JOURNAL LA MULTIAPP

VOL. 04, ISSUE 06 (253-259), 2023
DOI: 10.37899/journallamultiapp.v4i6.1005

# Prediction of Elementary School Students' Mental Health using Decision Tree Algorithm with K-Fold Cross-Validation in Bone Bolango Regency, Gorontalo Province

**Salahuddin Liputo[1], Franky Tupamahu[2], Wahyudin Hasyim[2], Sri Ariyanti Sabiku[3], Rahmawaty Parman[1], Aan Hanapi[1]**

*[1]Psychology Department, Universitas Muhammadiyah Gorontalo*
*[2]Information Systems Department, Universitas Muhammadiyah Gorontalo*
*[3]Tourism Department, Universitas Negeri Gorontalo*

**Corresponding Author: Salahuddin Liputo*

Article Info

*Article history:*
Received 15 November 2023
Received in revised form 8 December 2023
Accepted 27 December 2023

*Keywords:*
Mental Health 2
SEHS-S
Machine Learning
Decision Trees

***Abstract***

*Mental health is a fundamental component of the World Health Organization's definition of health, encompassing not only freedom from illness but also well-being in physical, mental, and social dimensions. In today's modern society, mental health has become a paramount issue, as its soundness enables individuals to realize their own potential, cope with normal life pressures, work productively, and contribute effectively to their communities. In Indonesia, mental health-related challenges are associated with the absence of a reliable mental health detection tool. Conversely, abroad, there has been a substantial amount of research focused on innovative technology-based mental health detection using Machine Learning. This study aims to predict mental health using the Social Emotional Health Survey-Secondary (SEHS-S) as the evaluation criterion for prediction through Machine Learning. The Decision Tree algorithm is employed, and the prediction model is tested using K-Fold Cross-Validation, resulting in 8 folds with an accuracy rate of 78.61%.*

## Introduction

Mental health is a fundamental component of the World Health Organization's (WHO) definition of health. Globally, over the past three decades, mental health has emerged as a central issue in health development. For several decades, the WHO has emphasized that the definition of health is integral, signifying more than just freedom from illness. It encompasses a state where an individual achieves complete well-being in physical, mental, and social aspects (Ridlo & Zein, 2018). In the context of mental health, a positive condition allows individuals to realize their own potential, cope with normal life pressures, work productively, and contribute effectively to the communities in which they reside (World Health Organization, 2013).

Referring to the statement above, it is imperative that mental health care receives serious attention from the government given the substantial impact of weakened mental conditions on society. Mental health issues can no longer be regarded as peripheral concerns in health development (Ridlo & Zein, 2018). Specifically addressing the problem of Emotional Mental Disorders, there is an increasing prevalence of affected individuals in Indonesia. According to the 2018 Riskesdas data for the population aged > 15 years, the prevalence was 9.8% (approximately 19 million people), compared to 6% in the 2013 Riskesdas data. The challenges include the limited capacity of human resources capable of implementing prevention and control measures for emotional mental disorders, the absence of a comprehensive recording



*ISSN: 2716-3865 (Print), 2721-1290 (Online)*

and reporting system related to emotional mental disorders, and suboptimal coordination across programs and sectors in efforts to prevent and control depression (Riset Kesehatan Dasar, 2018).

In addressing the issues of early detection of mental health and substance abuse, these serve as new indicators, necessitating widespread socialization efforts in each region. The challenge is associated with the absence of established guidelines for the early detection of mental health (Riset Kesehatan Dasar, 2018). Given that mental health has become a primary concern in modern society, research efforts have been directed towards creating digital methods for monitoring mental health and mood states (Mastoras et al., 2019). Presently, the utilization of innovative technologies such as machine learning, big data, and artificial intelligence (AI) has been developed as adopted approaches for treatment, intervention, and psychological diagnosis, experiencing dramatic growth over the past few years (Zhou et al., 2022).

For instance, Vaishnavi et al.'s 2021 study (Vaishnavi et al., 2022), which utilized five machine learning techniques, yielded a prediction accuracy of 81.75%. Another study by Bedi et al. in 2015 (Bedi et al., 2015) employed semantic coherence and syntactic complexity to predict psychosis development, resulting in a 100% accuracy. Further research by the same team, applying the same approach to a larger sample and developing similar machine learning classifications, achieved an accuracy of 83% in predicting the onset of psychosis and 72% accuracy in distinguishing the speech of patients with new-onset psychosis from that of healthy individuals (Corcoran et al., 2018).

One of the prevalent methods for predicting mental health using machine learning is the Decision Tree algorithm. Research on the application of the Decision Tree algorithm for mental health prediction has been conducted, including a study by Laijawala et al., 2020 (Laijawala et al., 2020). Their findings indicated that the Decision Tree algorithm is the most optimal, demonstrating low execution time and high accuracy, specifically 82.2%. Subsequently, another study by Vaishnavi et al., 2021, employed five machine learning techniques, one of which was the Decision Tree algorithm. The combined accuracy of these five techniques resulted in a mental health prediction accuracy of 81.75%.

The objective of this research is to identify the mental health of elementary school students using machine learning with the Decision Tree algorithm method. The rationale for selecting elementary school students aligns with the questionnaire used, namely the Social Emotional Health Survey-Secondary (SEHS-S) (Furlong et al., 2020), designed to identify the mental health of elementary school students. We endeavored to adapt the SEHS-S into a machine learning application. The steps we took included identifying SEHS-S parameters consisting of belief in self, belief in others, emotional competence, and engaged living. Subsequently, we integrated these parameters into a machine learning application using the Decision Tree algorithm method, which was then tested using K-Fold Cross-Validation.

The Decision Tree is employed to explore data and uncover hidden relationships between a set of potential input variables and the target variable. It is a structure used to partition a large dataset into smaller sets by applying a series of decision rules. With each decision branch, the members of the resulting set become more similar to each other. The Decision Tree is depicted as a flow-chart-like tree structure, where each internal node represents a test on an attribute, each branch indicates the result of the test, and leaf nodes represent classes or class distributions. In a decision tree, three types of nodes exist: a) Root Node, the topmost node that may have no input and can have zero or more outputs; b) Internal Node, a branching node with one input and at least two outputs; c) Leaf Node or Terminal Node, the final node with one input and no output.

K-Fold Cross-Validation (K-Fold CV) is a technique used when there is a limited amount of data. It involves dividing the dataset into K subsets, using K-1 subsets for training, and the



*ISSN: 2716-3865 (Print), 2721-1290 (Online)*

remaining subset for testing. This process is repeated K times, with each subset serving as the testing set exactly once. The performance metrics are then averaged over the K iterations to provide a more robust evaluation of the model's performance.

## Methodology

To predict the mental health of the subjects, we utilized the Social Emotional Health Survey-Secondary (SEHS-S), specifically developed for identifying the mental health of elementary school students. The SEHS-S comprises 36 items organized into 12 subscales representing positive social-emotional health constructs associated with four health domains. The first domain, belief in self, consists of three subscales based on the constructs of Social Emotional Learning (SEL): self-efficacy, self-awareness, and persistence. The second domain, belief in others, includes three subscales derived from constructs found in the literature on childhood resilience: school support, peer support, and family support. The third domain is emotional competence, comprising three subscales also based on SEL constructs, such as emotional regulation, empathy, and self-control. The final domain is engaged living, consisting of three subscales based on constructs from positive youth psychology literature, including gratitude, zest, and optimism.

All items in the SEHS-S were standardized to a four-point response scale (scored as follows: 1 = not true, 2 = somewhat true, 3 = mostly true, and 4 = very true). To ensure the validity of the research data, we initially conducted validity and reliability testing of the SEHS-S questionnaire by distributing it to a specific sample. Once the questionnaire was deemed valid, reliable, and exhibited a normal data distribution, we proceeded to redistribute it to predict the mental health of elementary school students in the Bone Bolango Regency. Upon collecting the data, we processed it using a Frequency Table Matrix and the Random Forest Algorithm. Following these steps, the classification and prediction processes were carried out using the Decision Tree Algorithm. Subsequently, testing was conducted using K-Fold Cross-Validation, and accuracy was measured using a confusion matrix to obtain prediction accuracy. The research flow described above is illustrated in the following diagram:

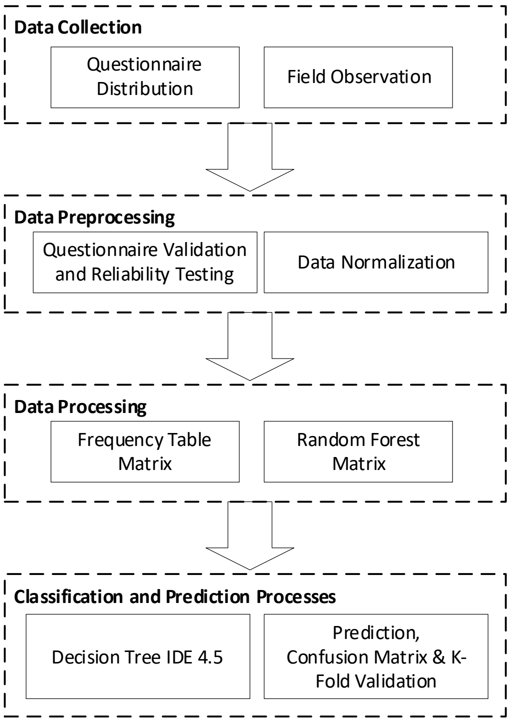



*Figure 1. Research flow*

### Data Processing



ISSN: 2716-3865 (Print), 2721-1290 (Online)

This methodology employs the decision tree to transform extensive datasets into rule representations in the form of decision trees. Decision trees serve as valuable tools for data exploration, revealing hidden relationships among numerous potential input variables and a target variable. Data within a decision tree is typically organized in tabular form with attributes and records, where attributes denote parameters known as criteria in tree formation.

A decision tree, an artificial intelligence method, takes the form of a tree where each branch offers choices and alternatives, with leaves indicating the decision made. This method is widely utilized due to its ability to amalgamate patterns, knowledge, and information into the structure of a decision tree.

In this research, data testing is conducted using RapidMiner, software utilized for classification employing various descriptive and classification techniques. This is intended to generate accurate information and knowledge for users, facilitating precise decision-making.

**Evaluation Method**

In testing the parameter "Number of Folds" in the Decision Tree algorithm, the range considered is from 2 to 100, as follows:

Table 1. Number of Folds and accuracy

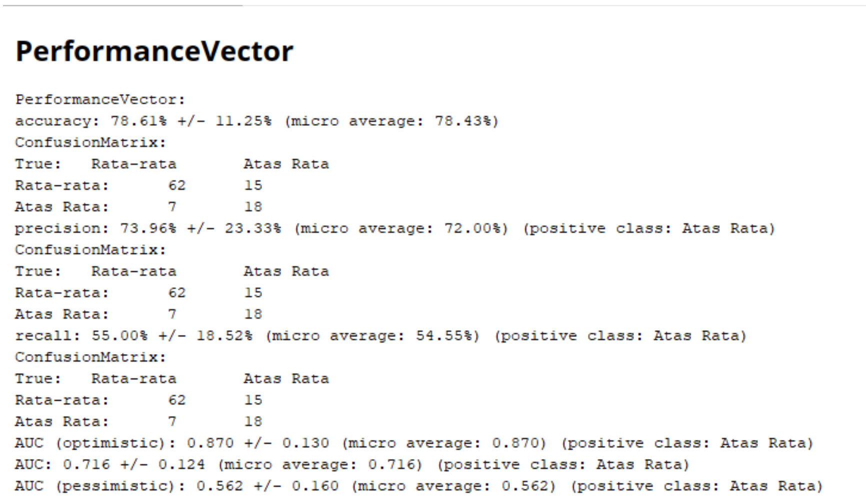
| **Number of Fold** | **Accuracy** |
|---|---|
| 2 | 76.47% |
| 3 | 72.55% |
| 4 | 77.50% |
| 5 | 76.52% |
| 6 | 78.43% |
| 7 | 77.62% |
| 8 | 78.61% |
| 9 | 73.40% |
| 10 | 78.18% |
| 20 | 75.33% |
| 30 | 74.44% |
| 40 | 70.42% |
| 50 | 69.33% |
| 60 | 69.17% |
| 70 | 70.00% |
| 80 | 67.50% |
| 90 | 67.22% |
| 100 | 67.50% |

Based on the test results presented in Table 1 above, the highest Accuracy value is achieved with 8 Folds, demonstrating an accuracy of 78.61%.

The performance metrics, measured using a confusion matrix to obtain prediction accuracy, indicate that the pattern obtained from testing the number of folds in the decision tree will be utilized to derive the decision tree.



*ISSN: 2716-3865 (Print), 2721-1290 (Online)*



## PerformanceVector

```
PerformanceVector:
accuracy: 78.61% +/- 11.25% (micro average: 78.43%)
ConfusionMatrix:
True:   Rata-rata       Atas Rata
Rata-rata:      62      15
Atas Rata:      7       18
precision: 73.96% +/- 23.33% (micro average: 72.00%) (positive class: Atas Rata)
ConfusionMatrix:
True:   Rata-rata       Atas Rata
Rata-rata:      62      15
Atas Rata:      7       18
recall: 55.00% +/- 18.52% (micro average: 54.55%) (positive class: Atas Rata)
ConfusionMatrix:
True:   Rata-rata       Atas Rata
Rata-rata:      62      15
Atas Rata:      7       18
AUC (optimistic): 0.870 +/- 0.130 (micro average: 0.870) (positive class: Atas Rata)
AUC: 0.716 +/- 0.124 (micro average: 0.716) (positive class: Atas Rata)
AUC (pessimistic): 0.562 +/- 0.160 (micro average: 0.562) (positive class: Atas Rata)
```

*Figure 2. Performance vector*

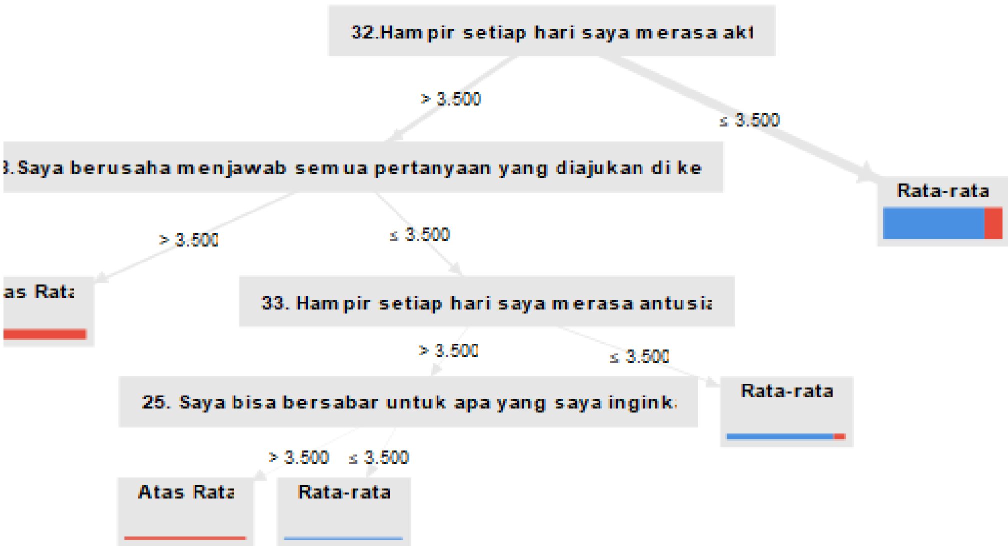



*Figure 3. Decision tree*

Based on all the research steps conducted in the data mining analysis of the decision tree algorithm for predicting the mental health of elementary school students, it can be concluded that from the 36 questionnaire items, a dataset was generated. This dataset was processed to produce rules as a reference for mental health prediction. The total data used consisted of 102 records with 36 variables. The results obtained from the RapidMiner testing align with the analysis of the decision tree algorithm, yielding an accuracy rate of 78.61%.

## Results and Discussion

Explain the results of the research in the form of problem-solving analyzed using relevant theories. The results of the study also revealed the findings of the research. Discussion is accompanied by logical arguments by linking the results of research with theory, the results of other studies.

The Decision Tree algorithm is utilized to detect the 12 subscales representing positive social-emotional health constructs associated with four health domains. Equations 1 through 4 are used to calculate precision, recall, specificity and F1 score values, resulting in the generation of a Confusion Matrix.



ISSN: 2716-3865 (Print), 2721-1290 (Online)

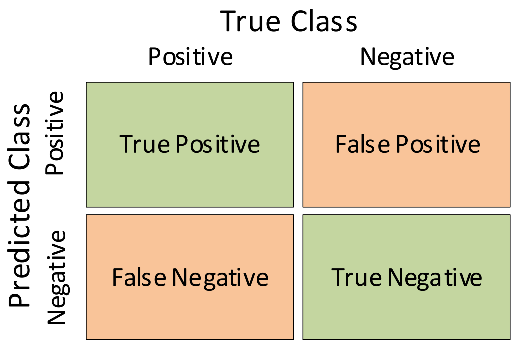



*Figure 4. Confusion Matrix*

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Postive} \tag{1}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2}$$

$$Specifity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{3}$$

$$F1\ Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{4}$$

**accuracy: 78.61% +/- 11.25% (micro average: 78.43%)**

| | true Rata-rata | true Atas Rata | class precision |
|---|---|---|---|
| pred. Rata-rata | 62 | 15 | 80.52% |
| pred. Atas Rata | 7 | 18 | 72.00% |
| class recall | 89.86% | 54.55% | |

*Figure 5. Confusion matrix*

The evaluation of the decision tree method can be assessed through the Confusion Matrix. In the above figure, it can be seen that the true positive prediction have a value of 62 and the false positive value is 15. The false negative is 7 and the true negative is 18. For the "Rata-rata" (average) class, the precision is 80.52%, and for the "Atas rata" (above average) class, the obtained precision is 72.00%.

## Conclusion

In this study, the algorithm utilized for determining the level of mental health is the Decision Tree machine learning algorithm. The dataset contains general and basic information about the community, along with the SEHS-S questionnaire. F1 scores are extracted to identify the most suitable model for predicting mental health. By applying the results of the RapidMiner testing, the same outcome is obtained as in the analysis of the decision tree algorithm, yielding an accuracy rate of 78.61% with the testing of 8 Folds.

## References

Ridlo, I. A., & Zein, R. A. (2018). Arah Kebijakan Kesehatan Mental: Tren Global dan Nasional Serta Tantangan Aktual. *Buletin Penelitian Kesehatan*, *46*(1), 45-52.

World Health Organization (2013). Mental Health Action Plan/MHAP.

Riset Kesehatan Dasar (Riskesdas) (2018). Badan Penelitian dan Pengembangan Kesehatan Kementerian RI tahun 2018.



*ISSN: 2716-3865 (Print), 2721-1290 (Online)*

Mastoras, R. E., Iakovakis, D., Hadjidimitriou, S., Charisis, V., Kassie, S., Alsaadi, T., ... & Hadjileontiadis, L. J. (2019). Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Scientific reports, 9*(1), 1-12.

Zhou S, Zhao J and Zhang L (2022) Application of Artificial Intelligence on Psychological Interventions and Diagnosis: An Overview. *Front. Psychiatry 13*:811665. doi: 10.3389/fpsyt.2022.811665

Vaishnavi, K., Kamath, U. N., Rao, B. A., & Reddy, N. S. (2022). Predicting mental health illness using machine learning algorithms. *In Journal of Physics: Conference Series. 161* (10), 012021. IOP Publishing.

Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., & Corcoran, C. M. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia, 1*(1), 1-7.

Corcoran CM, Carrillo F, Slezak DF, Bedi G, Klim C, Javitt DC, Bearden CE, Cecchi GA. (2018*). Prediction of psychosis across protocols and risk cohorts using automated language analysis*. World Psychiatry, 17:67-75

Laijawala, V., Aachaliya, A., Jatta, H., & Pinjarkar, V. (2020*). Classification algorithms based mental health prediction using data mining*. In 2020 5th international conference on communication and electronics systems (ICCES) (pp. 1174-1178). IEEE.

Furlong, M. J., Nylund-Gibson, K., Dowdy, E., Wagle, R., Hinton, T., & Carter, D. (2020). *Modification and standardization of Social Emotional Health Survey Secondary—2020 edition*. Santa Barbara, CA, University of California Santa Barbara, International Center for School Based Youth Development.



*ISSN: 2716-3865 (Print), 2721-1290 (Online)*