Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

Performance Comparison of CART And KNN Algorithms for Analyzing Early Predictions of Mental Health

Eling Sekar Anggraeni[1)*], Lulu Amnah Fitriya Maharani[2)], Desi Riyanti[3)], Ranggi Praharaningtyas Aji[4)], Pungkas Subarkah[5)]
[1)*2)3)4)5)] Universitas Amikom Purwokerto, Jawa Tengah, Indonesia
[1)*]elingsekara@email.com, [2)]luluamnah241@email.com, [3)]desiriyanti097@email.com,
[4)]ranggi.p.aji@amikompurwokerto.ac.id, [5)]subarkah@amikompurwokerto.ac.id

ABSTRACT

Currently, mental health is an unresolved mental health problem both at the national and international levels. Mental health disorders are conditions where a person has difficulty in adjusting to the conditions around them. Mental Health is an important aspect of overall health. Efforts to maintain and improve it can help a person achieve better well-being in everyday life. This research aims to conduct Early Prediction Analysis related to mental health problems experienced by students by measuring the accuracy level of the analysis. This research was conducted using the CART (Classification and Regression Trees) and KNN (K-Nearest Neighbor) algorithms with a set of Mental Health Datasets consisting of 11 attributes and 101 data. The data is processed using the Weka Application and the accuracy results of each algorithm are obtained, amounting to 94.0594% for the CART Algorithm and 91.0891% for the KNN Algorithm. From this achievement, it can be concluded that the performance of the CART and KNN algorithms falls into the Excellent Classification category. Judging from the accuracy obtained, the CART algorithm has a higher accuracy value than the KNN algorithm, so the CART algorithm has a high performance for analyzing early prediction of mental health of students who do not take steps in seeking mental health support.

Keywords: Early Prediction Analysis; Mental Health, CART Algorithm; KNN Algorithm

1. INTRODUCTION

Mental health plays an important role in a person's life, and by having good mental health, a person can carry out activities as a living being. A healthy mental state will support one's progress towards a better future. (Masyarakat et al., 2023). Mental health is a behaviour of a person that can trigger disturbances in the neural network of the human brain, which will further affect the mental state. The burden of mental health can also affect a person's ability to face the challenges of life, and therefore, a person's psychological level is influenced by emotional experiences, behaviour, environment, and family educational background (Kurnia, 2021). Mental health is closely linked to aspects of physical health. Research focuses on the impact of an individual's mental health on their health. Some of the complaints expressed by medical personnel include mental disorders and psychological illnesses, which can have a significant impact on a person's physical condition. The state of health or illness is a complexity that involves biological, psychological and social aspects in the same condition (Azania & Naan, 2021). A person who experiences illness is not only caused by internal factors, but can also be affected by a lack of ability to manage time, so that the energy in the body is drained.

One of the problems in mental health is Depression. Although Depression is common amongst adolescents, it is often overlooked, unlike the adult phase which receives special attention regarding its symptoms. Symptoms of depression in adolescents are often perceived as manifestations of: excessive emotional sensitivity, mood swings, and fluctuating depression symptoms (Praptikaningtyas et al., 2019). Depression can be caused by a variety of factors, including biological, psychological and socioeconomic. Biological factors involve neurobiological abnormalities, hormonal imbalances, and genetic factors that affect neurotransmitters such as serotonin and other neurons in the brain (Al Aziz, 2020). Depression is a serious problem that is ranked fourth globally. The impact of depression on a personal level can affect daily activities, characterised by excessive levels of sadness, lack of attention to the surrounding environment, stress, and has the potential to become a social problem with high suicide rates and tends to increase

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

every year (Sudarsono & Lestari, 2020). Caniaghi argues that depression is a mood disorder characterised by difficulty controlling feelings of severe suffering, while mood itself refers to a person's emotional state (Wisnugraha et al., 2023).

Therefore, making observations related to the increase in disease rates each year shows that early prevention and prediction of health has a positive impact on a person, such as the ability to anticipate disease and avoid other negative impacts. In the context of treatment, efforts to minimise dependence on psychologist treatment can be made, and the support of the family can provide a good enlightenment for a person and can help ease the burden of thoughts or overcome the problems being faced (Rahma et al., 2022). In addition, the importance of discussing mental health is a way to empower individuals, families and communities to find, maintain and optimise their mental health to face the challenges of everyday life (Radiani et al., 2019).

This research aims to analyse the early prediction of mental health problems experienced by adolescents using the KNN (K-Nearest Neighbor) and CART (Clasification and Regression Trees) algorithms. Comparison of these two methods is important because both have different advantages and disadvantages. KNN is a simple and effective algorithm for classification problems, but it can be sensitive to noisy data and outliers. On the other hand, CART is a decision tree-based algorithm that can handle complex data and provide interpretable results, but can be prone to overfitting. By comparing the performance of these two algorithms, this study aims to find out which algorithm is more accurate in predicting mental health problems in adolescents.

2. LITERATURE REVIEW

In order for this research to focus more on a research problem and produce research novelty, as well as map the position of the research to be carried out by the researcher, the researcher needs to conduct a stage study of previous studies that are similar to the theme of the research to be carried out. Based on this, the researcher conducted a literature study of the results of previous studies and the results were described as follows:

1. This research with the classification of mental health in a person was conducted by Daniel Dwi Kurnia, Septi Andryana, and Aris Gunaryati in 2021 (Kurnia, 2021). Quantitative approach and methods used Genetic Algorithm. This research can be concluded that many people suffer from mental disorders in a person, health consultations to specialists require quite expensive costs. The solution to deal with this is to diagnose the early symptoms of mental health.
2. Research on the classification of mental health in college students was conducted by Desti and Naan 2021 with research locations at Sunan Gunung Djati University students (Azania & Naan, 2021). This dissertation was prepared with a quantitative approach. This study concludes that mental health is students doing an activity by producing spiritual values in supporting mental health amid the covid-19 pandemic. Students have positive thoughts that this pandemic will end soon.
3. Research conducted by Gunawan Sudarsono and Sri Poedji Lestari in 2020 (Sudarsono & Lestari, 2020). Through a quantitative approach and the K-Nearest Neightbor method, this study concludes that some of them diagnose and make decisions based on the level of depression of final year students that affect mental health.
4. This research was conducted by Fuspita Dewi Savitri in 2022 (Savitri, 2022). Using a quantitative approach and the CART method, also known as a nonparametric method, concluded that cart is to obtain an accurate group of data as a characterisation of a classification.
5. This research was conducted by Ade Septi Binjori in 2020 (Binjori, 2020). Using a quantitative approach and the CART method concluded that CART is a method or algorithm from one of the data exploration techniques, namely the decision tree technique.

Research with the classification of mental health levels by Widya A Radiani in 2022 entitled "Present Mental Health and Islamic Handling of Disorders". Using a Qualitative approach by taking from several conclusions that mental health is a harmony in a psyche, with the benefit of a person's self in order to avoid anxiety and inner pressure. Islamic teachings by using therapy to people who have Islamic knowledge with the aim of curing someone or patient.

Based on the literature review conducted, the advantage of using the CART and KNN methods in this study is that the CART method can be used to produce accurate data groups to characterize a classification, and can be used as one of the data exploration techniques, namely the decision tree technique. Meanwhile, the KNN method can be used to diagnose and make decisions based on the level of depression of final year students that affect mental health.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

The difference between previous research and current research is that previous research focused more on the classification of mental health in individuals or students, and used different methods such as Genetic Algorithms and qualitative approaches. Meanwhile, the current research uses CART and KNN methods to classify mental health in college students, and uses a quantitative approach. Therefore, this research can provide more accurate and effective results in classifying mental health in college students.

3. METHOD

A. Research Procedure

This research was conducted by applying the CART and K-Nearest Neighbor Algorithms in Early Analysis and Prediction of Mental Health. The implementation of the CART algorithm itself is done through compiling candidate splits carried out on all predictor variables and KNN itself is carried out through a series of trials with variations in the value of k and the size of the Euclidean distance. At the Evaluation stage, CART performance testing is calculated to determine branches that have suitability and KNN itself is carried out using Accuracy, Precision and Recall values. The stages of the research to be carried out are shown in Figure 1 below:

:

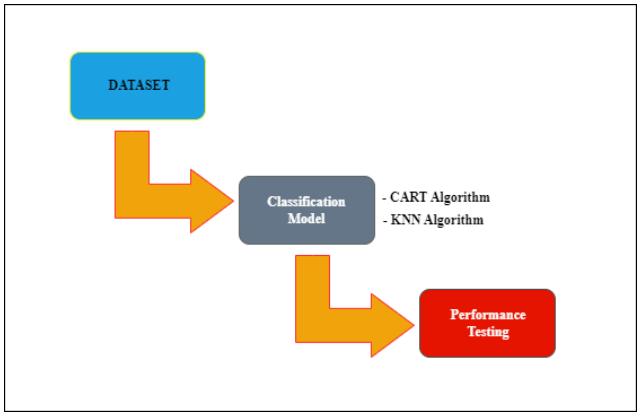

Fig. 1 Research Stages of Mental Health Prediction Analysis Using CART and K-Nearest Neighbor Algorithms

B. Mental Health Dataset

This stage uses a mental health dataset obtained from the kaggle.com website with the command https://www.kaggle.com/datasets/shariful07/student-mental health/versions/1/data?select=Student+Mental+health.csv . This dataset consists of 11 attributes and includes 101 records. The characteristics of the data obtained from the dataset are included in Quantitative data which is presented in the form of numbers or numerical and the data processing is carried out using statistical and mathematical methods with the help of supporting software.

Table 1.
Attribute Description

Atribut	Deskripsi
Timestamp	Date and time information at the time of data collection
Choose your gender	Gender consists of 2 categories Female and Male
Age	Student age at the time of data collection
What is your course	Student's study program at the time of data collection

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

Your Current Year of Student	Current year of study
What is your CGPA	Student Grade Point Average (GPA)
Marital Status	Marital Status consists of Yes : Married, No: Not Married
Do you have depresssion	Depression status experienced by students (Yes: Yes Depression, No: No Depression)
Do you have anxiety	Anxiety status experienced by students (Yes: Yes Anxious, No: Not Anxious)
Do you have panic attack	Panic attack status experienced by students (Yes: Yes experiencing panic attacks, No: Not experiencing panic attacks)
Did you seek any specialist a treatment	Status of students who seek specialists for treatment (Yes: Yes, looking for a treatment specialist, No: Not looking for treatment specialists)

C. *CART Algorithm (Classification and Regression Trees)*

CART (Classification and Regression Trees) is one of the algorithms of decision tree data processing techniques. The CART algorithm builds a decision tree by selecting the most optimal branching of each node. The CART algorithm is a decision tree algorithm that is certain in its branching always amounts to two or binary branching. Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone are experts who first discovered the concept of calculating the CART algorithm. Larose in Trees formed using training data (training set) may be able to place class values based on the values of other variables or independently on target variables from a new data set. Building a binary tree based on a function of input variables by separating records at each node is known as the CART algorithm(Melani & Sulastri, 2020).

The CART (Classification And Regression Tree) algorithm classifies a record into one of the available classifications on the destination variable based on the values of its predictor variables. The steps in the CART algorithm are as follows (Savitri, 2022) :

a. The first step is to construct the candidate split. This is done for all the predictor variables as a whole. The list containing the candidate splits is called the current candidate split list.
b. The second step (Iteration-1) is to evaluate the performance of all candidate branches in the current candidate branch list by calculating the suitability magnitude value of candidate branch s at decision point t, denoted by Φ(s|t) and defined as follows. can be identified by the following equation (Santiko & Subarkah, 2019):

$$\emptyset(s|t) = 2.PL.PR.Q(s|t) \quad (1)$$

$$Q(s|t) = \sum_{j=1}^{jumlah\ kategori} |P(j|tL) - P(j|tR)| \quad (2)$$

t_L = the left branch of the decision point t

t_R = the right branch of the decision point t

$$P_L = \frac{left\ branch\ candidate\ (t_L)}{training\ data} \quad (3)$$

$$P_R = \frac{right\ branch\ candidate\ (t_R)}{training\ data} \quad (4)$$

$$P(j|t_L) = \frac{j\ left\ branch\ candidate\ (t_L)}{decision\ point\ t} \quad (5)$$

$$P(j|t_R) = \frac{j\ right\ branch\ candidate\ (t_R)}{decision\ point\ t} \quad (6)$$

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

D. *K-Nearst Neighbor Algorithm* (KNN)

K-Nearest Neighbor (K-NN) algorithm is an algorithm method used to classify objects based on learning data that has the closest distance to the object. Or in other words, K-Nearest Neighbor is one of the most frequently used simple algorithms, where data points are classified based on similarity with other adjacent groups of data points. As a result, this KNN algorithm is able to provide competitive results in the classification process (Mardiani et al., 2023).

$$d(x,y) = \sqrt{\sum_i^n (xi - yi)^2} \quad (7)$$

Explanation of the formula above is as follows:
d(x,y) : x,y data distance
xi : sample data
yi : testing data or test data
i : variable data
n : data dimension (Pratama et al., 2022).

Stages of the K-Nearest Neigbor (KNN) algorithm are as follows:

1. Determine the parameter K = i.e. the number of nearest neighbours.
2. Calculate the distance between the new data and all training data.
3. Sort the distances and assign nearest neighbours according to the to-K minimum distance.
4. Checking the class of the nearest neighbour
5. Using the simple majority of the nearest neighbour class as the new data prediction value (Prakoso et al., 2020).

4. RESULT

This research was conducted using the Mental Health dataset obtained from Kaggle.com. This dataset consists of 11 Attributes and has 101 records with 1 target attribute. The target attribute has 2 categories, namely Yes (The student is looking for a treatment specialist) and No (The student is not looking for a treatment specialist). Before the data is processed, the data is entered into the Weka Application to carry out the pre-processing stage which aims to adjust the attributes to be processed and to identify empty values (Missing Values) in each attribute. The following is a table explaining the attributes in the dataset :

Table 2
Description Of Attributes In The Dataset

Atribut	Description
Timestamp	Date and time information at the time of data collection
Choose your gender	Gender consists of 2 categories Female and Male
Age	Student age at the time of data collection
What is your course	Student's study program at the time of data collection
Your Current Year of Student	Current year of study
What is your CGPA	Student Grade Point Average (GPA)
Marital Status	Marital Status consists of Yes : Married, No: Not Married
Do you have depresssion	Depression status experienced by students (Yes: Yes Depression, No: No Depression)
Do you have anxiety	Anxiety status experienced by students (Yes: Yes Anxious, No: Not Anxious)
Do you have panic attack	Panic attack status experienced by students (Yes: Yes experiencing panic attacks, No: Not experiencing panic attacks)

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

Did you seek any specialist a treatment	Status of students who seek specialists for treatment (Yes: Yes, looking for a treatment specialist, No: Not looking for treatment specialists)

Next is to do the Pre-Processing stage, which is the stage for selecting data in order to obtain data that is ready for use. Then after the dataset has been adjusted using the Weka Application, the results of the adjustment can be seen as follows :
So at the age attribute there is a missing value of 1% so to solve this problem, replace the missing value for the age attribute. Because this attribute is numeric data, replace missing value using mean or average.

The next step involves processing the dataset using the Weka Application. The purpose of this stage is to produce a Confusion Matrix value. In this process, users involve the CART and KNN algorithms in Weka with the SimpleCART and IBk methods to process the dataset.

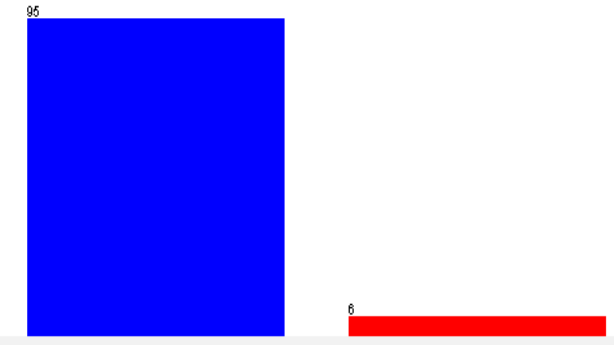

Fig.2 Dataset Visualization

Figure 2, is an image of the Dataset Visualization. In the Dataset Visualization, it displays two groups of data with different colors. The blue graph shows 95 data, representing the NO category, indicating that the students did not seek specialists for treatment. On the other hand, the red graph with 6 data represents the YES category, indicating that the students sought specialists for treatment.

The next step in this research stage is the Dataset Performance Testing stage. At this stage, the dataset is tested using the CART and KNN algorithms with the Simple CART and lBk methods through the Weka Application. The test results using Weka, resulted in an Accuracy value of 91.0891% for the KNN Algorithm and 94.0594% for the CART Algorithm. This Accuracy value is calculated from the Accuracy, Precision, Recall, and F-Measure parameters. Details of the calculations for the KNN and CART algorithms will be explained as follows :

A. Result Algorithem CART

The candidate branches for the Age predictor variable are as follows :

1. Age = 18, Age = 19,20,21,22,23,24
2. Age = 18, Age = 19,20,21,22,23,24
3. Age = 19, Age = 18,20,21,22,23,24
4. Age = 20, Age = 18,19,21,22,23,24
5. Age = 21, Age = 18,19,20,22,23,24
6. Age = 22, Age = 18,19,20,21,23,24
7. Age = 23, Age = 18,19,20,21,22,24
8. Age = 24, Age = 18,19,20,21,22,23

The candidate branches for the predictor variable What is your CGPA? are as follows:

1. What is your CGPA? = 0 - 1.99, What is your CGPA? = 2.00 - 2.49,2.50 - 2.99,3.00 - 3.49,3.50 - 4.00
2. What is your CGPA? = 2.00 - 2.49, What is your CGPA? = 0 - 1.99,2.50 - 2.99,3.00 - 3.49,3.50 - 4.00

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

3. What is your CGPA? = 2.50 - 2.99, What is your CGPA? = 0 - 1.99,2.00 - 2.49,3.00 - 3.49,3.50 - 4.00
4. What is your CGPA? = 3.00 - 3.49, What is your CGPA? = 0 - 1.99,2.00 - 2.49,2.50 - 2.99,3.50 - 4.00
5. What is your CGPA? = 3.50 - 4.00, What is your CGPA? = 0 - 1.99,2.00 - 2.49,2.50 - 2.99,3.00 - 3.49

The candidate branches for the Do you have Depression? predictor variable are as follows:
Do you have Depression? = Yes, Do you have Depression? = No

The candidate branches for the Do you have Anxiety? predictor variable are as follows:
Do you have Anxiety? = Yes, Do you have Anxiety? = No

The candidate branches for the predictor variable Do you have Panic attack? are as follows:
Do you have Panic attack? = Yes, Do you have Panic attack? = No

Table 3
Data Description Do you look for a specialist doctor for treatment?

Nama Kategori	Nilai
Yes	6
No	95

Given the characteristic of the CART algorithm that each decision point is binary, the candidate branches will be named candidate left branch and candidate right branch.

Table 4
List of current branch candidates (iteration-1)

Branch Candidate	Left Branch Candidate (tL)	Right Branch Candidate (tL)
1	Age = 18	Age = 19,20,21,22,23,24
2	Age = 19	Age = 18,20,21,22,23,24
3	Age = 20	Age = 18,19,21,22,23,24
4	Age = 21	Age = 18,19,20,22,23,24
5	Age = 22	Age = 18,19,20,21,23,24
6	Age = 23	Age = 18,19,20,21,22,24
7	Age = 24	Age = 18,19,20,21,22,23
8	What is your CGPA? = 0 - 1.99	What is your CGPA? = 2.00 - 2.49,2.50 - 2.99,3.00 - 3.49,3.50 - 4.00
9	What is your CGPA? = 2.00 - 2.49	What is your CGPA? = 0 - 1.99,2.50 - 2.99,3.00 - 3.49,3.50 - 4.00
10	What is your CGPA? = 2.50 - 2.99	What is your CGPA? = 0 - 1.99,2.00 - 2.49,3.00 - 3.49,3.50 - 4.00
11	What is your CGPA? = 3.00 - 3.49	What is your CGPA? = 0 - 1.99,2.00 - 2.49,2.50 - 2.99,3.50 - 4.00

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

12	What is your CGPA? = 3.50 - 4.00	What is your CGPA? = 0 - 1.99,2.00 - 2.49,2.50 - 2.99,3.00 - 3.49
13	Do you have Depression? = Yes	Do you have Depression? = No
14	Do you have Anxiety? = Yes	Do you have Anxiety? = No
15	Do you have Panic attack? = Yes	Do you have Panic attack? = No

The second step of the algorithm will assess the overall performance of the branch candidates in the current branch candidate list. For now, the list of prospective branches is up-to-date. Then, how to measure the performance of each candidate branch listed on the list? The result of calculating the suitability score for each candidate branch that is still on the current branch candidate list is as follows:

Table 5
Calculation of suitability scores for candidate branches (iteration-1)

Branc Candidate	PL	PR	Did You Seek Any Specialist For A Treatment?	P(j|tL)	P(j|tR)	2.PL.PR	Q(s|t)	ϕ(s|t)
1	32/101 = 0,317		Yes	2/32 = 0,063	4/69 = 0,058	0,433	0,009	0,004
		69/101 = 0,683	No	30/32 = 0,938	65/69 = 0,942			
2	21/101 = 0,208		Yes	1/21 = 0,048	5/80 = 0,063	0,329	0,030	0,010
		80/101 = 0,792	No	20/21 = 0,952	75/80 = 0,938			
3	7/101 = 0,069		Yes	0/7 = 0	6/94 = 0,064	0,129	0,128	0,016
		94/101 = 0,931	No	7/7 = 1	88/94 = 0,936			
4	3/101 = 0,030		Yes	0/3 = 0	6/98 = 0,061	0,058	0,122	0,007
		98/101 = 0,970	No	3/3 = 1	92/98 = 0,939			
5	2/101 = 0,020		Yes	0/2 = 0	6/99 = 0,060	0,039	0,121	0,005
		99/101 = 0,980	No	2/2 = 1	93/99 = 0,939			
6	13/101 = 0,129		Yes	1/13 = 0,077	5/88 = 0,057	0,224	0,040	0,009
		88/101 = 0,871	No	12/13 = 0,923	83/88 = 0 943			
7	23/101 = 0,228		Yes	2/23 = 0,087	4/78 = 0,051	0,352	0,071	0,025
		78/101 = 0,772	No	21/23 = 0,913	74/78 = 0,949			
8	1/101 = 0,010		Yes	0/1 = 0	6/100 = 0,06	0,020	0,12	0,002
		100/101 = 0,990	No	1/1 = 1	94/100 = 0,94			
9	2/101 = 0,020		Yes	0/2 = 0	6/99 = 0,060	0,039	0,121	0,005
		99/101 = 0,980	No	2/2 = 1	93/99 = 0,939			
10	5/101 = 0,050		Yes	2/5 = 0,4	4/96 = 0,042	0,094	0,717	0,067
		96/101 = 0,950	No	3/5 = 0,6	92/96 = 0,958			
11	44/101 = 0,436		Yes	0/44 = 0	6/57 = 0,105	0,492	0,211	0,104

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

		57/101 = 0,564	No	44/44 = 1	51/57 = 0,895			
12	49/101 = 0,485		Yes	4/49 = 0,082	2/52 = 0,038	0,500	0,086	0,043
		52/101 = 0,515	No	45/49 = 0,918	50/52 = 0,962			
13	35/101 = 0,347		Yes	29/35 = 0,829	0/66 = 0	0,453	1,657	0,750
		66/101 = 0,653	No	6/35 = 0,171	66/66 = 1			
14	34/101 = 0,337		Yes	3/34 = 0,088	3/67 = 0,045	0,447	0,087	0,039
		67/101 = 0,663	No	31/34 = 0,912	64/67 = 0,955			
15	33/101 = 0,327		Yes	4/33 = 0,121	2/68 = 0,029	0,440	0,184	0,080
		68/101 = 0,673	No	29/33 = 0,879	66/68 = 0,971			

A. KNN Algorithm

In this stage, the KNN algorithm is used to produce a Confusion Matrix value which consists of Accuracy, Recall and Possion values. The following is an explanation of the Confusion Matrix calculation :

Table 6
Confusion Matrix

Correct Classification	Classification As	
	Yes	No
Yes	TP	FN
No	FP	TN

Accuracy is a value that represents how precise a model is in performing the correct classification or how precise the success rate of the model is in predicting the actual data. In other words, Accuracy is how close the predicted results of a model are to the actual data and this process is referred to as the process of measuring precision and error. The Accuracy equation can be seen as follows (Saleh et al., 2019) :

$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{Total\ Data} \text{ X } 100\% \quad (8)$$

Precision is a measure of the accuracy of the classification process by comparing positive classification results (TP) to all positive predictions. In other words, Precision is used to measure how precise the model is in identifying objects that are actually positive. The equation of Precision itself can be seen as follows (Kurniadi et al., 2021) :

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

Recall is how many correct predictions from all positive classes. Recall value can be calculated using the following equation (Informasi et al., 2023) :

$$Recall = \frac{\text{TP}}{(\text{TP+FN})} \quad (10)$$

Details of the Confusion Class (Yes and NO) calculation can be found in the table below:

Table 7

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

"Yes" Class Confusion Knn Algorithm

4 (TP)	2 (FN)
90(FP)	5 (TN)

Table 8
"No" Class Confusion Knn Algorithm

90 (TP)	5 (FN)
4 (FP)	2 (TN)

Calculation of Accuracy, Precision, Recall and F-Measure values, namely:

$$\text{Accuracy} = \frac{TP+TN}{(TN+TP+FP+FN)} = \frac{90+2}{(2+90+4+5)} = \frac{92}{101} = 0{,}9108091 = 91{,}0891$$

$$\textit{Precision} = \frac{TP}{(TP+FP)} = \frac{90}{90+4} = \frac{90}{94} = 0{,}9574$$

$$\textit{Recall} = \frac{TP}{(TP+FN)} = \frac{90}{90+5} = \frac{90}{95} = 0{,}9473$$

$$\textit{F-Measure} = \frac{2xPrecisionxRecall}{Precision+Recall} = \frac{2x\ 0{,}9574\ x\ 0{,}9473}{0{,}9574+0{,}9473} = \frac{2\ x\ 0{,}9069}{1{,}9047} = \frac{1{,}8138}{1{,}9047} = 0{,}9522$$

After getting the results of each Precision, Recall and F-Measure value, the next step is to calculate the average value of the class (Weighted Avg). Weighted Avg calculation can be done using the following formula :

$$\textit{Weighted Avg (Precision)} = \frac{0{,}957\times A+0.286\times B}{101} = \frac{0{,}957\times 95+0.286\times 6}{101} = 0{,}917139$$

$$\textit{Weighted Avg (Recall)} = \frac{0{,}947\times \text{A}+0.333\times \text{B}}{101} = \frac{0{,}947\times 95+0.333\times 6}{101} = 0{,}910525$$

$$\textit{Weighted Avg (F-Measure)} = \frac{0{,}952\times \text{A}+0.308\times \text{B}}{101} = \frac{0{,}952\times 95+0.308\times 6}{101} = 0{,}913743$$

Weighted Avg calculation is the average result of each Precision, Recall and F-Measure value to get the Confusion Matrix value. Table 7 is a table of Accuracy Value based on Confusion Matrix of KNN Algorithm.

Table 9
Accuracy Value Based On Confusion Matrix Knn Algorithm

Class	*Precision*	*Recall*	*F-Measure*
Yes	0.957	0.947	0.952
No	0.286	0.333	0.308
Weighted Avg	0.918	0.911	0.914
Yes	0.957	0.947	0.952

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

5. DISCUSSIONS

From the calculations that have been carried out on the two algorithms, the accuracy results of each algorithm are 94.0594% on the CART Algorithm with a Precision value of 0.941%, a Recall value of 1.000% and an F-Measure value of 0.969%. The accuracy value of the KNN algorithm is 91.0891% with a Precision value of 0.957%, 0.947% for the Recall value and 0.952% for the F-Measure value. Judging from the accuracy value obtained, the CART algorithm has a higher value than the KNN algorithm, with the accuracy value of the CART algorithm being 94.0594% and the accuracy value of the KNN algorithm being 91.0891%. Based on the research review above, it also has the same opinion that the use of the CART method has several advantages for predicting compared to other classification methods, because the results are easy to interpret.

Table 10
Comparison of Accuracy Value of CART and KNN Algorithms

Class	*Precision*	*Recall*	*F-Measure*
CART	0.941	1.000	0.969
KNN	0.957	0.947	0.952

6. CONCLUSION

So, after processing the data using the CART and KNN algorithms, results can be obtained which show that the performance of the two algorithms is very good in predicting mental health. The comparison of the two methods is in accordance with the characteristics of the data used, which is relatively complex and has high variation. In this case, the CART algorithm showed higher performance in analyzing early mental health predictions in students who did not take steps to seek mental health support. This can be explained because the CART algorithm is more effective in handling data that has many variables and complex interactions. Based on the results of the research that has been conducted, this study provides an alternative mental health prediction method that is more accurate and effective, and provides recommendations for mental health researchers to use the CART algorithm in predicting early mental health. Suggestions for future research are to conduct further development and testing on the CART and KNN algorithms, as well as comparing with other methods to improve the accuracy and effectiveness of mental health prediction.

7. REFERENCES

Al Aziz, A. A. (2020). Hubungan Antara Intensitas Penggunaan Media Sosial dan Tingkat Depresi pada Mahasiswa. *Acta Psychologia*, *2*(2), 92–107. https://doi.org/10.21831/ap.v2i2.35100

Azania, D., & Naan. (2021). Humanistika: Jurnal Keislaman. *Humanistika: Jurnal Keislaman*, *7*(1), 26–45.

Binjori, A. S. (2020). Implementasi Data Mining Untuk Pengembangan Sistem Rekomendasi Pemilihan SMK Dengan Menggunakan Algoritma Cart. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *1*(2), 42–48. http://djournals.com/klik/article/view/46

Informasi, D. A. N. T., Muallif, I. S., Budiman, H., & Ransi, N. (2023). *Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor*. *1*(1), 297–306.

Kurnia, D. D. (2021). Sistem Pakar Untuk Mendiagnosa Gangguan Kesehatan Mental Menggunakan Algoritma Genetika. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *8*(3), 1171–1187. https://doi.org/10.35957/jatisi.v8i3.1079

Kurniadi, R. P., Saedudin, R. R., & Widartha, V. P. (2021). Perbandingan Akurasi Algoritma K-Nearest Neighbor Dan Logistic Regression Untuk Klasifikasi Penyakit Diabetes. *E-Proceeding of Engineering*, *8*(5), 9757–9764.

Mardiani, E., Rahmansyah, N., Ningsih, S., & ... (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of …*, *3*(2), 13880–13892. http://j-innovative.org/index.php/Innovative/article/view/1949%0Ahttp://j-innovative.org/index.php/Innovative/article/download/1949/1468

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4232

Submitted : 5 July 2024
Accepted : 14 July 2024
Published : 22 July 2024

Masyarakat, J. P., Putri, A., Maria, C., Syahrias, L., & Mustika, I. (2023). *Penyuluhan Mental Health*. *6*(1), 154–161.

Melani, F., & Sulastri. (2020). Analisis Perbandingan Klasifikasi Algoritma CART dengan Algoritma C 4 . 5 Pada Kasus Penderita Kanker Payudara. *Jurnal TEKNO KOMPAK*, *17*(1), 171–183.

Prakoso, R. D. Y., Wiriaatmadja, B. S., & Wibowo, F. W. (2020). Sistem Klasifikasi Pada Penyakit Parkinson Dengan Menggunakan Metode K-Nearest Neighbor. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, *2016*, 63–68.

Praptikaningtyas, A. A. I., Wahyuni, A. A. S., & Aryani, L. N. A. (2019). Hubungan Tingkat Depresi Pada Remaja Dengan Prestasi Akademis Siswa SMA Negeri 4 Denpasar. *Jurnal Medika Udayana*, *8*(7), 1–5. https://ojs.unud.ac.id/index.php/eum

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

Radiani, W. A., Ushuluddin, F., Islam, U., & Antasari, N. (2019). *KESEHATAN MENTAL MASA KINI DAN PENANGANAN Pendahuluan Setiap hari melalui media informasi baik cetak ataupun elektronik , kekerasan selalu dalam muncul rumah berita tangga , tragedi pelecehan seksual , prostitusi , dan beragam bentuk kejahatan yang lain* . 87–113.

Rahma, Y., Prasetiadi, A., & Wibowo, M. (2022). Identification of Mental Ilness From Patient Diseases Using Knn and Levenshtein Distance Algorithm. *Jurnal Teknik Informatika (Jutif)*, *3*(5), 1363–1372. https://doi.org/10.20884/1.jutif.2022.3.5.371

Saleh, H., Faisal, M., & Musa, R. I. (2019). Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor. *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, *4*(2), 120–126. https://doi.org/10.51876/simtek.v4i2.60

Santiko, I., & Subarkah, P. (2019). Comparison of Cart and Naive Bayesian Algorithm Performance to Diagnose Diabetes Mellitus. *IJIIS: International Journal of Informatics and Information Systems*, *2*(1), 9–16. https://doi.org/10.47738/ijiis.v2i1.9

Savitri, F. D. (2022). Penerapan Metode Cart Dalam Memprediksi Penjualan Produk Fast Moving Dan Slow Moving. *Journal of Informatics, Electrical and Electronics Engineering*, *1*(4), 119–125. https://djournals.com/jieee

Subarkah, P., Ikhsan, A. N., & Setyanto, A. (2018). The Effect of The Number of Attributes On The Selection of Study Program Using Classification and Regression Trees Algorithms. *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 1–5. https://doi.org/10.1109/ICITISEE.2018.8721030

Sudarsono, B. G., & Lestari, S. P. (2020). Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Penelitian Ilmiah Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Media Informatika Budidarma*, *4*(4), 1094–1099. https://doi.org/10.30865/mib.v4i4.2448

Wisnugraha, W. S., Farida, I. N., & ... (2023). Implementasi Algoritma Naïve Bayes Dalam Menentukan Diagnosa Tingkat Depresi Mahasiswa Akhir Terhadap Pengerjaan Skripsi. *Prosiding SEMNAS …*, *7*, 919–928. https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/3517%0Ahttps://proceeding.unpkediri.ac.id/index.php/inotek/article/download/3517/2321

* Corresponding author