ILKOM Jurnal Ilmiah Vol. 15, No. 2, August 2023, pp. 262-270
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Naïve Bayes and K-Nearest Neighbor Approaches in Data Mining Classification of Drugs Addictive Diseases

Dadang Priyanto [a,1,*]; Ahmad Robbiul Iman [a,2]; Deny Jollyta [b,3]
[a] *Department of Computer Science, Universitas Bumigora, Ismail Marzuki Street, Mataram, 83127, Indonesia.*
[b] *Department of Informatics Engineering, Institut Bisnis Dan Teknologi Pelita Indonesia, Pekanbaru Riau, 28127, Indonesia*
[1] *dadang.priyanto@universitasbumigora.ac.id;* [2] *robbishimozuki@gmail.com;* [3] *deny.jollyta@lecturer.pelitaindonesia.ac.id*
** Corresponding author*

Article history: Received December 06, 2022; Revised January 24, 2023; Accepted June 09, 2023; Available online August 16, 2023

Abstract

Indonesia, with its very large population, is a potential market for drugs trafficking. Hence, seriousness is needed in cracking down or preventing drug trafficking. Narcotics are substances or drugs that can cause dependence or addicted and other negative impacts on users. The problem is that drug users do not realize and even ignore diseases caused by drug addiction. The diseases can be life-threatening for users, such as inflammation of the liver, heart disease, hypertension, stroke, and others. The prevalence rate of drug abuse in West Nusa Tenggara (NTB) is included in the high category, reaching 292 cases or around 37.24% cases. This study aimed to create an application that classify various diseases of drug users using the naïve bayes and KNN methods. The results of this study indicated that there was a very close relationship between drug users and various deadly diseases. The prediction results showed that the naïve bayes method provided a prediction accuracy of 94.5% while the KNN showed a prediction accuracy of 92.5%. This shows that the naïve bayes method provides superior predictive performance than the KNN in the data set of drug addicts in NTB.

Keywords: Drug Addiction; Drug Addiction Disease; Naïve Bayes; Narcotics; K-Nearest Neighbor.

Introduction

Indonesia as an archipelagic country with more than 276 million people. This is a very large population and has great potential for the marketing and distribution of narcotics. *Narkoba* in Indonesia is an abbreviation of narcotics, psychotropics and other addictive substances. The term narcotics itself is a substance or drug that is natural, synthetic, or semi-synthetic if consumed by the human, it can affect the central nervous system, namely the brain, and influence the analysis and thinking ability [1], [2]. Drug or narcotic users will be addicted to these illegal drugs. Drug users can feel excessive anxiety, increased sexual appetite, paranoia, experience delusions, behavior changes towards aggressiveness. Trafficking of narcotics is prohibited because it can cause dependence and various diseases [3], [4]. The problem is that drug users do not realize and even ignore the diseases caused by drug addiction, these diseases can be life-threatening for users, such as inflammation of the liver, heart disease, hypertension, stroke and others. The prevalence rate of drug abuse in West Nusa Tenggara is in the high category, reaching 292 cases or around 37.24% of drug abuse cases, out of a population of approximately 5.1 million people.

This research aimed to create an application that can classify various diseases of drug users based on the naïve bayes and K-Nearest Neighbor (KNN) methods. Classification is a systemic arrangement in groups according to established rules or standards. Literally classification can be said as the division of something according to classes. According to science, classification is the process of grouping objects or data based on their characteristics, similarities, and differences. Processing data with the classification of data mining processing categories, and many methods that can be used in data mining. Data mining is grouped into descriptive data mining and predictive data mining [5]. The naïve bayes and KNN methods can be used in classification. The naïve bayes classifier is based on the static bayes probability theorem which was put forward by the English scientist Thomas bayes. It improves future opportunities based on past experience with the main characteristic of a very strong (naïve) assumption of the independence of each condition or event. [6]–[8]. The KNN method is an algorithm used to classify an object, based

http://dx.doi.org/10.33096/ilkom.v15i2.1544.262-270

on the k training data that is closest to the object. The condition for the value of k is that it cannot be greater than the number of training data, and the value of k must be odd and more than one. The distance of an object determine its classification, a close distance of an object classified in one group by calculating using the Euclidean Distance [9], [10].

Previous studies have used a variety of data mining methods such as disease diagnosis by using the fuzzy, naïve bayes and KNN comparative methods for smallpox, dengue fever and flu [11]. Another study using the naïve bayes and KNN methods for liver disease classification, which gave the best performance was the KNN algorithm [12], another comparison of classification algorithms in coronary heart disease showed that the random forest algorithm provided better performance in terms of accuracy [13] , then a comparative analysis of the classification method for Hepatitis shows that the naïve bayes method has better performance [14], Other comparative studies of data mining algorithms on medical data performance Support Vector Machine (SVM) provide the best accuracy [15], and the use of genetic algorithms and machines learning in the classification and prediction of heart disease [16].

Previous studies have shown that various classification methods provide different best performance results based on the data set used, therefore this study provides a novelty in classification using the naïve bayes and KNN methods on drug addict disease data sets in West Nusa Tenggara region. From the results of the comparison showed the performance of the best method which differed in accuracy from previous studies.

Method

The data mining methods used in this study were a classification category and a grouping of an object into a particular class. Various issues related to grouping objects can be solved by applying classification techniques. In general, classification performance was carried out using the confusion matrix, and this study used the naïve bayes and KNN methods.

A. Naïve Bayes method

Naïve bayes is a simple classifying probabilistic method based on Bayes' theorem where classification is carried out through training sets of a number of data efficiently [17]. Naïve Bayes assumes that the value of an input attribute in a given class does not depend on the values of other attributes [18]. Bayes' theorem itself was put forward by British scientist Thomas bayes, namely predicting future opportunities based on previous experience so that it is known as Bayes' theorem. Where the Bayes theory Equation 1 is.

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

Where :

X	: Data with unknown class
C	: Hypothesis data X is a specific class
$P\ (C\|X)$	: Probability of hypothesis C based on condition X (posterior probability)
$P(C)$	: Probability of hypothesis C (prior probability)
$P(X\|C)$	: Probability of X based on conditions in hypothesis C
$P(X)$	: Probability of X

B. K-Nearest Neighbor method

KNN is a method using a supervised learning algorithm in which a number of new attributes whose class is unknown can be searched for by looking for similarities based on the majority of KNN as a reference in determining the class of an attribute [19]–[21]. The KNN method can be calculated using the Euclidean distance in determining the distance between datasets. The Euclidean distance formula can be seen in Equation 2.

$$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + \cdots + (N_1 - N_2)^2} \quad (2)$$

Where:

d : distance between the two attributes
X_1 : actual data
X_2 : data testing
Y_1 : nth actual data
Y_2 : nth testing data
N_1 : first nth attribute
N_2 : second nth attribute

The flowchart of the calculation process using the K-NN method can be seen in **Figure 1**.

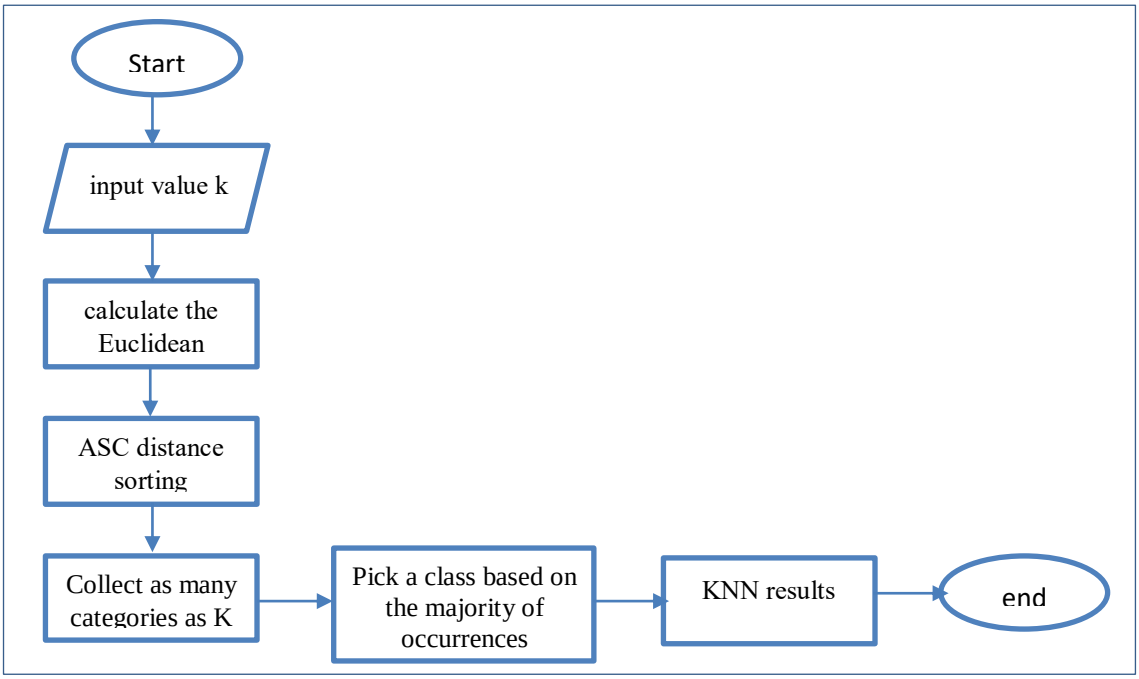

Figure 1. Flowchart of K-Nearest Neighbor

C. Data Set

This study used data taken from the National Narcotics Agency of West Nusa Tenggara Province(BNNP), with data collection using interview techniques and in the form of data files. From the results of the interviews, information was obtained about the symptoms experienced by drug addicts, the types of drugs consumed, the level of addiction, and history of drug addict disease. The data obtained from the results of a survey regarding the types of drug addict disease obtained 336 data, and can be partially presented in tabular form as in **Table 1**.

Table 1. Drug Addiction Disease Dataset Table

Drug Type	Addiction Level	G001	…	G024	Disease History	Disease
Cocaine	Recreational	Yes	…	No	Stomach Inflammation	GERD
Marijuana	Trying	Yes	…	Yes	Heart disease	Hepatitis
Crystal Meth	Active dependency	No	…	No	Asthma	Acute Asthma
Morfin	Trying	Yes	…	No	Hypertension	Cyst
Gorilla tobacco	Recreational	Yes	…	No	Malaria	Thypoid
…	…	…	…	…	…	…
Gorilla tobacco	Recreational	Yes		Yes	Diabetes	Hepatitis

From **Table 1** it can be explained the description of the specified symptom variables as in **Table 2** below.

Table 2. Description of symptom variables

Symptoms Code	Symptoms
G001	Nausea and vomiting
G002	Sore throat
G003	Expulsion of stomach contents involuntarily
G004	Bad breath
G005	Coughs
G006	Hard to breathe
G007	Hoarseness
G008	Speech disorder
G009	Hard to sleep
G010	Heart rate increases
G011	Dizzy
G012	Easily Tired
G013	Chest feels tight
G014	Redness of the skin around the cyst area
G015	Blood or pus that smells bad from the lump
G016	Fever

Symptoms Code	Symptoms
G017	Mild or severe pain in the lower abdomen
G018	Stomach ache
G019	Loss of appetite
G020	Joint pain
G021	Diarrhea
G022	Dry Cough
G023	Jaundice
G024	Dark urine

Results and Discussion

A. Preprocessing

Before processing data with the Algorithm, pre-processing was a must. This was to make the data is easier to read, reduce the burden of representation in the data, reduce the duration of data processing, and simplify the process of data analysis by the system. The dataset that was formed must be a dataset that was free from missing values and inconsistent attributes so there would not be interference with the process at the modeling stage. The initial stage of preprocessing was by initializing attributes using the variables "X_train" and "Y_train" as class labels. In the manual calculation of naïve bayes and KNN the attribute must have a numeric value. This is because naïve bayes and KNN utilize Probability and statistical calculations. So the data preprocessing stage for the naïve bayes and KNN algorithms is done by changing the value of the numeric "X_train" attribute. The results of the preprocessing for the naïve bayes and KNN algorithms can be partially seen in Table 3 below.

Table 3 Preprocessing Results

Drugs	Addicted	G001	…	G024	History	Disease
0.84	1.29	1	…	-1	0.25	GERD
0.18	0.06	1	…	1	1.75	Hepatitis
1.54	1.18	-1	…	-1	0.72	Acute Asthma
0.48	0.06	1	…	-1	0.96	Cyst
1.15	1.29	1	…	-1	0.35	Thypoid
…	…	…	…	…	…	…
1.54	1.29	1	…	-1	0.96	Cyst
0.48	0.06	-1	…	-1	0.81	Acute Asthma
1.15	1.29	1	…	1	0.81	Hepatitis

Based on Table 3, the results of data processing describe the condition of the disease based on the type of drug used, the level of addiction, the symptoms experienced ranging from 1 to 24, as well as a history of previous illnesses. For example, record data set 1 indicates that the type of addiction was 0.84, the type of drug used was 'Cocaine', an addiction level was 1.29 categorized as 'Recreational' addiction, and symptoms that were felt according to code G001 were 'Nausea and Vomiting' and so on until symptom 24. If it was combined with history of previous illness it can be concluded the disease was 'GERD'. The same thing can be done in reading the next record in Table 3 above. Information about the attributes used in reading Table 3 can be seen with the attributes of the type of drug, the level of addiction, symptoms of the disease and history of the disease which can be explained in the following Table 4-7.

Table 4. Description of Drug Type Attributes

Initial Attribute Value	Numerical Attribute Values
Crystal meth	1.54
Cocaine	0.84
Morphine	0.48
Gorilla Tobacco	1.15
Marijuana	0.18

Table 4 Provides information regarding the attribute values for the drug types used and numerical values according to the types of drugs used. For example, narcotics of the methamphetamine type have a numerical attribute value of 1.54 and so on.

Table 5. Addiction Level Attribute Description

Initial Attribute Value	Numerical Attribute Values
Try Use	0.06
recreational	1.29
Active Dependence	1.18

Table 5 Provides a description of the level of addiction experienced by users such as a value of 0.06 indicating addiction at the 'Trying' stage and so on.

Table 6. Symptom Description

Initial Attribute Value	Numerical Attribute Values
Yes	1
Not	2

Table 6 Describes information on the symptoms experienced by addicts with a determination value of 1 'Yes' experience' and 2 'No' experience', while the symptoms experienced refer to Table 2 above.

Table 7. Description of Disease History

Initial Attribute Value	Numerical Attribute Value
Asthma	0.72
Stomach Inflammation	0.25
Heart disease	1.75
Malaria	0.35
Pneumonia	1.66
Tuberculosis	0.44
Diabetes	0.81
Hypertension	0.96
There isn't any	0

Table 7 describes the previous history of the disease experienced by the addict as an additional variable in determining the conclusion of the illness.

B. Calculation results of the K-Nearest Neighbor method

The results of calculations with the KNN from the amount of existing data, 10 testing data were needed to be calculated with the results as in Table 8.

Table 8. Data Testing

Drug Type	Addiction Level	G001	…	G024	Disease History	Disease
0.48	1.18	1	…	-1	0	Thypoid
1.54	1.29	1	…	-1	1.75	GERD
1.15	0.06	-1	…	-1	0.96	Hepatitis
0.48	1.29	-1	…	-1	0	GERD
0.84	1.29	1	…	-1	0.81	Cyst
0.18	0.06	-1	…	1	1.75	Cyst
0.18	1.29	1	…	-1	0.81	Acute Asthma
1.15	0.06	1	…	1	1.75	Hepatitis
1.15	1.29	-1	…	-1	1.75	Cyst
1.54	0.06	1	…	-1	0.35	GERD

As seen in Table 8, it can be concluded that the results of testing from 10 data, with varying levels of disease experienced by addicts such as typhoid, GERD, and others according to the drugs consumed and their addiction level. From table 8 it can be calculated to determine the closest majority *K* value similarity or based on the distance between data sets with the Euclidean distance formula and the results were shown in Table 9 .

Table 9. Euclidean distance result

Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
0.677495387	1.18156676	1.64693655	1.834557167	1.4066272	1.731964203	0.43104524	1.614218077	1.0694391	1.33574
1.094257739	1.34331679	1.52715422	1.698293261	1.5270887	2.021113554	0.67749539	1.771242502	0.8258935	1.22793
1.923174459	0.88977525	1.09535382	0.849352695	0.8713208	2.141144554	1.11919614	1.345696846	0.5374942	0.47223
1.718924082	0.44922155	1.83338485	1.697910481	0.4114608	2.255393536	0.94175368	1.67002994	0.9393615	0.54452
1.479932431	1.70120545	0.66272166	1.04517941	1.8645643	1.638566447	1.21268298	1.315028517	1.1157509	1.59706
1.706399719	0.27037012	1.53590364	1.3713497	0.2014944	2.021014597	0.91869473	1.336674979	0.868159	0.50359
1.263566381	1.59176003	1.56626945	1.762498227	1.7652762	2.167694628	0.93648278	1.934760967	0.9404786	1.39893
0.992219734	1.22711043	1.16357209	1.380543371	1.4259383	1.683627037	0.55506756	1.401784577	0.7598684	1.19537
1.158015544	1.3095419	1.66679333	1.824883558	1.4974645	0.671490879	1.23798223	0.820548597	1.8714967	1.91348

Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
2.067147793	1.10498869	0.93193347	0.624579859	1.0901835	1.801887899	1.43405021	0.872811549	1.1983322	1.13723
0.517010638	1.05763888	1.42369238	1.637101096	1.3042622	1.288953064	0.37802116	1.206067991	1.1540364	1.38351
0.933380951	1.31244047	1.20307938	1.44934468	1.5182226	1.65562677	0.60712437	1.434747365	0.8673523	1.30407
1.741895519	0.53037722	1.77476759	1.634441801	0.4987986	2.298543017	0.93455872	1.693664666	0.8190849	0.41833
1.293290377	1.40730238	0.74040529	1.04890419	1.5836666	0.881873007	1.1319894	0.516623654	1.3944533	1.6533
…	…	…	…	…	…	…	…	…	…
1.443502684	1.64972725	0.52316345	0.962756459	1.8177184	1.251958466	1.257417989	0.940638081	1.3234047	1.70678

From **Table 9** the selection of the smallest distance value that was included in the ranking of a number of *K* values is shown. In **Table 9** the value of *K* is 10, so you have to find the smallest distance value of 10. Furthermore, the results of the highest number of classes from a predetermined *K* value can be obtained as predicted in **Table 10** below.

Table 10. The highest number of classes from the value of *K*

Nth Test Data	Original Class	Acute Asthma	Cyst	GERD	Hepatitis	Typoid
1	Thypoid	False	False	False	False	True
2	Hepatitis	True	False	False	False	False
3	Asma Akut	False	False	False	True	False
4	Kista	False	False	False	True	False
5	Asma Akut	True	False	False	False	False
6	Hepatitis	False	False	False	False	False
7	Kista	True	False	False	False	False
8	Asma Akut	False	False	False	False	False
9	GERD	False	False	True	False	False
10	GERD	False	False	True	False	False

Table 10 explains that the test results indicate a 'True' value, so addicts are diagnosed with a disease according to the predicted results as presented in **Table 9** test results.

C. Calculation results with Naive Bayes

The calculation process with the naïve bayes method was started by retrieving training data from the data that had been obtained. The variables that were used in the classification of drug addicts were disease, drug types, level of drug addiction, illness records, and the results of the calculation showed that the average probability of the disease class was 20% while for the probability of drug type attributes as shown in **Table 11**.

Table 11. Attribute Probability of Drug Types

Drug Type	Acute Asthma	GERD	Typhoid	Cyst	Hepatitis
Crystal meth	10%	30%	30%	30%	10%
Cocaine	50%	20%	20%	20%	20%
Morphine	0%	10%	0%	10%	10%
Gorilla Tobacco	30%	30%	40%	10%	0%
Marijuana	10%	10%	10%	30%	60%
Total number	100%	100%	100%	100%	100%

Table 11. Attribute Probability of Drug Types

Addiction Disease	Acute Asthma	GERD	Typhoid	Cyst	Hepatitis
Recreational	40%	30%	20%	30%	40%
Active dependency	0%	30%	50%	30%	40%
Try it on	60%	40%	30%	40%	20%
Total	100%	100%	100%	100%	100%

The resulting model was evaluated with the aim of assessing the performance of the naïve bayes algorithm and the KNN algorithm. Evaluation was conducted by calculating the accuracy, precision, and recall of each algorithm used using the Confusion Matrix Technique. If the evaluation results showed good performance, then the two algorithms were feasible in making predictions on new data whose labels were unknown.

D. Testing the Naïve Bayes

At this stage the test was carried out on the performance of the naïve bayes classification model made when making predictions on data testing, using the confusion matrix and the confusion matrix plot. The test results can be seen in **Figure 2** of the confusion chart.

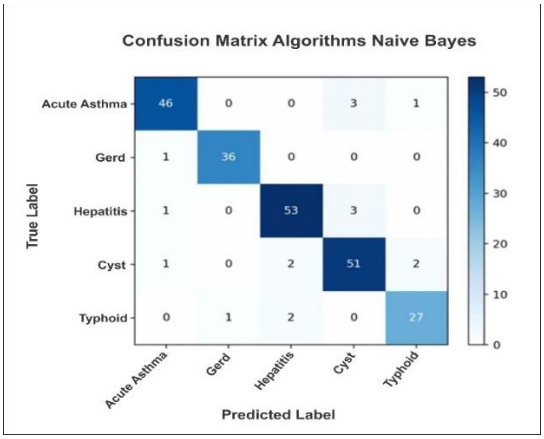

Figure 2. Naive Bayes Confusion Matrix Graph

In the contusion matrix f above, it can be seen that the prediction for the Acute Asthma class has a true value of 46, the prediction for the GERD has a value of 36, the prediction for the Hepatitis has a value of 53, the prediction for the class of Cyst is 51, and the prediction for the Typhoid class with a value of 27 , wrong prediction in the Acute Asthma class by 4, GERD class by 1, Hepatitis class by 4, Cyst class by 2, and Typhoid class by 3. The results of the calculation of the Naive Bayes confusion matrix can be seen in **Table 12**.

Table 12. Confusion Matrix Naïve Bayes

		Acute Asthma	Gerd	Typhoid	Cyst	Hepatitis
Predictions	**Acute Asthma**	46	0	0	3	1
	GERD	1	36	0	0	0
	Hepatitis	1	0	53	3	0
	Cyst	1	0	2	51	0
	Typhoid	0	1	2	0	27

Based on **Table 12**, prediction accuracy can be calculated using the naïve bayes algorithm which produces a prediction accuracy of 93% as in the following Equation 3 calculation.

$$\text{Prediction Accuracy} = \frac{46+36+53+51+27}{230} = 0{,}93\% \quad (3)$$

E. Comparison Results of the Naïve Bayes and K-Nearest Neighbor Methods

This test was carried out by dividing training data and data testing into 4 variations of data, and testing was carried out on naïve bayes and KNN modeling on variations in the percentage of training data and testing data. The results of testing with the confusion matrix can be seen in the comparison of accuracy, precision and recall of the two algorithms as shown in **Table 13**.

Table 13. Results of Comparison of Accuracy, Precision and Recall

Data *Training*	Data *Testing*	Accuracy (%)		Precision (%)		Recall (%)	
		Naïve Bayes	*K-NN*	*Naïve Bayes*	*K-NN*	*Naïve Bayes*	*K-NN*
335	201	93	93	93	93	93	93
335	180	96	90	95	90	95	90
335	230	93	91	93	91	93	91
335	181	96	96	96	96	96	96
Aaverage		94.5	92.5	94.25	92.5	94.25	92.5

Based on **Table 13**, the calculations after being validated showed that the naïve bayes method had a higher accuracy than the KNN, with a value of 94.5%, while KNN after being validated had an accuracy value of 92.5%.

Conclusion

Narcotics are drugs that are prohibited from being widely distributed, because they can lead to addiction for their users and can cause various deadly diseases. This research was conducted to classify various diseases of Narcotic addicts in the form of application programs using two methods, namely naïve bayes and KNN. This research has shown that there was a very close relationship between drug users and various deadly diseases. The prediction results showed that the naïve bayes method provided a prediction accuracy of 94.5% while the KNN showed a prediction accuracy of 92.5%. This shows that the naïve bayes method provides better predictive performance than the KNN with the data set used in this study.

References

[1] B. Priambodo, Y. Jumaryadi, and U. Salamah, "Penggunaan Aplikasi Deteksi Pecandu Narkoba Di Meruya Utara," *J. Pasopati Pengabdi. Masy. dan Inov. Pengemb. Teknol.*, vol. 4, no. 3, 2022, doi: 10.14710/pasopati.2022.14468.

[2] M. Mustaqim, H. Dafitri, and D. Dharmawati, "Edukasi Digital Pengenalan Bahaya Narkoba Bagi Anak Usia Dini Berbasis 3D Dan Augmented Reality," *Djtechno J. Teknol. Inf.*, vol. 2, no. 2, pp. 170–176, 2021, doi: 10.46576/djtechno.v2i2.1623.

[3] B. C. Laksono and N. W. K. Projo, "Pemodelan Analisis Rantai Markov untuk Mengestimasi Potensi Kasus Narkoba di Indonesia," *Semin. Nas. Off. Stat.*, vol. 2021, no. 1, pp. 715–722, 2021, doi: 10.34123/semnasoffstat.v2021i1.1016.

[4] A. Sinjar and T. Sahuri, "Bahaya Narkoba Terhadap Masa Depan Generasi Muda," vol. 2, no. 2, pp. 154–160, 2021.

[5] D. Priyanto, M. Zarlis, H. Mawengkang, and S. Efendi, "Analysis of earthquake hazards prediction with multivariate adaptive regression splines," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2885–2893, 2022, doi: 10.11591/ijece.v12i3.pp2885-2893.

[6] H. Hairani, G. S. Nugraha, M. N. Abdillah, and M. Innuddin, "Komparasi Akurasi Metode Correlated Naive Bayes Classifier dan Naive Bayes Classifier untuk Diagnosis Penyakit Diabetes," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 3, no. 1, pp. 6–11, 2018, doi: 10.30743/infotekjar.v3i1.558.

[7] Y. F. Safri, R. Arifudin, and M. A. Muslim, "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor," *Sci. J. Informatics*, vol. 5, no. 1, p. 18, 2018, doi: 10.15294/sji.v5i1.12057.

[8] R. M. Abarca, "Sistem Pakar Diagnosis Penyakit Mata," *Nuevos Sist. Comun. e Inf.*, no. sistem pakar penyakit mata, pp. 2013–2015, 2021.

[9] A. H. Bisri Merluarini, Diah Safitri, "Perbandingan Analisis Klasifikasi menggunakan Metode K-Nearest Neighbor (K-Nn) dan Multivariate Adaptive Regression Spline (Mars) pada Data Akreditasi Sekolah Dasar Negeri di Kota Semarang," vol. 3, pp. 313–322, 2014.

[10] P. I. Lestari, D. E. Ratnawati, and L. Muflikhah, "Implementasi Algoritme K-Means Clustering Dan Naive Bayes Classifier Untuk Klasifikasi Diagnosa Penyakit Pada Kucing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 968–973, 2019.

[11] A. Rajkumar, "Comparison of Fuzzy Diagnosis with K-Nearest Neighbor and Naïve Bayes-2010.pdf," *Glob. J. Comput. Sci. Technol.*, vol. 10, no. 10, pp. 38–43, 2010.

[12] A. Desiani, "Perbandingan Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Klasifikasi Penyakit Hati," *Simkom*, vol. 7, no. 2, pp. 104–110, 2022, doi: 10.51717/simkom.v7i2.96.

[13] A. B. Wibisono and A. Fahrurozi, "Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr.2019.v24i3.2393.

[14] S. Sulastri, K. Hadiono, and M. T. Anwar, "Analisis Perbandingan Klasifikasi Prediksi Penyakit Hepatitis Dengan Menggunakan Algoritma K-Nearest Neighbor, Naïve Bayes Dan Neural Network," *Dinamik*, vol. 24, no. 2, pp. 82–91, 2020, doi: 10.35315/dinamik.v24i2.7867.

[15] A. Suragala, P. Venkateswarlu, and M. China Raju, "A Comparative Study of Performance Metrics of Data Mining Algorithms on Medical Data," *Lect. Notes Electr. Eng.*, vol. 698, no. February 2021, pp. 1549–1556, 2021, doi: 10.1007/978-981-15-7961-5_139.

[16] V. Biksham, V. Srujana, I. Meghana, B. Harshath, and G. Tarun, "Heart Disease Prediction Using Machine Learning," *Ymer*, vol. 21, no. 4, pp. 489–494, 2022, doi: 10.37896/YMER21.04/48.

[17] E. Wijaya, "Implementation Analysis of GLCM and Naive Bayes Methods in Conducting Extractions on Dental Image," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 407, no. 1, 2018, doi: 10.1088/1757-899X/407/1/012146.

[18] J. N. Peksi, B. Yuwono, and Y. M. Florestiyanto, "Classification of Anemia with Digital Images of Nails and Palms using the Naive Bayes Method," *J. Inform. dan Teknol. Inf.*, vol. 18, no. 1, pp. 118–130, 2021, doi: 10.31515/telematika.v18i1.4587.

[19] P. Chalekar, S. Shroff, S. Pise, and S. Panicker, "Use of K-Nearest Neighbor in Thyroid Disease Classification," *Int. J. Curr. Eng. Sci. Res.*, vol. 1, no. 2, pp. 2394–0697, 2014.

[20] J. M. Martínez-Otzeta and B. Sierra, "Analysis of the iterated probabilistic weighted k nearest neighbor method, a new distance-based algorithm," *ICEIS 2004 - Proc. Sixth Int. Conf. Enterp. Inf. Syst.*, no. January, pp. 233–240, 2004, doi: 10.5220/0002605402330240.

[21] A. K. Saputro, K. A. Wibisono, and F. P. Pratiwi, "Identification of Disease Types on Tea - Plant Varieties Based Image Processing with K-Nearest Neighbor Method," *J. Phys. Conf. Ser.*, vol. 1569, no. 3, 2020, doi: 10.1088/1742-6596/1569/3/032078.