JOURNAL LA MULTIAPP

VOL. 05, ISSUE 01 (019-024), 2024
DOI: 10.37899/journallamultiapp.v5i1.853

Prediction of Mental Health of Elementary School (SD) Students using the Decision Tree Algorithm with K-Fold CV testing in Bone Bolango Regency, Gorontalo Province

Salahuddin Liputo[1], Franky Tupamahu[1]

[1]*Psychology Department, Universitas Muhammadiyah Gorontalo*

**Corresponding Author: Salahuddin Liputo*

Article Info

Article history:
Received 7 October 2023
Received in revised form 3 November 2023
Accepted 29 December 2023

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

Abstract

Mental health is a fundamental component of the WHO definition of health, which means not only being free from disease but also being physically, mentally and socially healthy. Currently, mental health has become a major issue in modern society because if it is good it will enable us to realize our own potential, overcome the normal stresses of life, work productively, and be able to contribute to the society in which we live. In Indonesia, problems related to mental health are related to the lack of mental health detection tools. Meanwhile abroad, much research has been developed regarding mental health detection based on innovative technology using Machine Learning. This research aims to predict mental health using the Social Emotional Health Survey-Secondary (SEHS-S) as a prediction evaluation criterion using Machine Learning with the Decision Tree algorithm method with K-Fold CV testing. The sample in this research was elementary school students in Bone Bolango Regency, Gorontalo Province

Introduction

Mental health is a fundamental component of the WHO definition of health. Globally, over the last three decades, mental health has been a central issue in health development. Since several decades ago, WHO has emphasized that the definition of health is an integral definition; it means not just being free from disease,

but a condition where a person achieves complete well-being physically, mentally and socially (Ridlo & Zein, 2018). Regarding mental health, if it is in good condition, it allows us to realize our own potential, overcome the normal stresses of life, work productively, and be able to contribute to the society in which we live, World Health Organization, (2013).

Referring to the statement above, the management of mental health should receive serious attention from the Government considering how big the impact of the weak mental condition of society is. Mental health problems can no longer be considered a peripheral issue in health development (Ridlo & Zein, 2018). The problem of mental emotional disorders (GME) is the increasing prevalence of sufferers in Indonesia. Based on the 2018 Riskesdas, the population aged > 15 years was 9.8% (around 19 million people), compared to the 2013 Riskesdas of 6%, limited human resource capacity capable of carrying out efforts to prevent and control GME, a recording and reporting system related to GME has not yet been developed, there is not yet optimal coordination between programs and sectors in efforts to prevent and control depression (Riset Kesehatan Dasar, 2018).

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

The problem of early detection of mental health (Keswa) and drugs is a new indicator, so it needs to be disseminated to every region. The problem is related to the lack of guidelines for early detection of children (Riset Kesehatan Dasar, 2018). Since mental health is becoming a major issue in modern society, research efforts have been made in a big way

created digital methods for monitoring mental health and mood states (Mastoras et al., 2019). Currently, the use of innovative technologies such as machine learning, big data, and artificial intelligence (AI) has developed as an approach adopted for treatment, intervention and psychological diagnosis and has developed dramatically over the last few years (Zhou et al., 2022). For example, research by Vaishnavi et.al, (2021), which uses 5 (five) machine learning techniques produces a prediction accuracy of 81.75%. Then research using semantic coherence and syntactic complexity to predict the development of psychosis by Bedi et.al, (2015), resulted in 100% accuracy.

One method of predicting mental health using machine learning that is currently widely used is the Decision Tree algorithm. Research regarding the use of decision tree algorithms to predict mental health has been carried out, including by Laijawala et.al, 2020 [8], the results of which are that the decision tree algorithm is the most optimal because of its low execution time and

high accuracy, namely 82.2%. Then the next one is by Vaishnavi et.al, (2021), which uses 5 (five) machine learning techniques, one of which is the decision tree algorithm, where these five techniques produce mental health prediction accuracy of 81.75%.

The aim of this research is to identify mental health in elementary school students using machine learning with the Decision Tree algorithm method. The reason for selecting elementary school students was in accordance with the questionnaire used, namely the Social Emotional Health Survey-Secondary (SEHS-S) developed by Furlong (2020), to identify the mental health of elementary school students. We try to adapt SEHS-S to machine learning applications. The steps we took were to identify the SEHS-S parameters which consist of belief in yourself, belief in others, emotional competence and engaged living.

Decision trees are a data mining tool that may uncover latent connections between several potential inputs and outputs. With a decision tree, you may break down massive datasets into more manageable chunks by applying a series of decision rules; as you go down the tree, the records in each subset become more and more similar to one another. A decision tree is a kind of tree structure similar to a flowchart in which each internal node represents an attribute test, each branch represents the results of the attribute test, and the leaf nodes represent the classes or class distribution. There are three distinct kinds of nodes in a decision tree, and they are: The root node, or the top node, may have no inputs and several outputs, or none at all. There is only one input to an Internal Node, but at least two possible outputs. There is just one input and no output at the leaf node, also known as the terminal node.

K-fold cross validation (K-Fold CV) is a technique that can be used if you have a limited amount of data.

Methods

We utilized the Social Emotional Health Survey-Secondary (SEHS-S), a tool designed to assess adolescents' emotional well-being, to make inferences about the participants' likely state of mind. The SEHS-S measures positive social-emotional health components across four health domains with 36 items distributed over 12 subscales. Confidence in oneself is measured along three subscales based on Social Emotional Learning (SEL) concepts: self-efficacy, self-awareness, and perseverance. Trust in others is the second domain, and it has three subscales: school support, peer support, and family support, all of which are based on concepts from the literature on childhood resilience. Emotional competence is the third dimension, and it has three subscales based on SEL characteristics including emotional regulation, empathy, and self-

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

control. Finally, the engaged living domain includes three subscales based on concepts from the positive youth psychology literature: appreciation, energy, and optimism, as well as regulation and self-discipline. Finally, the engaged living domain includes three subscales based on concepts from the positive youth psychology literature: appreciation, energy, and optimism, as well as regulation and self-discipline. Engaged life is the last domain, and it consists of three subscales derived from positive adolescent psychology literature, including thankfulness, energy, and hope.

Value 1 = not true, value 2 = slightly true, value 3 = moderately true, and value 4 = extremely true) is used for all items on the SEHS-S. Before collecting any data for the study, we made sure the SEHS-S questionnaire was legitimate by administering it to representative samples of the population. In order to forecast the mental health of primary school kids in Bone Bolango Regency, we redistributed the questionnaire after ensuring its validity, relabeling, and normal data distribution. The data is processed using the Frequency Table Matrix and the Random Forest Algorithm once it has been gathered. The Decision Tree Algorithm is then used to do classification and prediction, followed by K-Fold CV validation and confusion matrix measurement to determine the accuracy of the predictions made. The following diagram illustrates the aforementioned study procedure:

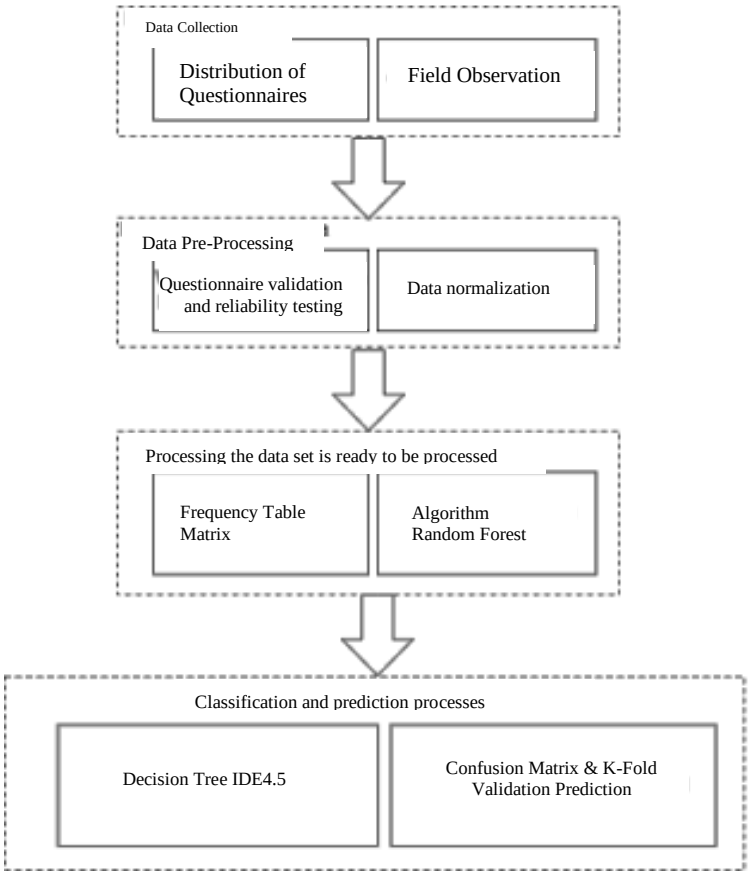

Figure 1. Study Procedure

Data Processing

Pre-processing is an important phase in handling data sets with Machine Learning. This preprocessing phase focuses on data cleaning, i.e. removing unfilled rows from the data set. This technique is not appropriate for handling this particular data set. Another technique is to calculate the average of a particular column or columns to fill the NAN cases. There are other techniques to fill in missing values. Namely median and mode. Here the mean is applied to the data set. Initial data processing is carried out by filling in the NAN value in the dataset with the Mean in a certain field.

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

Evaluation Method

The preprocessed data is retrieved, and the features are extracted. The dataset is divided into an 80:20 ratio representing 80% of the data used to train the algorithm and the remaining 20% for testing data. A classification algorithm is applied to a data set to classify the data.

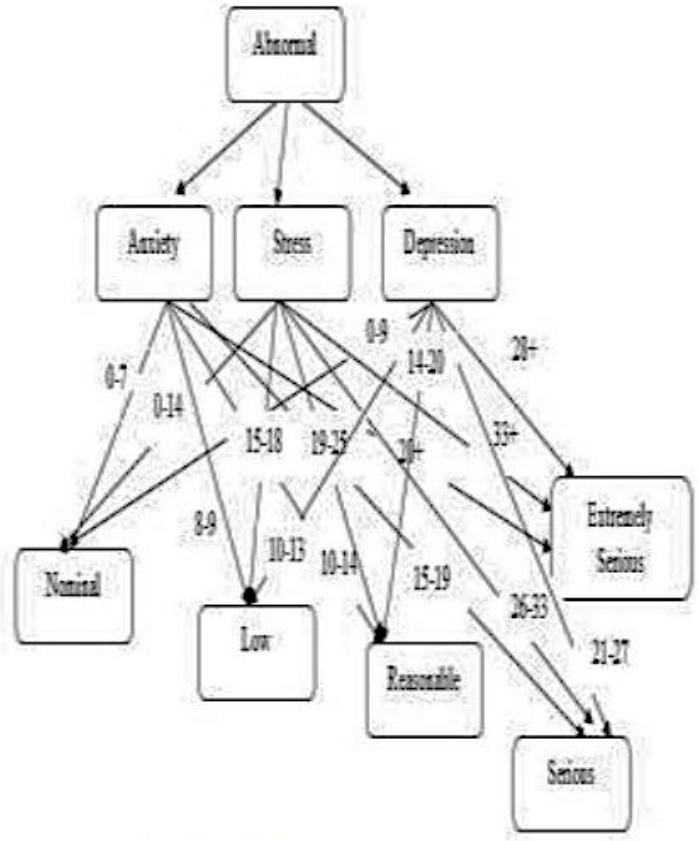

Figure 2. Decision Tree Classifier

Results and Discussion

Decision tree and Random Forest algorithms were used to detect 12 subscales representing positive social emotional health constructs related to four health domains. Based on this, a confusion matrix is generated.

Equations 1 to 6 are used to manipulate the values of accuracy, precision, recall, error rate, and specificity to produce CM - confusion matrix.

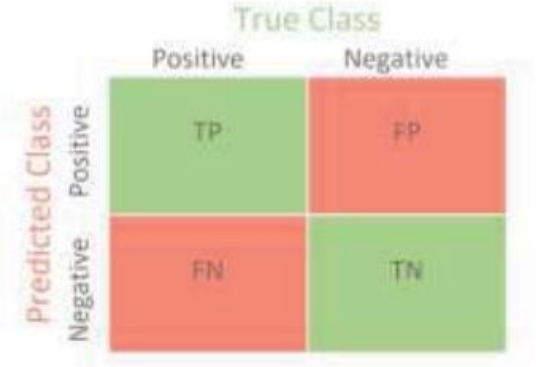

Figure 3. Confusion Matrix

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

1, 2, 3, 4, 5 represent severity levels. namely normal, low, reasonable, serious, very serious. in the table below.

$$\text{Αχχυραχψ Ρατε (AP)} = \frac{Summazation of Diagonals(TP)}{Total count} \quad (1)$$

$$\text{Ερρορ Ρατε (EP)} = 1 - \text{AP} \quad (2)$$

$$\text{Πρεχισιον} = \frac{True Positive}{True Positive + False Positive} \quad (3)$$

$$\text{Ρεχαλλ} = \frac{True Positive}{True Positive + False Negative} \quad (4)$$

$$\text{Σπεχιφιχιτψ} = \frac{True Negative}{True Negative + False Positive} \quad (5)$$

$$\text{Φ1 Σχορε} = \frac{2 * (\text{Pr}ecision * \text{Re}call)}{\text{Pr}ecision + \text{Re}call} \quad (6)$$

In fact, True positive = Diagonal of the matrix

False Negative = Leaving the TP of that class, the number of rows is stable for that class

False Positive = Does not include the TP of the class, the sum of the Equivalent classes.

True Negative = Does not include classes, complete number of rows and columns.

From the answers of 100 subjects, we estimate that on average all subjects have very good self-confidence referring to the percentage of answers given, namely 48% have good self-efficacy, 50% have very good self-awareness, and 45% have perseverance (persistence) is very good.

From the answers of 100 subjects, we estimate that on average all subjects have very good trust in other people, referring to the percentage of answers given, namely 55% have very good school support, 55% have peer support) which is very good, and 50% family support which is very good.

From the answers of 100 subjects, we estimate that on average all subjects have very good emotional competence referring to the percentage of answers given, namely 53% have very good emotional regulation, 50% have very good empathy, and 62% have very good self-control.

From the answers of 100 subjects, we estimate that on average all subjects have very good emotional competence referring to the percentage of answers given, namely 67% have very good gratitude, 36% have very good zest, and 62% have very good optimism.

Conclusions

In this research, to determine the level of mental health, a machine learning algorithm, namely the Decision Tree algorithm, was used. The dataset contains general and basic community information along with the SEHS-S questionnaire. The F1 score is taken to identify which model is best suited for mental health prediction.

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

References

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.

Furlong, M. J., Nylund-Gibson, K., Dowdy, E., Wagle, R., Hinton, T., & Carter, D. (2020).

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

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.

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.

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

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*, *2161* (1), 012021. IOP Publishing.

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

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

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