Classification for Determining the Level of Drugs Dependence Using the Naïve Bayes Classifier

Novianti Puspitasari[1], Muhammad Ajay[2], Masna Wati[3], Anindita Septiarini[4]
Department of Informatics, Faculty of Engineering, Mulawarman University, Indonesia[1,2,3,4]
novipuspitasari@unmul.ac.id[1], muhammadajay093@gmail.com[2], masnawati@fkti.unmul.ac.id[3], anindita@unmul.ac.id[4]

Article Info

Article history:

Received Feb 11, 2024
Revised Jul 9, 2024
Accepted Dec 14, 2024

Keyword:

Classification
Confusion Matrix
Drugs
Level
Naïve Bayes Classifier

ABSTRACT

Drug users or abusers are people who use narcotics or psychotropic drugs without supervision or medical indication from a doctor. Before undergoing rehabilitation, drug users must first undergo an examination to determine their level of drug dependence so that they can receive medical treatment according to their level of drug dependence. Determining the level of drug dependence requires a technique that can provide labels or categories of data for drug users based on the user's condition or influential criteria. This study applies the Naïve Bayes Classifier method to a system to determine the level of drug dependence. This study uses medical record data from 220 drug users. The user's medical record data is processed using data mining stages consisting of data selection, data cleaning, data transformation, and division of training and test data to produce 120 training data and 100 test data. The results of the Naive Bayes Classifier method calculation resulted in 29 users having a trial level of dependence (mild), 42 identified as having a regular level of dependence (moderate), and 29 others as users with a severe level of dependence. The confusion matrix testing was very accurate, namely, 94% accuracy, 95% precision value, and 92% recall. Meanwhile, the system that has been built can run very well. Based on the results of the research that has been conducted, this research can contribute to determining the level of dependence of drug addicts objectively so that related parties can provide rehabilitation or appropriate treatment to drug addicts.

Corresponding Author:

Novianti Puspitasari
Department of Informatics, Faculty of Engineering
Mulawarman University
Sambaliung Street Number 9 Kampus Gunung Kelua, Samarinda 75119, Indonesia
Email: novipuspitasari@unmul.ac.id

1. INTRODUCTION

Drug addiction or dependence is still a problem experienced by many developing countries and a significant problem for public health. On the health side, drug addiction has been linked to an increased risk of cancer, psychological complications, heart, liver, and lung diseases, and infections [1]. Drug users are individuals who use narcotics or psychotropic substances without medical indications and are not under the supervision of a doctor. Drug users are pathological (causing abnormalities) and create obstacles in carrying out daily activities [2], so they can easily destroy the life of a nation [3]. It is based on the BNN Puslitdatin report regarding the 2019 prevalence survey of drug abuse in 13 provinces in Indonesia; it was found that large cities such as Surabaya, Yogyakarta, Bandung, Medan, and Samarinda had the largest prevalence of drug abuse [4]. From the survey results, Samarinda ranks first with a percentage of 60%, followed by Balikpapan with a rate of 20%, and the rest are other areas in the province of East Kalimantan [5]. From January 2013 to March 2017, 632 recorded drug users attended rehabilitation and inpatient treatment at the National Narcotics Agency Rehabilitation Center in Samarinda City [6]. The rehabilitation program is one of the government's efforts to overcome the problem of drug abuse [7]. Before undergoing rehabilitation, drug users must first undergo an examination to determine their level of drug dependence so that they receive medical treatment according to the level of drug dependence. The level of drug dependence is classified into three categories, namely, trying to use (light category), regularly using (medium category), and people with an addiction (severe category) [8]. Determining the category of the level of drug dependence in drug users must require appropriate methods so that drug users receive appropriate rehabilitation programs. So far, the East Kalimantan Province National Narcotics Agency is still determining the level of drug dependence using simple methods. Therefore, in assessing the level of drug dependence, a technique is needed to determine labels or classes from drug user data based on the user's condition or influential criteria, and a system is also required that can process the requirements considered in determining the level of drug dependence. Based on existing problems, this research aims to assess the level of drug dependence by applying classification techniques in data mining using the Naïve Bayes Classifier method.

The Naïve Bayes Classifier method is one method of classification techniques [9], [10]. The Naïve Bayes Classifier is known to be better than several other classification methods; this is due to several reasons. First, the main characteristic of Naïve Bayes is the very strong (naive) independence assumption of each condition or event. Second, the model is simple and easy to make. Third, the model can be implemented for large data sets and has good accuracy [11]. The classifier from the Naïve Bayes method also has significant advantages compared to other methods, namely that it only requires a small-scale training data set for estimation [12]. Many studies have been carried out regarding classification methods that apply the Naïve Bayes method, including research classifying the level of internet addiction in students with an accuracy rate of 90% and can be said to be excellent [13]. This method also obtained an accuracy of 92.31% with an error rate of 7.69% for classifying student personalities [14].

Furthermore, other research concluded that the Naïve Bayes method could obtain classification results with a high accuracy value of 98.6%, a precision value of 98.6%, and a recall value of 89.5% [15]. Naïve Bayes also produces high accuracy values to evaluate company security performance with a perfect accuracy value of 100%, so this method is categorized as an appropriate method for solving problems [16]. This statement is in line with the results of research [17] which has compared the Naïve Bayes and K-Nearest Neighbor (K-NN) methods. The research results on both methods show that Naïve Bayes produces more accurate classification with an accuracy value of 89.04% compared to K-NN of 87.67%. The high accuracy value of Naïve Bayes in carrying out classification means that this method has been applied in various studies [18]–[23].

From various studies that have been conducted with Naïve Bayes, it can be seen that this method is still rarely used for classifying the level of drug addicts. That is because the classification of the level of drug addicts is generally determined subjectively. This study provides more contribution to help the government to be more targeted in providing treatment and rehabilitation to drug addicts. Furthermore, the application of Naïve Bayes in software will make it easier for stakeholders to determine the level of drug addiction through identification based on criteria or symptoms experienced by drug users objectively. Ultimately, the findings of this study have an impact on the health and social fields to determine the exact level of drug addicts so that the government can provide appropriate rehabilitation and reduce the number of drug addicts.

2. RESEARCH METHOD

This research on drug dependence levels uses data mining stages to process raw data into research data ready to be used. Next, the data is classified using the Naïve Bayes Classifier method and applied to the system.

2.1. Data Mining Stages

The stages in data mining consist of [24]:

1) Data Selection

The data selection stage is carried out on medical record data of drug users. Drug user medical record data initially had eleven attributes, namely the user's name or initials, gender, age, occupation, frequency of drug use, method of drug use, duration of drug use, psychological symptoms, number of types of substances used, alcohol screening score, screening score. Drugs and level of dependency. After going through the data selection process, criteria that do not influence the determination of the level of drug dependence, such as gender, age, and occupation, are ignored so that the results of the data selection process leave seven attributes that influence the level of drug dependence, namely the criteria for frequency of drug use, method of drug use, duration of drug use, amount of drug use, psychological symptoms of drug users, alcohol screening score, and drug screening score. These seven criteria are the criteria for training data and testing data. The seven criteria can be seen in Table 1.

Table 1. Criteria for Determining the Level of Drug Dependence

No	Criteria	Sub Criteria
1	Frequency of drug use	≤4 times/month ≤12 times/month ≥13 times/month
2	Ways of Using Drugs	Inhaled/smoked Swallowed Smoke Inject
3	Duration of drug use	≤1 year ≤5 years ≥6 years
4	Psychological Symptoms	There isn't any Difficulty remembering Anxiety Frequent hallucinations Depression
5	Number of types of substances used	One type of drug Two Types of Drugs Three Types of Drugs Four Types of Drugs
6	Narcotics Screening Score	0-3 (no intervention) 4-26 (brief intervention) ≥27 (intensive treatment)
7	Alcohol Screening Score	0-10 (no intervention) 11-26 (brief intervention) ≥27 (intensive treatment)

2) Data Cleaning

Data cleaning is a basic action like noise removal. In this process, duplicate and empty data are deleted, then checked for inconsistent data, and errors in the data are corrected. In drug user data from the data cleaning process carried out on 220 data, there were no duplicate or empty data, so no data needed to be deleted or cleaned.

3) Transformation

In this research, the transformation process was carried out on several attributes or criteria from the data set that the data selection and cleaning process had carried out. The transformation process for several attributes or criteria can be seen in Table 2.

Table 2. Transformation of Research Criteria

No	Criteria	Sub Criteria	Transformation
1	Frequency of drug use	≤4 times/month	Small dose
		≤12 times/month	The dose is quite high

No	Criteria	Sub Criteria	Transformation
		≥13 times/month	High dosage
2	Duration of drug use	≤1 year	New user
		≤5 years	Long enough
		≥6 years	Long
3	Narcotics Screening Score	0-3 (no intervention)	No intervention
		4-26 (brief intervention)	Brief intervention
		≥27 (intensive treatment)	Intensive treatment
4	Alcohol Screening Score	0-10 (no intervention)	No intervention
		11-26 (brief intervention)	Brief intervention
		≥27 (intensive treatment)	Intensive treatment

4) Modelling
Data mining algorithms are selected to solve problems using various techniques, methods, or algorithms in the modeling stage. This research uses the Naïve Bayes Classifier to determine the level of drug dependence based on predetermined criteria based on a data set of drug users.
5) Evaluation
The evaluation process in this research was carried out by analyzing the modeling results that had been carried out to measure or calculate the level of accuracy of the Naïve Bayes Classifier method in determining the level of drug dependence. Measuring the level of accuracy of the process uses the confusion matrix technique.

2.2. Research Data

The research uses medical record data of drug users from 2019 to 2020, totaling 220 data as in Table 3.

Table 3. Drug User Data

Drug Users	Frequency of Use	Method of Use	Duration of Use	Psychological Symptoms	Number of types of substances	Alcohol Screening Score	Narcotics Screening Score	Level
User 1	4 times/month	smoke	6 months	There isn't any	one kind	0 (no intervention)	7 (short intervention)	Try using it (lightly)
User 2	13 times/month	smoke	7 years	frequent hallucinations	two types	20 (short intervention)	27 (intensive treatment)	Addict (heavy)
User 3	8 times/month	smoke	4 years	difficulty remembering	two types	10 (no intervention)	16 (brief intervention)	Regular use (moderate)
User 4	30 times/month	inject	13 years old	frequent hallucinations	one kind	0 (no intervention)	31 (intensive treatment)	Addict (heavy)
User 5	10 times/month	smoke	5 years	feeling anxious	two types	15 (brief intervention)	20 (short intervention)	Regular use (moderate)
User 6	8 times/month	inject	3 years	feeling anxious	two types	10 (no intervention)	15 (brief intervention)	Regular use (moderate)
User 7	10 times/month	smoke	3 years	feeling anxious	two types	15 (brief intervention)	20 (short intervention)	Regular use (moderate)
User 8	7 times/month	smoke	4 years	feeling anxious	one kind	0 (no intervention)	20 (short intervention)	Regular use (moderate)
User 9	10 times/month	smoke	4 years	feeling anxious	two types	10 (no intervention)	22 (short intervention)	Regular use (moderate)

Drug Users	Frequency of Use	Method of Use	Duration of Use	Psychological Symptoms	Number of types of substances	Alcohol Screening Score	Narcotics Screening Score	Level
User 10	3 times/month	smoke	1 year	There isn't any	one kind	0 (no intervention)	10 (short intervention)	Try using (light)
:	:	:	:	:	:	:	:	:
User 210	3 times/month	Smoke	3 years	Difficulty Remembering	One Kind	0 (No Intervention)	10 (Brief Intervention)	Regularly Use (Moderate)
User 211	10 times/month	Smoke	3 years	Difficulty Remembering	One Kind	0 (No Intervention)	20 (Brief Intervention)	Regularly Use (Moderate)
User 212	4 times/month	Inhaled	4 years	There isn't any	Two Types	12 (Brief Intervention)	12 (Brief Intervention)	Regularly Use (Moderate)
User 213	1 times/month	Smoke	1 year	There isn't any	Two Types	2 (No Intervention)	6 (Brief Intervention)	Try Using (Lightly)
User 214	15 times/month	Smoke	17 years	Frequent hallucinations	Two Types	23 (Brief Intervention)	33 (Intensive Treatment)	Addict (Severe)
User 215	5 times/month	Inhaled	4 years	Frequent hallucinations	Two Types	5 (No Intervention)	5 (Brief Intervention)	Regularly Use (Moderate)
User 216	5 times/month	Swallowed	3 years	Anxiety	One Kind	0 (No Intervention)	5 (Brief Intervention)	Regularly Use (Moderate)
User 217	6 times/month	Smoke	1 year	Anxiety	One Kind	0 (No Intervention)	3 (No Intervention)	Regularly Use (Moderate)
User 218	14 times/month	Inject	17 years	Depression	Two Types	30 (Intensive Treatment)	30 (Intensive Treatment)	Addict (Severe)
User 219	12 times/month	Swallowed	2 years	Difficulty Remembering	Two Types	5 (No Intervention)	5 (Brief Intervention)	Regularly Use (Moderate)
User 220	10 times/month	Inhaled	3 years	Anxiety	Three Types	11 (Brief Intervention)	11 (Brief Intervention)	Regularly Use (Moderate)

From the research data contained in Table 3, the data was then divided into training data and testing data. The training data in this research amounted to 120 data, and 100 data became testing data. The testing data for this research can be seen in Table 4.

Table 4. Drug User Testing Data

Drug Users	Frequency of Use	Method of Use	Duration of Use	Psychological Symptoms	Number of types of substances	Alcohol Screening Score	Narcotics Screening Score	Level
User 1	Small Dose	Swallowed	Long enough	There isn't any	Two Types	No Intervention	No Intervention	?
User 2	High dosage	Smoke	Long enough	Depression	Two Types	Intensive Treatment	Intensive Treatment	?
User 3	Quite High Dosage	Inhaled	Long	Anxiety	Two Types	No Intervention	Brief Intervention	?

Drug Users	Frequency of Use	Method of Use	Duration of Use	Psychological Symptoms	Number of types of substances	Alcohol Screening Score	Narcotics Screening Score	Level
User 4	Small Dose	Smoke	New User	There isn't any	One Kind	No Intervention	No Intervention	?
User 5	Small Dose	Inhaled	Long enough	Difficulty Remembering	One Kind	No Intervention	No Intervention	?
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
User 96	Quite High Dosage	Swallowed	Long enough	Anxiety	One Kind	No Intervention	Brief Intervention	?
User 97	Quite High Dosage	Smoke	New User	Anxiety	One Kind	No Intervention	No Intervention	?
User 98	High dosage	Inject	Long	Depression	Two Types	Intensive Treatment	Intensive Treatment	?
User 99	Quite High Dosage	Swallowed	Long enough	Difficulty Remembering	Two Types	No Intervention	Brief Intervention	?
User 100	Quite High Dosage	Inhaled	Long enough	Anxiety	Three Types	Brief Intervention	Brief Intervention	?

2.3 Naïve Bayes Classifier

Naïve Bayes Classifier performs classification based on Bayes' theorem by calculating a set of probabilities and then adding up the frequencies and combinations of values from the given dataset [19], [25]. The main characteristic of this method is the number of very strong (naive) assumptions regarding the independence of each event condition. Naïve Bayes Classifier assumes that all independent and non-attributes are interdependent based on the values given to the class variables. This method has a high level of accuracy with fast computing time [9], [26] so it can predict future opportunities based on previous experience. In addition, the model is simple and easy to create, can be implemented for large data sets, and has good accuracy, making this method sufficient for identifying the level of drug users [11]. Probability calculations in this method use equation 1 [9]:

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

Where X is data with an unknown class, H is the hypothesis that the data is a specific class, $P(H|X)$ is the probability H based on the condition in the hypothesis X (posteriori probability), $P(H)$ is the probability in the hypothesis H (prior probability), $P(X|H)$ is the probability of X based on the conditions in the hypothesis H and $P(X)$ is the probability of X.

2.4 Confusion Matrix

This research uses the Confusion matrix to measure the performance of the classification model. The confusion matrix works by comparing the classification results with the application of the algorithm carried out with the results. The classification should produce information on the accuracy level of the algorithm's application [26]. Not only the accuracy value, the Confusion matrix can also produce the percentage of data records with valid (correct) values in predictions by the algorithm and the percentage of predicted data records with incorrect values by the algorithm. The Confusion Matrix table can be seen in Table 5 [26].

Table 5. *Confussion Matrix*

Actual Classification	Classification	
	Correct Prediction	Incorrect Prediction
Positive Actual	True Positive	False Negative
Negative Actual	False Positive	True Negative

Based on Table 5, True Positive (TP) shows a positive value based on the number of positive data records classified. False Positives (FP) shows a positive value based on the number of negative data records classified. False Negative (FN) shows a negative value based on the number of positive data records classified. In contrast, True Negative (TN) shows a negative value based on the number of negative data records classified.

3 RESULTS AND ANALYSIS

3.1. Data Classification Using the Naïve Bayes Classifier Method

The Naïve Bayes Classifier method divides two stages in the classification process: learner and classifier. The steps in this research's learner and classifier process are described as follows:

1. Learner

   The Learner process trains the training data to produce a probabilistic model (classifier). In the learner process, there is a process to determine the category or class of drug dependency level. This research has three classes of drug dependence levels, namely tried using (light), regularly using (moderate), and dependence (severe). After getting the class labels, calculate the initial probability of each category or class (prior probability for each class). The conditional probability of variables or criteria in the class used in determining the level of drug dependence is calculated from these calculations.

   The prior probability value for each class is obtained by calculating the appropriate amount of data from the class label. The probability of class labels with training data for drug users is 120 data, consisting of 43 data for the tried-to-use class (light), 54 data for the regularly used class (moderate), and 23 data for the dependency class (severe). The prior probability value for each class is calculated by dividing the incident class data by the total case data in the training data using Equation (1). The overall calculation of the probability values for each class is shown in Table 6.

Table 6. Probability Value of Each Class

Class	Amount of data	Probability Value
Try Using (Lightly)	43	0.358333333
Regularly Use (Moderate)	54	0.45
Dependency (Severe)	23	0.191666667

The next stage of the learner process is to calculate the conditional probability of the criteria in the class used in determining the level of drug dependence. Table 7 displays the conditional probability values for the requirements for frequency of use, method of drug use, duration of drug use, psychological symptoms, amount of substance, alcohol, and narcotics screening scores.

Table 7. Criteria for Determining the Level of Drug Dependence

Criteria	Category	Number of events			Probability Value		
		Light	Moderate	Severe	Light	Moderate	Severe
Frequency of Drug Use	Small Dose	42	8	1	0.976744186	0.148148148	0.043478261
	Quite a High Dosage	1	45	8	0.023255814	0.833333333	0.347826087
	High dosage	0	1	14	0	0.018518519	0.608695652
How to Use Drugs	Swallowed	17	13	1	0.395348837	0.240740741	0.043478261
	Inhaled/smoked	5	9	1	0.11627907	0.166666667	0.043478261
	Smoke	21	31	12	0.488372093	0.574074074	0.52173913
	Inject	0	1	9	0	0.018518519	0.391304348
Length of Drug Use	New user	33	1	0	0.76744186	0.018518519	0
	Long enough	10	43	4	0.23255814	0.796296296	0.173913043
	Long	0	10	19	0	0.185185185	0.826086957
Psychological Symptoms	There isn't any	40	1	0	0.930232558	0.018518519	0
	Difficulty Remembering	2	25	1	0.046511628	0.462962963	0.043478261
	Anxiety	1	26	2	0.023255814	0.481481481	0.086956522
	Frequent hallucinations	0	1	9	0	0.018518519	0.391304348
	Depression	0	1	11	0	0.018518519	0.47826087
Amount of Substance used	One Kind	33	21	2	0.76744186	0.388888889	0.086956522
	Two Types	10	32	12	0.23255814	0.592592593	0.52173913
	Three Types	0	1	7	0	0.018518519	0.304347826
	Four Types	0	0	2	0	0	0.086956522

Alcohol Screening Score	No Intervention	40	29	6	0.930232558	0.537037037	0.260869565
	Brief Intervention	3	24	10	0.069767442	0.444444444	0.434782609
	Intensive Treatment	0	1	7	0	0.018518519	0.304347826
Narcotics Screening Score	No Intervention	20	1	0	0.465116279	0.018518519	0
	Brief Intervention	23	52	5	0.534883721	0.962962963	0.217391304
	Intensive Treatment	0	1	18	0	0.018518519	0.782608696

The results of the initial probability values for each class and the conditional probability values for each criterion calculated in the learner process are shown in Table 7. In the classifier process, they are used to classify user testing data.

2. Classifier
   The Classifier process determines the drug user testing data's category (classification). The calculation to determine the level of drug dependence uses the 1st user testing data sample in Table 4. The 1st user has the frequency of drug use in "small doses," the method of drug use is "swallowed," the duration of drug use is "quite a long time," the psychological symptoms are "not present," several substances "two types," alcohol screening score "no intervention" and narcotics screening score "no intervention." The calculation stage starts with calculating each conditional probability value for each attribute.

1. Conditional Probability Value of Frequency of Use of "Small Dose"
   The probability value for the frequency of small dose drug use criteria is calculated using equation 1. This calculation is also based on training data and the conditional probability value for the small dose criterion, which is in Table 3 with the following calculations.

$$P(Frequency\ of\ Use = Small\ Dose|Class = "light" = \frac{42}{43} = 0.976744186$$

$$P(Frequency\ of\ Use = Small\ Dose|Class = "moderate" = \frac{8}{54} = 0.148148148$$

$$P(Frequency\ of\ Use = Small\ Dose|Class = "severe" = \frac{1}{23} = 0.043478261$$

2. Conditional Probability Value of How to Use Drugs "Swallowed"
   After getting the probability value for the frequency of drug use criteria, the next calculation is to determine the conditional probability value of how to use drugs using Equation 1.

$$P(How\ to\ Use\ Drugs = Swallowed|Class = "Light" = \frac{17}{43} = 0.395348837$$

$$P(How\ to\ Use\ Drugs = Swallowed|Class = Moderate" = \frac{13}{54} = 0.240740741$$

$$P(How\ to\ Use\ Drugs = Swallowed|Class = "Severe" = \frac{1}{23} = 0.043478261$$

3. Conditional Probability Value of Length of Drug Use "Long Enough"
   The conditional probability criteria for quite a long time uses the same equation, namely equation 1, as follows:

$$P(Length\ of\ Drug\ Use = Long\ Enough|Class = "Light" = \frac{10}{43} = 0.23255814$$

$$P(Length\ of\ Drug\ Use = Long\ Enough|Class = "Moderate" = \frac{43}{54} = 0.796296296$$

$$P(Length\ of\ Drug\ Use = Long\ Enough|Class = "Severe" = \frac{4}{23} = 0.173913043$$

4. Probability Value of Psychological Symptoms "None"
   The calculation to determine the probability value for the psychological symptom criterion "None" is carried out in the same way as the calculation for the previous criterion with Equation 1:

$$P(Psychological\ Symptoms = None|Class = "Light" = \frac{40}{43} = 0.930232558$$

$$P(Psychological\ Symptoms = None|Class = "Moderate" = \frac{1}{54} = 0.018518519$$
$$P(Psychological\ Symptoms = None|Class = "Severe" = \frac{0}{23} = 0$$

5. Conditional Probability Value of the Number of Types of Substance "Two Types"
The criteria for the number of types of substances in the two types category use the following probabilities:
$$P(Amount\ of\ Substance\ used = Two\ types|Class = "Light" = \frac{10}{43} = 0.23255814$$
$$P(mount\ of\ Substance\ used = Two\ types|Class = "Moderate" = \frac{32}{54} = 0.592592593$$
$$P(Amount\ of\ Substance\ used = Two\ types|Class = "Severe" = \frac{12}{23} = 0.52173913$$

6. Conditional Probability Value of "No Intervention" Alcohol Screening Score
Search for probability values for alcohol screening scores, namely:
$$P(Alcohol\ Screening\ Score = No\ Intervention|Class = "Light" = \frac{40}{43} = 0.930232558$$
$$P(Alcohol\ Screening\ Score = No\ Intervention|Class = "Moderate" = \frac{29}{54} = 0.537037037$$
$$P(Alcohol\ Screening\ Score = No\ Intervention|Class = "Severe" = \frac{6}{23} = 0.260869565$$

7. Conditional Probability Value of Narcotics Screening Score "No Intervention"
The next calculation looks for the probability of the narcotics screening score criteria using Equation 1 as the following calculation:
$$P(Narcotics\ Screening\ Score = No\ Intervention|Class = "Light" = \frac{20}{43} = 0.465116279$$
$$P(Narcotics\ Screening\ Score = No\ Intervention|Class = "Moderate" = \frac{1}{54} = 0.018518519$$
$$P(Narcotics\ Screening\ Score = No\ Intervention|Class = "Severe" = \frac{0}{23} = 0$$

After calculating the conditional probability value for each criterion, calculate the final probability for each class by multiplying all the conditional probability values for each attribute obtained according to the event class, calculating the final probability for each class as follows:
$$P(X|Light) = 0.976744186 * 0.395348837 * 0.23255814 * 0.930232558 * 0.23255814 * 0.930232558 * 0.465116279$$
$$P(X|Light) = 0.008405605$$
$$P(X|Moderate) = 0.148148148 * 0.240740741 * 0.796296296 * 0.018518519 * 0.592592593 * 0.537037037 * 0.018518519$$
$$P(X|Moderate) = 3.09951\text{E} - 06$$
$$P(X|Severe) = 0.043478261 * 0.043478261 * 0.173913043 * 0 * 0.52173913 * 0.260869565 * 0$$
$$P(X|Severe) = 0$$

The next process is to calculate the final probability value by multiplying the class variable obtained in calculating the final probability for each criterion with the prior probability value for each class. The prior probability value for each class can be seen in Table 3.
$$P(X|Class = Light) * P(Class = Light)$$
$$= 0.008405605 * 0.358333333$$
$$= \mathbf{0.003012009}$$
$$P(X|Class = Moderate) * P(Class = Moderate)$$
$$= 3.09951\text{E} - 06 * 0.45$$
$$= 1.39478\text{E} - 06 = 1.3 * 10^{-5} = 0.000013$$
$$P(X|Class = Moderate) * P(Class = Moderate)$$
$$= 0 * 0.191666667$$
$$= 0$$

Based on the results of the final probability calculation, it can be seen that the highest probability value is in the try-to-use (light) class with a value of 0.003012009 so that the 1st user's data falls into the try-to-use (light) level classification. The calculation process for all other testing data is the same as the calculation process that has been carried out. The overall classification results of testing data can be seen in Table 8.

Table 8. Drug User Classification Results

Drug Users	Final Probability Value			Classification
	Light	Moderate	Severe	
User 1	0.003012009	0.000013	0	Try Using (lightly)
User 2	0	0.000014	0.000629166	Addict (heavy)
User 3	0	0.001707796	0.000061	Regularly Used (medium)
User 4	0.040518603	0.0000507	0	Try Using (lightly)
User 5	0.000146171	0.000015	0	Try Using (Lightly)
⋮	⋮	⋮	⋮	⋮
User 96	0.000068	0.006961046	0.000021	Regularly Used (medium)
User 97	0.0000241	0.000074	0	Try Using (lightly)
User 98	0	0.0000107	0.002241405	Addict (heavy)
User 99	0.000041	0.010199335	0.000064	Regularly Used (medium)
User 100	0	0.000189919	0.000012	Regularly Used (medium)

Based on Table 8, from one hundred testing data classified using the Naïve Bayes Classifier method, 29 data were classified as drug dependence levels (light), 42 data were classified as regular use (moderate) levels of dependence, and 29 data were classified as regular use (moderate) levels of dependence. Level of dependence (severe). The results of the calculations that have been carried out are then applied to the system being built, as shown in Figures 1.

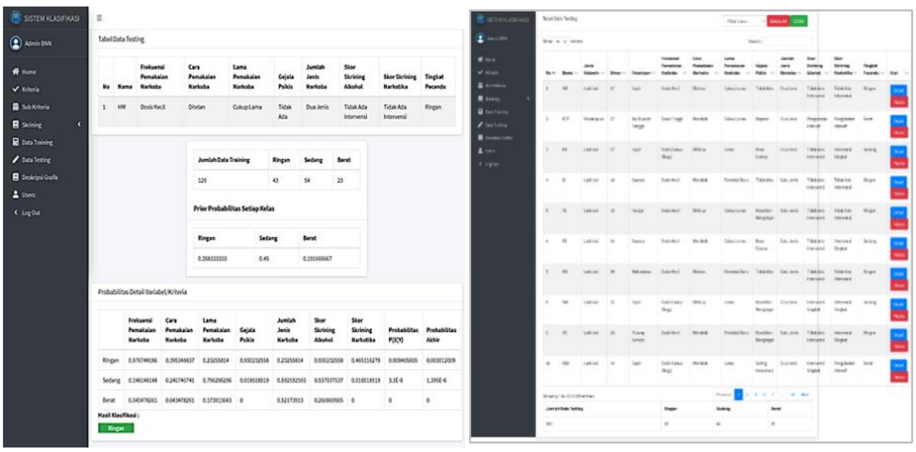

Figure 1. Classification Calculation Results System Page

3.2. Classification Performance Testing

Classification performance testing using a confusion matrix was conducted to assess the classification results by the Naïve Bayes Classifier method. Testing is based on accuracy, precision, and recall values obtained from differences in the amount of actual data and classification results in drug user testing data. Of the 100 testing data used, 94 were successfully classified correctly, and six were not classified correctly. Table 9 displays drug user testing data's actual data and classification results.

Table 9. Total actual data and classification results data

Actual	Classification Results			
	Light	Moderate	Severe	Total

Light	26	0	0	26
Moderate	3	42	3	48
Severe	0	0	26	26
Total	29	42	29	100

Based on the data in Table 9, the accuracy, precision, and recall values can be found in the following way.

$$Accuracy = \frac{TP}{Total\ Testing\ Data} * 100 = \frac{(26+42+26)}{100} * 100 = 94\ \%$$

$$Precision = \frac{Precision\ light + moderate + severe}{Total\ Class} * 100 = \frac{(1+0{,}87+1)}{3} * 100 = 95\ \%$$

$$Recall = \frac{Recall\ light + moderate + severe}{Total\ Class} * 100 = \frac{(0{,}89+1+0{,}89)}{3} * 100 = 92\ \%$$

The calculation results show that the resulting accuracy value is 94%, precision is 95%, and recall is 92%. Based on this, the Naïve Bayes Classifier method has succeeded in classifying the level of drug dependence with very high accuracy.

4. CONCLUSION

The Naïve Bayes Classifier method has been applied to the system to determine the level of dependence of drug users. From the classification results, information was obtained that the largest number of drug users were drug users with a moderate level of dependence or regularly used illegal drugs, and the rest were heavy and light drug users. The classification results that have been carried out are based on the probability values for each criterion in each different class. The probability values of these criteria are optimized to determine the level of drug dependence based on the classification process carried out by the Naive Bayes Classifier method. Furthermore, the classification performance carried out by the Naïve Bayes Classifier was tested with a confusion matrix and obtained very good accuracy results of 94%, precision with a percentage of 95% and 92% for recall. It indicates that, overall, this method produces accurate classification, and the system that has been built can run very well.

From the conclusions obtained, this study can help the national narcotics agency in the process of determining the level of drug dependence using the system that has been built so that the diagnostic results obtained can provide objective information related to appropriate treatment and rehabilitation for drug users based on symptoms and categories of drug addiction of users. Further research is expected to use other classification methods such as K-Nearest Neighbor or add criteria and levels of dependence to contribute to science in the field of diagnosing the level of drug addicts.

REFERENCES

[1] R. Ghanbari and S. Sumner, “Using Metabolomics to Investigate Biomarkers of Drug Addiction,” *Trends Mol. Med.*, vol. 24, no. 2, pp. 197–205, 2018, doi: 10.1016/j.molmed.2017.12.005.

[2] A. Sonjaya, “Construction of the Rehabilitation Model for Drug Abuse in Non-Penal Criminal Policy Perspective,” *Open J. Leg. Stud.*, vol. 3, no. 2, pp. 111–124, 2020, doi: 10.32591/coas.ojls.0302.03111s.

[3] M. A. I. Arif, S. I. Sany, F. Sharmin, M. S. Rahman, and M. T. Habib, “Prediction of Addiction to Drugs and Alcohol using Machine Learning: A Case Study on Bangladeshi population,” *Int. J. Electr. Comput. Eng.*, vol. 11, no. 5, pp. 4471–4480, 2021, doi: 10.11591/ijece.v11i5.pp4471-4480.

[4] R. Jabar and S. Nurhayati, “The Effect of Drug Hazard Counselling in Improving Public Knowledge Level of Hazardous Drugs,” *J. SPEKTRUM*, vol. 9, no. 4, pp. 455–461, 2021, doi: 10.24036/spektrumpls.v9i4.114106.

[5] R. H. Kusuma and S. A. Rayhaniah, “The Role of Counseling and Islamic Communication in Client Recovery Effort at the National Narcotics Agency Rehabilitation Center (BNN) Tanah Merah-Samarinda,” *J. Al-Tazkiah*, vol. 11, no. 2, pp. 123–140, 2022, doi: https://doi.org/10.20414/altazkiah.v11i2.6002 THE.

[6] E. Sukamto, R. Rasmun, P. Andi, and S. Sutrisno, “The Effect of Family Support Toward Motivation in Following the Drugs Rehabilitation Program,” *J. Glob. Res. Public Heal.*, vol. 4, no. 1, pp. 7–14, 2019.

[7] T. Alamsyah, A. Candra, and D. Marianthi, “Pageu Gampong Model in Aceh Culture on Drug Handling,” *Int. J. Health Sci. (Qassim).*, vol. 4, no. 3, pp. 49–59, 2020, doi: 10.29332/ijhs.v4n3.458.

[8] A. Hadiyanto, “Effectiveness of Narcotic Addict Rehabilitation Share to Suppress Crime narcotics (Study in Loka Rehabilitation of the National Narcotics Agency Riau Islands Province),” *4th Int. Call Pap.*, vol. 1, no. 1, pp. 64–80, 2018.

[9] S. Prabha P.M and S. B, “Sentimental Analysis using Naive Bayes Classifier,” *Int. Conf. Vis. Towar. Emerg. Trends Commun. Netw.*, pp. 1–5, 2019, doi: 10.1109/ViTECoN.2019.8899618.

[10] T. B. Shahi and A. K. Pant, “Nepali News Classification using Naïve Bayes, Support Vector Machines and Neural Networks,” *Int. Conf. Commun. Inf. Comput. Technol.*, pp. 1–5, 2018, doi: 10.1109/ICCICT.2018.8325883.

[11] N. Salmi and Z. Rustam, “Naïve Bayes Classifier Models for Predicting the Colon Cancer,” *J. Annu. Basic Sci. Int. Conf.*, pp. 1–9, 2019, doi: 10.1088/1757-899X/546/5/052068.

[12] V. Rawat and Suryakant, “A Classification System for Diabetic Patients with Machine Learning Techniques,” *Int. J. Math. Eng. Manag. Sci.*, vol. 4, no. 3, pp. 729–744, 2019, doi: 10.33889/IJMEMS.2019.4.3-057.

[13] F. Z. Parinduri, R. Dewi, and S. Susiani, “Classification of Internet Addiction Levels in Students using the Naïve Bayes Algorithm,” *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 3, pp. 257–264, 2022, doi: 10.55123/jomlai.v1i3.965.

[14] D. Fahrudy *et al.*, “Intelligent System for Classification of Student Personality With Naive Bayes Algorithm,” *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 1, pp. 1–9, 2022, doi: 10.31598/sintechjournal.v5i1.969.

[15] G. Liang, Y. G. Yan, M. Wang, X. L. Lian, M. S. Li, and W. H. Tang, “Classification for Text Data from the Power System Based on Improving Naive Bayes,” *Asia-Pacific Power Energy Eng. Conf. APPEEC*, pp. 1–6, 2020, doi: 10.1109/APPEEC48164.2020.9220634.

[16] F. Handayani and R. W. Sembiring, “Application of Nave Bayes Algorithm for Security Performance Evaluation at PT . Sei Mangke Nusantara 3,” *J. Artif. Intell. Eng. Appl.*, vol. 1, no. 1, pp. 3–8, 2021.

[17] E. Firasari, N. Khasanah, U. Khultsum, D. N. Kholifah, R. Komarudin, and W. Widyastuty, “Comparation of K-Nearest Neighboor (K-NN) and Naive Bayes Algorithm for the Classification of the Poor in Recipients of Social Assistance,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 1–6, 2020, doi: 10.1088/1742-6596/1641/1/012077.

[18] A. Hermawan, “Implementation of Naïve Bayes Algorithm for Classification of Mental Health of Social Media Users,” *bit-Tech*, vol. 4, no. 2, pp. 61–70, 2021, doi: 10.32877/bt.v4i2.282.

[19] M. O. Mughal and S. Kim, “Signal Classification and Jamming Detection in Wide-band Radios using Naïve Bayes Classifier,” *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1398–1401, 2018, doi: 10.1109/LCOMM.2018.2830769.

[20] A. W. Syaputri, E. Irwandi, and M. Mustakim, “Naïve Bayes Algorithm for Classification of Student Major’s Specialization,” *J. Intell. Comput. Heal. Informatics*, vol. 1, no. 1, p. 17, 2020, doi: 10.26714/jichi.v1i1.5570.

[21] T. Sajana and M. R. Narasingarao, “Classification of Imbalanced Malaria Disease using Naïve Bayesian Algorithm,” *Int. J. Eng. Technol.*, vol. 7, no. 2.7, pp. 786–790, 2018, doi: 10.14419/ijet.v7i2.7.10978.

[22] V. R. Balaji, S. T. Suganthi, R. Rajadevi, V. Krishna Kumar, B. Saravana Balaji, and S. Pandiyan, “Skin Disease Detection and Segmentation using Dynamic Graph Cut Algorithm and Classification Through Naive Bayes Classifier,” *Meas. J. Int. Meas. Confed.*, vol. 163, p. 107922, 2020, doi: 10.1016/j.measurement.2020.107922.

[23] M. Andrejiova and A. Grincova, “Classification of Impact Damage on a Rubber-Textile Conveyor Belt using Naïve-Bayes Methodology,” *Wear*, vol. 414–415, pp. 59–67, 2018, doi: 10.1016/j.wear.2018.08.001.

[24] O. Moscoso-zea, P. Saa, and S. Luján-mora, “Evaluation of Algorithms to Predict Graduation Rate in Higher Education Institutions by Applying Educational Data Mining,” *Australas. J. Eng. Educ.*, pp. 1–10, 2019, doi: 10.1080/22054952.2019.1601063.

[25] N. L. W. S. R. Ginantra and N. W. Wardani, “Measurement of the Similarity of Indonesian Papers on One Journal Topic with the Naive Bayes Algorithm and Vector Space Model,” *Ijconsist Journals*, vol. 1, no. 1, pp. 20–26, 2019.

[26] N. Puspitasari, R. Rosmasari, F. W. Pratama, and H. Sulastri, “Quality Classification of Palm Oil Varieties Using Naive Bayes Classifier,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 11–23, 2022.