Diet, Obesity, Gender, Body Fat Level, and Race. Of these 17 criteria, there are 8 criteria that exist in the Framingham Heart Study data.

C. Discretization

This data discretization process involves converting the continuous values in the dataset into categorical groups or 'bins'. Using the Python module [12] DecisionTree https://scikitlearn.org/stable/modules/tree.html, each continuous value in the dataset is analyzed and grouped into the appropriate bin based on certain criteria determined by the Classification and Regression Trees (CART) algorithm. For obesity features, the author takes from BMI data which is in accordance with the regulation of the Minister of Health of the Republic of Indonesia Number 41 of 2014 concerning balanced nutrition guidelines.

D. Split Dataset

The dataset is split into training and testing data with a ratio of 80:20, which is 80% of training data and 20% of testing data. With a total of 11,099 data records after preprocessing. For Train data 11,099 * 80% = 8879 records, and for test data 11,099 * 20% = 2220 records.

E. Naive Bayes Analysis

The naive bayes method is a statistical approach to perform induction inference on classification problems. Naive Bayes assumes that the relationship between one attribute and another is conditionally free for class Y. It is called naive because this assumption is quite difficult to fulfill in real life. However, it turns out that this method has a fairly high accuracy rate for most cases. This method uses conditional probability as its basis. In statistics, conditional probability is expressed as shown below.

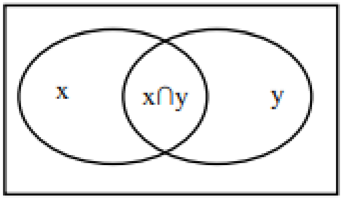


Figure 2. Conditional Probability

The probability of X in Y is the probability of intersection of X and Y from the probability of Y, or in other words P(X|Y) is the percentage of the number of X in Y. Chance or probability can be defined as the likelihood of an event occurring. The way to calculate the probability can be seen in Equation (1).

$$P(A) = \frac{n(A)}{n(S)} \qquad (1)$$

Description:
P(A): Probability of event A
n(A): Number of Occurrences of A
n(s): number of sample points S (number of universes)

Where the probability of event A is denoted by P(A), which is the sum of the probabilities of A from all members of the sample point S. Conditional probability is the probability of an event occurring on the condition that another event has occurred first. The probability of event A occurring on the condition that event B has occurred is calculated by Equation (2).

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \qquad (2)$$

Description:
P(A|B): The probability of Event A given that Event B has occurred
P(A∩B): probability of the event A intersects B
P(B): Probability of event B

Naive Bayes is a set of supervised learning algorithms based on the application of Bayes' theorem with a "naive" assumption of conditional independence between each pair of features assigned a class variable value. The equation for calculating Naive Bayes probabilities can be expressed as follows:[12].

$$P(A \mid B_1, \dots, B_n) = \frac{P(A)P(B_1,\dots,B_n|A)}{P(\mathrm{B},\dots,B_n)} \qquad (3)$$

Description:

P(A|$B_1, \dots, B_n$): Conditional probability A given$B1, \dots, B\ n$

P($B_1, \dots, B_n$|A): Conditional probability $B1, \dots, Bn$ given A

P(A): Prior probability of class (A)

P($B_1, \dots, B_n$): Prior probability of the class ($B1, \dots, Bn$)

By utilizing the independence assumption, the Naive Bayes method in disease diagnosis to generate probability estimates related to each possible disease. We assume that each factor is considered independent of each other, allowing us to estimate the probability of a condition based on the factors that appear. Therefore, the Naive Bayes approach can be an effective tool in disease diagnosis by providing relevant probability calculations based on factor observations.

In the classification process, the approach taken by the Naïve Bayes Classifier method is HMAP (Hypothesis Maximum Aposteriori Probability), which is to obtain a decision by finding the largest probability of all attributes in all possible decisions. So if there are as many events B as k and as many events A as i then the probability of the occurrence of event B if it is known that state A at i P(B|Ai) is expressed by the Naïve Bayes Classifier equation in the following equation[14] [19] .

$$\text{HMAP} = \arg max \prod_{i=1}^{n} P(B_K \mid Ai) \qquad (4)$$

with :

$P(Bk|Ai)$: Probability of B happening k times if A happens i times.

F. Evaluation

Calculation of the accuracy of the results obtained from the model is the final stage in assessing the performance of the algorithm used for classification. This process represents an important evaluation step that provides insight into the success of the applied algorithm. The evaluation results can be seen through the confusion matrix which consists of four combinations: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). The confusion matrix provides an accuracy value, which measures the accuracy of the model in classification using the following equation [20] [21]:

$$Akurasi = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \qquad (5)$$

Precision is another important metric that measures the accuracy of the model in predicting true positives out of the overall predicted positives, calculated by the equation:

$$\text{Presisi} = \frac{\text{TP}}{\text{TP+FP}} \qquad (6)$$

Recall measures the success of the model in finding true positive predictions from all true positive data, with the formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (7)$$

F1-score, which calculates the average of precision and recall, is calculated using the equation:

$$\text{F1-score} = \frac{2 \cdot \text{Presisi} \cdot \text{Recall}}{\text{Presisi+Recall}} \qquad (8)$$

Description:
True Positive (TP) = The amount of data that is positive and correctly predicted as positive.
False Positive (FP) = The amount of data that is negative but predicted as positive.
False Negative (FN) = The amount of data that is positive but predicted as negative.

True Negative (TN) = The amount of data that is negative and correctly predicted as negative. Using this method, a holistic evaluation can be obtained, providing a deep understanding of the effectiveness of the classification algorithm used.

III. RESULTS AND DISCUSSION

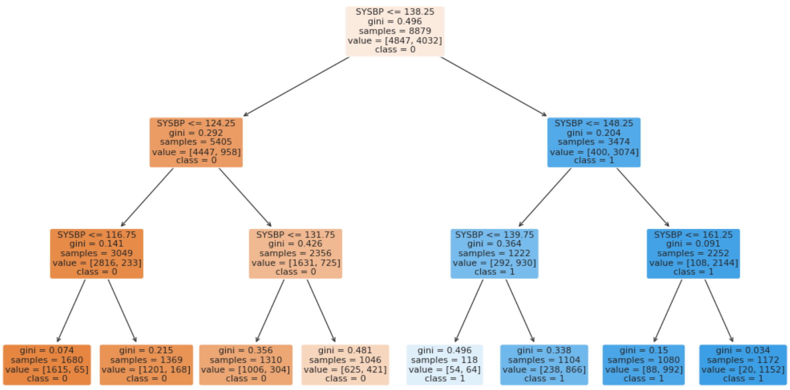
A. CART Discretization

Here the authors use the Pyton DecisionTree module[12] and performs a discretization loop on each feature variable of numeric type.

```
tree = DecisionTreeClassifier(max_depth=3)

for column in df.columns:
     if df[column].dtype != 'object':
         X = df[[column]]
         tree.fit(X, y)
         df[column] = tree.apply(X)
```

This code creates a decision tree with a maximum depth of 3 using DecisionTree from the sklearn.tree library, Here is the decision tree created with this model on the SYSBP feature.



Gambar 3. CART Decision Tree

There are a total of 15 nodes, 7 split nodes and 8 leafnodes, here the author takes the threshold value used to divide the dataset at the node so as to form the data range data, namely SYSBP <= 118.75, <= 127.75, <= 132.75, <= 139.25, <= 142.25, <= 148.25 and <= 161.25. The following SYSBP features after discretization

Table 3 Data After Discretization

Attributes	Number of Cases	hypertension Yes	hypertension No
TOTAL	8879	4032	4847
SYSBP 119.0-127.5	1636	256	1380
SYSBP 83.5-118.5	1954	89	1865

SYSBP 142.5-148.0	704	563	141
SYSBP 148.5-161.0	1080	992	88
SYSBP 161.5-295.0	1172	1152	20
SYSBP 139.5-142.0	405	306	99
SYSBP 133.0-139.0	977	422	555
SYSBP 128.0-132.5	951	252	699

B. Naïve Beyes Analysis

There are several stages in this analysis

1.) Prior Probability Calculation

Hypothesis Probability P(H) / Prior for each diagnosis class, i.e. patients suffering from hypertensive disease and patients not suffering from hypertensive disease. This uses equation (2).

2.) Likehood Calculation

Based on equations (2) and (3), to obtain the posterior probability value, the likelihood value P(A|B) or the probability of each diagnosis criterion A in each diagnosis class B is needed first. The occurrence value of each diagnosis criterion Ai that affects the occurrence of diagnosis class B1 or hypertension YES and class B2 or hypertension NO, The calculation of the likelihood value of the SYSBP feature in diagnosis classes B1 and B2 in the training database is shown in the following table.

Tabel 1. SYSBP Feature likehood table

Feature	P (Ai | B1)	P (Ai | B2)
SYSBP 119.0-127.5	0.0634921	0.2847122
SYSBP 83.5-118.5	0.0220734	0.3847741
SYSBP 142.5-148.0	0.1396329	0.0290902
SYSBP 148.5-161.0	0.2460317	0.0181556
SYSBP 161.5-295.0	0.2857143	0.0041263
SYSBP 139.5-142.0	0.0758929	0.020425
SYSBP 133.0-139.0	0.1046627	0.1145038
SYSBP 128.0-132.5	0.0625	0.1442129

3.) Diagnosis Class Calculation

In this step, the Naive Bayes formulation given in equation (2) is used. After obtaining both diagnosis class values P(A|B1) and P(A|B2), each of these data will be classified into the presence or absence of potential hypertension using the Naive Bayes formulation. So based on equation (4), the classification system will make a decision based on the largest value of P(A|Bi).

C. Validation

The test results on the testing data can be seen in the Confusion Matrix image.

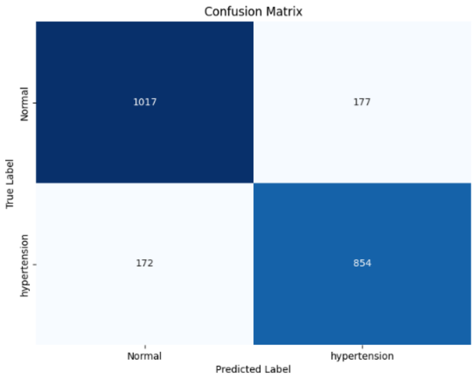


Figure 4. Confussion Matrix

True Positive Value: 845, False Positive Value: 172, False Negative Value: 177, and True Negative Value: 1017 and get values based on equations (5), (6), (7), (8), as follows:
Accuracy=0.8428,
Precision (Positive Predictive Value) =0.8283220174587779,
Recall (Sensitivity or True Positive Rate)=0.8323586744639376,
Specificity (True Negative Rate)=0.8517587939698492,
F1 Score=0. 8303354399611085.

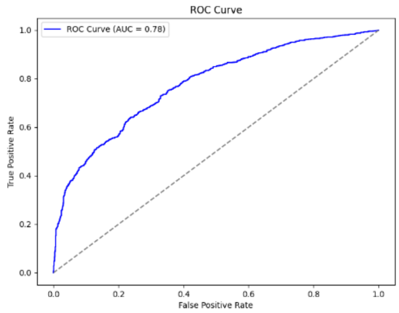


Figure 5. ROC curve

The results of testing the testing data for the Naive method on the ROC value show that the ROC Curve value is 0.78 for the evaluation of the Naive Bayes method.

D. App Implementation

After the evaluation stage, then implement it into a web-based program. The results of the implementation of the expert system are as follows:

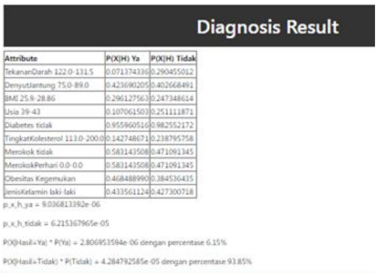


Figure 6. Diagnosis Results Page

Figure 7. Admin Page

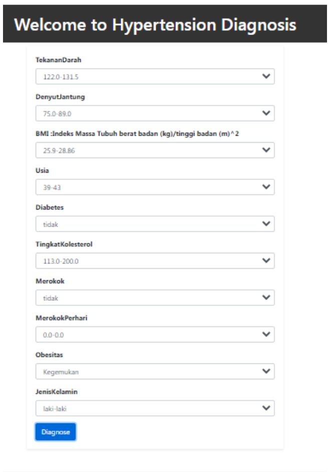


Figure 8. Diagnosis page

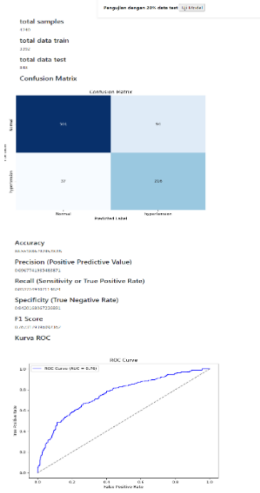


Figure 9 . Model Test Page

E. CONCLUSIONS

The Naïve Bayes method has been successfully applied to classify hypertension by utilizing a discretization approach on the risk factors of hypertension in the Framingham dataset. The analysis results indicate that this model is capable of predicting hypertension classification with an accuracy of 84.28%. This accuracy value has increased by 2.70% compared to before the data discretization. Binary classification testing of hypertension with cardiovascular disease data yielded accuracies of 99.05% and 99.94% after discretization. Similarly, binary classification testing of high blood pressure with The Behavioral Risk Factor Surveillance System (BRFSS) 2015 data resulted in accuracies of 71.55% and 71.58% before and after discretization, respectively. For binary classification testing of heart disease with BRFSS 2020 data, accuracies were 77.15% and 76.94% before and after discretization. Lastly, for binary classification testing of heart attacks with BRFSS 2022 data, accuracies were 77.55% and 76.47% before and after discretization. These results demonstrate that discretization has an impact on accuracy outcomes with the Naïve Bayes method. With the implementation of a web-based application, this model can provide binary classification on other datasets with minor configuration adjustments to the prediction class.

REFERENCES

[1] Supriyono, S., & Fadila, E. (2022). Sistem pakar penegakan diagnosa penyakit hipertensi dengan inferensi forward chaining menggunakan metode Support Vector Machine (SVM). Jurnal Kesehatan Medika Udayana, 8(02), 207–221. https://doi.org/10.47859/jmu.v8i02.283

[2] Rizky, R., Susilawati, S., Hakim, Z., & Sujai, L. (2020). Sistem Pakar Deteksi Penyakit Hipertensi Dan Upaya Pencegahannya Menggunakan Metode Naive Bayes Pada RSUD Pandeglang Banten. Jurnal Teknik Informatika Unis, 7(2), 138–144. https://doi.org/10.33592/jutis.v7i2.395

[3] Puspa, M. A. (2018). sistem pakar diagnosa penyakit hipertensi menggunakan metode naive bayes pada rsud aloe saboe kota gorontalo. Ilkom Jurnal Ilmiah, 10(2), 166–174. https://doi.org/10.33096/ilkom.v10i2.304.166-174

[4] Surorejo, S., Chaeriko, Y. P., Ananda, S., Studi, P., Informasi, S., Ymi Tegal, S., & Kunci, K. (2022). Penerapan Metode Naïve Bayes Pada Sistem Pakar Untuk Diagnosa Penyakit Hipertensi. 3(1), 8–17.

[5] Sani, M. I. M., Abdullah, N. a. S., & Rosli, M. M. (2022). Review on hypertension diagnosis using expert system and wearable devices. International Journal of Power Electronics and Drive Systems, 12(3), 3166. https://doi.org/10.11591/ijece.v12i3.pp3166-

3175

[6] Santoni, M. M., Chamidah, N., & Matondang, N. (2020). Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada 29 software KNIME. Techno.Com, 19(4), 353–363. https://doi.org/10.33633/tc.v19i4.3872

[7] Nugroho, W. E., Prihandoyo, T., & Somantri, O. (2022). Optimalisasi Metode Naive Bayes untuk Menentukan Program Studi bagi Calon Mahasiswa Baru dengan Pendekatan Unsupervised Discretization. Infotekmesin, 13(1), 161– 167. https://doi.org/10.35970/infotekmesin.v13i1.1048

[8] Changpetch, P., Pitpeng, A., Hiriote, S., & Yuangyai, C. (2021). Integrating Data mining Techniques for Naïve Bayes Classification: Applications to medical datasets. Computation (Basel), 9(9), 99. https://doi.org/10.3390/computation9090099

[9] Yanti Apriyani, Iskandar, I. D. I., Mira Kusmira, Melisa Winda Pertiwi, Imam Amirulloh, & Taufik Wibisono. (2021). Implementasi Sistem Pakar dengan Algortima Naïve Bayes dengan Laplace Correction untuk Diagnosis Tuberkulosis Paru. INFORMASI (Jurnal Informatika Dan Sistem Informasi), 13(1), 24–46. https://doi.org/10.37424/informasi.v13i1.72

[10] Nurlelah, E., Abdilah, A., & Ghani, M. A. (2019). implementasi algoritma naive bayes pada sistem pakar untuk diagnosis penyakit demam berdarah dengue berbasis website. Repository Universitas Bina Sarana Informatika (RUBSI), 1–8. https://repository.bsi.ac.id/index.php/unduh/item/311052/implementasialgoritma-naive-bayes-pada-sistem-pakar-untuk-diagnosis-penyakit-demamberdarah-dengue-berbasis-website.pdf

[11] Haffandi, M. Y., Haerani, E., Syafria, F., & Oktavia, L. (2022). klasifikasi penyakit paru-paru dengan menggunakan metode naïve bayes classifier. Jurnal Teknik Informasi Dan Komputer, 5(2), 176. https://doi.org/10.37600/tekinkom.v5i2.649

[12] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). SciKit-Learn: Machine Learning in Python. HAL (Le Centre Pour La Communication Scientifique Directe). https://hal.inria.fr/hal-00650905

[13] K. A. Khalim, U. Hayati, and A. Bahtiar, "Perbandingan Prediksi Penyakit Hipertensi Menggunakan Metode Random Forest Dan Naïve Bayes," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 1, pp. 498–504, Mar. 2023, doi: 10.36040/jati.v7i1.6376

[14] S. Misra and H. Li, "Noninvasive fracture characterization based on the classification of sonic wave travel times," in Elsevier eBooks, 2020, pp. 243–287. doi: 10.1016/b978-0-12-817736-5.00009-0.

[15] S. Krishnan, "Machine learning for biomedical signal analysis," in Elsevier eBooks, 2021, pp. 223–264. doi: 10.1016/b978-0-12-813086-5.00006-2.

[16] D. Maslove, T. Podchiyska, and H. J. Lowe, "Discretization of continuous features in clinical datasets," Journal of the American Medical Informatics Association, vol. 20, no. 3, pp. 544–553, May 2013, doi: 10.1136/amiajnl-2012-000929.

[17] J. Du et al., "Developing a hypertension visualization risk prediction system utilizing machine learning and health check-up data," Scientific Reports, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-46281-y

[18] Fitri, D. R, "Diagnose Enforcement And Treatment Of High Blood Pressure".Jurnal Majority, Vol 4,no 3,https://juke.kedokteran.unila.ac.id/index.php/majority/article/download/549/550

[19] W. Setiawan and S. Ratnasari, "Sistem Pakar Diagnosis Penyakit Mata Menggunakan Naïvebayes Classifier," Setiawan | Prosiding Semnastek, Nov. 12, 2014. https://jurnal.umj.ac.id/index.php/semnastek/article/view/345

[20] A. Sabrani, W. Wedashwara, and F. Bimantoro, "Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia," Jurnal Teknologi Informasi, Komputer Dan Aplikasinya, vol. 2, no. 1, pp. 89–100, Mar. 2020, doi: 10.29303/jtika.v2i1.87.

[21] Y. E. Fadrial, "Algoritma Naive Bayes untuk mencari perkiraan waktu studi mahasiswa," Intecoms, vol. 4, no. 1, pp. 20–29, May 2021, doi: 10.31539/intecoms.v4i1.2219.

[20] A. Sabrani, W. Wedashwara, and F. Bimantoro, “Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” Jurnal Teknologi Informasi, Komputer Dan Aplikasinya, vol. 2, no. 1, pp. 89–100, Mar. 2020, doi: 10.29303/jtika.v2i1.87.