2nd International Conference on Information Science and Technology Innovation (ICoSTEC)
February, 25 2023. Yogyakarta, Indonesia

Predicting Children's Talent Based On Hobby Using C4.5 Algorithm And Random Forest

Sugeng Riyadi[1], Hartono[2], Wanayumini[3]

[1,2,3]*Magister of Computer Science, Potensi Utama University*
JL. KL. Yos Sudarso Km. 6,5 No. 3-A, Medan

[1]adhie_ogenk@yahoo.co.id

[2]hartonoibbi@gmail.com

[3]wanayumini@gmail.com

***Abstract*— A person's talent is closely related to intelligence, hobbies, and interests. These factors are the best features to be used in a dataset to predict a children's talent, such as in an academy, arts, or sports. This research uses the C4.5 and random forest algorithms in 8 different models to predict a children's talent based on a dataset gained from a survey involving 1601 parents. Each model contains four training-testing data ratios, such as 50:50, 60:40, 70:30, and 80:20. We calculate each model prediction performance using 10-fold and 20-fold cross-validation, with the accuracy, f-score, precision, and recall values as a comparison. The best result for the training evaluation we get is 91.5% for each comparison value from the random forest model (70:30 ratio) using a 20-fold cross-validation. For the testing evaluation, we get 92.7%, 92.8%, 92.8%, and 92.7% from the random forest model (50:50 ratio). The worst testing evaluation we get is 81.7% for each comparison value from the C4.5 model (50:50 ratio) using a 20-fold cross-validation. For the testing evaluation, we get 89.2%, 89.2%, 89.3%, and 89.2% from the C4.5 model (50:50 ratio).**

***Keywords*___ C4.5 algorithm, machine learning, random forest algorithm, talent prediction**

I. INTRODUCTION

Hobby (avocation) is one of the benchmarks in assessing a person's talent, where the activities, skills, and knowledge that a person has when doing his favorite will encourage the formation of creative readiness [1]. There are types of intelligence found in people, such as naturalist intelligence, existential intelligence, spiritual intelligence, linguistic aptitude, mathematical logic, kinesthetic intelligence, and intrapersonal-interpersonal communication, which are closely related to someone's talent [2]. The types of intelligence that people possess will influence their talents in academics, arts, and sports. This research focuses on predicting these three talents based on children's hobbies and intellectual levels.

In predicting a person's talent, especially in children, machine learning is often used based on a dataset containing data related to their level of intelligence [3]. Studies about the implementation of machine learning in predicting or classifying someone's academic talent are enough to show that it is feasible to use this method to help determine a children's talent based on their hobbies and activities. Some of these studies include college students' talent classification based on classroom behavior using the convolutional neural network algorithm [4], predicting academic talent capacity using the decision tree algorithm [5], and and using a neural network to predict students' academic performance [6].

Some studies on sports talent assessment using machine learning show that machine learning is an excellent tool to predict and analyze someone's sports talent. These studies include soccer talent assessment with the support vector machine algorithm [7]; netball players' talent assessment with the decision tree, neural network, and linear regressions algorithm [8]; and professional goalkeeper classification with the logistic regression, gradient boosting, and random forest algorithm [9].

In the field of children's talent prediction using machine learning, we use some research as references in this study: predicting middle school students' programming talent with the Artificial Neural network (ANN) algorithm, with the best results being the R-value for training data = 9.72284e-1 and R-value for testing data = 9.12687e-1 [10]; predicting youth tennis players' talent based on their motoric abilities using multilayer perceptron (MLP), achieving an accuracy value of 91%, sensitivity value of 75%, and specificity value of 84% [11]; and using the random forest algorithm to predict students' academic achievement, with an overall accuracy value of 85.03% [12]. From this research, we saw the feasibility of using machine learning algorithms to predict children's talents based on their hobbies and abilities.

This research uses the C4.5 and random forest algorithm to predict children's talent based on their hobbies and activities into three categories: academic, art, and sports talent We chose the C4.5 algorithm because the rules it produces are easy to interpret, proven to have excellent accuracy, and can handle both discrete and numeric variables [13]. We also chose the random forest algorithm to compare with the C4.5 algorithm because it uses a combination of decision trees that work similarly to the C4.5 algorithm, can process data quickly, and has high accuracy [14].

This research uses the C4.5 and random forest algorithm to predict children's talent based on their hobbies and activities into three categories: academic, art, and sports talent. Both algorithms are used in 8 models to process the dataset in 4 different ratio formats, such as 50:50, 60:40, 70:30, and 80:20. We then calculate and analyze the prediction results using 10-fold and 20-fold cross-validation to choose the best model.

II. METHODS

A. C4.5

The C4.5 algorithm is a type of decision tree-based machine learning built recursively until each of the tree's sections consists of data from the same class [15]. The C4.5 algorithm creates a decision tree by selecting the attribute to be used as the tree's root, creating branches for each value and then dividing the problem into one of the branches, and finally repeating the process until all cases in each tree's branch have the same class [16]. In determining the features used as breaking nodes in the tree, the C4.5 algorithm uses a gain criterion with a formula shown in equation (1) [17].

$$\mathrm{Gain}\ (S, A) = \mathrm{Entropy}(S) - \sum_{i=1}^{n} \frac{S_i}{S} * Entropy\ (S_i) \quad (1)$$

This algorithm offers many benefits in solving prediction problems, such as the ability to process numerical and discrete data, manage missing attribute values, produce simple rules that are easy to understand, and it is one of the fastest performance algorithms [18]

B. Random Forest

Random forest is one of the machine learning algorithms that are easy to apply and has a low computational load but still have high accuracy [19]. The Random Forest algorithm uses the bagging method to improve estimation value, adding a random sub-setting stage before each tree formation [20]. Random forest uses multiple decision trees on different subsets of the dataset to improve the accuracy of its decisions, where each decision tree will aggregate the estimates from each tree and produce a final output based on the majority of the projection votes [21].

The decision tree in the random forest algorithm consists of root nodes, internal nodes, and leaf nodes, built by calculating the entropy value (shown in equation (2)) as a determinant for the attribute's impurity and the information gain value (shown in equation (3)).

$$Entropy\ (Y) = - \sum_i p\ (c|y) * log_2\, p\ (c|y) \quad (2)$$

$$Inf.Gain\ (Y, A) = Entropy\ (Y) - \sum_{v \in A} \frac{|Y_v|}{|Y_A|} Entropy\ (Y_v) \quad (3)$$

The decision-making process using the random forest algorithm has the following steps [22]:

1. Randomly select K data points from the training set.
2. Build a decision tree model corresponding to the selected data points known as subsets.
3. Select the number N for the decision tree to be built.
4. Repeat the process by randomly selecting K data points and building a decision tree with the selected subsets.
5. Based on the higher votes for the new data points, it finds the predictions for each decision tree and assigns new data points to each section.

C. Dataset

The data used in this study comes from the Kaggle.com website, which contains 1601 data [23]. The author of this data collects it from a survey involving parents with kids and tabulates the results as a dataset with 13 feature categories. Table I shows the configuration of the dataset's features based on the survey questions.

We first normalize the data in the dataset by changing its values to numerical (Table II), with the sample result shown in Table III.

TABLE I
CONFIGURATION OF DATASET FEATURES

Features	Question
Olympiad Participation	Has your child participated in any Science/Maths Olympiad?
Scholarship	Has he/she received any scholarship?
School	Love's going to school?
Favorite Subject	What is his/her favorite subject?
Projects	Has done any projects under academics before?
Grasping Power	His/Her Grasping power (1-6)
Time Sport	How much time does he/she spend playing outdoor/indoor games?
Medals	Medals won in Sports?
Career Sport	Want's to pursue his/her career in sports?
Acting Sport	Regular in his/her sports activities?
Fantasy Arts	Love creating fantasy paintings?
Won Arts	Won art competitions?
Time Art	Time utilized in Arts?

TABLE II
NORMALIZING RULE

Features	**Old Values**	**New Values**
Olympiad Participation	Yes, No	1, 0
Scholarship	Yes, No	1, 0
School	Yes, No	1, 0
Favorite Subject	Mathematics, Science, History/Geography, Any Language	1, 2, 3, 4
Projects	Yes, No	1, 0
Grasping Power	1 to 6	1 to 6
Time Sport	1 to 6	1 to 6
Medals	Yes, No	1, 0
Career Sport	Yes, No	1, 0
Acting Sport	Yes, No	1, 0
Fantasy Arts	Yes, No	1, 0
Won Arts	Yes, Maybe, No	1, 2, 0
Time Art	1 to 6	1 to 6

TABLE III
SAMPLE OF NORMALIZING RESULT

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**

1	1	1	1	1	5	1	1	0	0	0	2	3	AC
1	1	1	1	1	3	2	0	0	0	0	0	1	AC
1	1	1	2	1	5	1	1	0	0	0	0	1	AC
1	1	1	1	1	5	1	1	0	0	0	2	3	AC
1	1	1	2	1	5	3	0	0	0	0	0	2	AC
0	0	1	4	1	3	1	0	0	1	1	1	4	AR
0	0	1	4	0	4	1	1	1	0	1	1	1	AR
0	1	0	2	1	4	3	1	1	0	1	2	5	AR
0	0	1	1	0	3	2	1	0	1	1	0	3	AR
0	0	1	4	1	4	3	0	0	1	1	2	4	AR
1	0	1	2	1	3	4	1	1	1	0	0	3	SP
0	0	1	1	0	5	4	0	1	1	0	0	3	SP
1	0	1	1	1	5	5	1	1	1	0	0	1	SP
0	0	1	1	0	3	6	1	1	1	0	0	4	SP
0	0	1	4	1	3	4	1	1	1	0	0	2	SP

Notes: A = Olympiad Participation, B = Scholarship, C = School, D= Favorite Subject, E = Projects, F= Grasping Power, G = Time Sport, H= Medals, I= Career Sport, J = Acting Sport, K= Fantasy Arts, L = Won Arts, M= Jasmine, N = Time Art, AC = Academic, AR = Art, SP = Sport

D. Classification Model

We use different models with various data training and testing ratios to classify the children's talent using the C4.5 and random forest algorithms, with Table IV showing the models' configuration.

TABLE IV
CONFIGURATION OF THE CLASSIFICATION MODEL

Model	Algorithm	Training Data Ratio	Testing Data Ratio
C4.5 (50)	C4.5	50%	50%
RF (50)	Random Forest	50%	50%
C4.5 (60)	C4.5	60%	40%
RF (60)	Random Forest	60%	40%
C4.5 (70)	C4.5	70%	30%
RF (70)	Random Forest	70%	30%
C4.5 (80)	C4.5	80%	20%
RF (80)	Random Forest	80%	20%

E. Evaluation

To evaluate the classification results of each model, we use 10-fold and 20-fold cross-validation with the value of accuracy, f-score, precision, and recall as a comparison. We calculate these values using the true positive, false positive, false negatives, and true negatives, using equations (4) to (7) [24]:

$$Accuracy = \frac{True\ Positive + True\ Negatives}{Actual + Predicted} \quad (4)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (5)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negatives} \quad (6)$$

$$F - Score = \frac{Precision\ .\ Recall}{Precision + Recall} \quad (7)$$

The values from equations (4) to (7) of each model are the final evaluation to see which model performs the best, where the highest values mean the best model.

III. RESULTS

From the dataset used in this research, we split the data according to the configuration in Table IV, with Table V showing the splitting result.

TABLE V
RESULTS OF TRAINING AND TESTING DATA SPLITTING

Model	Algorithm	Training Data	Testing Data
C4.5 (50)	C4.5	801	800
RF (50)	Random Forest	801	800
C4.5 (60)	C4.5	961	640
RF (60)	Random Forest	961	640
C4.5 (70)	C4.5	1121	480
RF (70)	Random Forest	1121	480
C4.5 (80)	C4.5	1281	320
RF (80)	Random Forest	1281	320

Each model performs classification on the dataset, using the 10-fold and 20-fold evaluation, according to the data splitting in Table V, resulting in the confusion matrix value shown in Table VI (for the training evaluation) and Table VII (for the testing evaluation).

TABLE VI
CONFUSION MATRIX OF TRAINING DATA

Model	Actual	10-Fold Prediction			20-Fold Prediction		
		A1	A2	S	A1	A2	S
C4.5 (50)	A1	309	17	24	312	15	23
	A2	14	181	10	14	184	7
	S	32	6	208	30	5	211
RF (50)	A1	316	13	21	316	12	22
	A2	16	179	10	20	180	5
	S	18	4	224	26	3	217
C4.5 (60)	A1	369	26	25	369	26	25
	A2	25	214	7	22	215	9
	S	28	6	261	31	6	258
RF (60)	A1	383	16	21	385	15	20
	A2	20	217	9	19	218	9
	S	16	9	270	19	5	271
C4.5 (70)	A1	445	18	26	443	21	25
	A2	24	254	9	22	258	7
	S	39	12	294	42	8	295
RF (70)	A1	447	16	26	454	11	24
	A2	25	253	9	21	257	9
	S	20	7	318	22	8	315

C4.5 (80)	A1	509	28	22	509	27	23
	A2	29	293	6	27	291	10
	S	37	8	349	35	8	351
RF (80)	A1	516	15	28	519	13	27
	A2	24	295	28	27	291	10
	S	29	6	359	26	8	360

Notes: A1 = Academic Talent, A2 = Art Talent, S = Sport Talent

TABLE VII
CONFUSION MATRIX OF TESTING DATA

Model	**Actual**	**Testing Prediction**		
		A1	**A2**	**S**
C4.5 (50)	A1	317	18	14
	A2	14	185	6
	S	31	3	212
RF (50)	A1	324	9	16
	A2	11	191	3
	S	13	6	227
C4.5 (60)	A1	253	12	14
	A2	7	152	5
	S	24	2	171
RF (60)	A1	253	14	12
	A2	6	153	5
	S	9	4	184
C4.5 (70)	A1	194	13	3
	A2	5	115	3
	S	11	3	133
RF (70)	A1	195	7	8
	A2	5	117	1
	S	9	6	132
C4.5 (80)	A1	131	7	2
	A2	3	76	3
	S	10	3	85
RF (80)	A1	130	7	3
	A2	3	77	2
	S	6	3	89

From the result shown in Table VI, we use equations (4) to (7) to calculate the evaluation values (accuracy, f-score, precision, and recall) of each model for the training data, as shown in Table VIII.

TABLE VIII
EVALUATION VALUES FOR TRAINING DATA

Model	**Accuracy (%)**		**F-Score (%)**		**Precision (%)**		**Recall (%)**	
	10F	**20F**	**10F**	**20F**	**10F**	**20F**	**10F**	**20F**
C4.5 (50)	87.1	88.3	87.1	88.3	87.1	88.3	87.1	88.3
RF (50)	89.8	89	89.8	89	89.8	89.1	89.8	89
C4.5 (60)	87.8	87.6	87.8	87.6	87.8	87.6	87.8	87.6
RF (60)	90.5	90.9	90.5	90.9	90.5	91	90.5	90.9
C4.5 (70)	88.6	88.8	88.6	88.8	88.6	88.8	88.6	88.8
RF (70)	90.8	91.5	90.8	91.5	90.8	91.5	90.8	91.5
C4.5 (80)	89.9	89.9	89.9	89.9	89.9	89.9	89.9	89.9
RF (80)	91.3	91.3	91.3	91.3	91.4	91.4	91.3	91.3

Notes: 10F = 10-Fold Cross Validation Evaluation, 20F = 20-Fold Cross Validation Evaluation

From the result shown in Table VII, we use equations (4) to (7) to calculate the evaluation values (accuracy, f-score, precision, and recall) of each model for the testing data, as shown in Table IX.

TABLE IX
EVALUATION VALUES FOR TESTING DATA

Model	**Accuracy (%)**	**F-Score (%)**	**Precision (%)**	**Recall (%)**
C4.5 (50)	89.2	89.2	89.3	89.2
RF (50)	92.7	92.8	92.8	92.7
C4.5 (60)	90	90	90	90
RF (60)	92.2	92.2	92.3	92.2
C4.5 (70)	92.1	92.1	92.2	92.1
RF (70)	92.5	92.5	92.6	92.5
C4.5 (80)	91.2	91.2	91.4	91.2
RF (80)	92.5	92.5	92.6	92.5

From the results in Table VIII above, we observed that the highest accuracy, f-score, precision, and recall values for training the data are 91.5%; all these values came from the RF (70) model using the 20-fold cross-validation. We also observed that the lowest accuracy, f-score, precision, and recall values for training the data are 81.7%; all these values came from the C4.5 (50) model using the 20-fold cross-validation. From these results, we conclude that the RF (70) model using 20-fold cross-validation is the best, and the C4.5 (50) model using the 10-fold cross-validation is the worst model for training the data.

From the results in Table IX above, we observed that the highest accuracy, f-score, precision, and recall values for testing the data are 92.7%, 92.8%, 92.8%, and 92.7%; all these values came from the RF (50) model. We also observed that the lowest accuracy, f-score, precision, and recall values for testing the data are 89.2%, 89.2%, 89.3%, and 89.2%; all these values came from the C4.5 (50) model. From these results, we conclude that the RF (50) model is the best, and the C4.5 (50) model is the worst model for testing the data.

IV. CONCLUSIONS

The results of this study show that a child's hobbies and activities can be used as reference material to predict the child's talent in academics, arts, or sports. The prediction model employing the C4.5 algorithm and random forest demonstrated excellent performance based on the dataset gathered from the survey responses regarding a child's interests and activities. Results from 10-fold and 20-fold cross-validation tests on the training set demonstrate that, on average, the random forest algorithm outperforms C4.5. The tabulated evaluation results for the training data (10-fold and 20-fold cross-validation) and the test data prove this statement. The evaluation results also show that the ratio of training and test data affects the performance of the model's prediction results. The experiment results using different data training and data testing ratio shows that the model performs better when the training data ratio is larger than the test data

REFERENCES

[1] R. J. Sternberg and D. Ambrose, Eds., *Conceptions of Giftedness and Talent*. Cham: Springer International Publishing, 2021.

[2] C. Yohana, I. Agung, N. S. Perdana, and S. Silisabon, "A study of factors influencing the development of student talent," *Int. J. Educ. Pract.*, vol. 8, no. 3, pp. 441–456, 2020, doi: 10.18488/journal.61.2020.83.441.456.

[3] M. Gonzalez, J. Capman, F. Oswald, E. Theys, and D. Tomczak, "'Where's the I-O?' Artificial Intelligence and Machine Learning in Talent Management Systems," *Pers. Assess. Decis.*, vol. 5, no. 3, pp. 33–44, 2019, doi: 10.25035/pad.2019.03.005.

[4] J. Han, "Classroom Behaviour of Talent Cultivation in Colleges and Universities Supported by Deep Learning Technology," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/6579995.

[5] S. Shi, M. Chen, R. Feng, H. Zhang, and S. Zhang, "Prediction of Academic Talent Capacity Based on Gradient Boosting Decision Tree," *Appl. Comput. Math.*, vol. 8, no. 4, p. 75, 2019, doi: 10.11648/j.acm.20190804.12.

[6] Yuhelmi, Taslim, Syamsidar, and Machdalena, "Prediction of Academic Achievement of Part-Time Working Students Using Artificial Neural Network," *J. Inovtekpolbeng- Seri Inform.*, vol. 7, no. 1, pp. 61–70, 2022, [Online]. Available: http://ejournal.polbeng.ac.id/index.php/ISI/article/view/2368.

[7] S. Jauhiainen, S. Aÿrämö, H. Forsman, and J. P. Kauppi, "Talent identification in soccer using a one-class support vector machine," *Int. J. Comput. Sci. Sport*, vol. 18, no. 3, pp. 125–136, 2019, doi: 10.2478/ijcss-2019-0021.

[8] N. H. Jasni, A. Mustapha, S. S. Tenah, S. A. Mostafa, and N. Razali, "Prediction of player position for talent identification in association netball: a regression-based approach," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 1, pp. 84–96, 2022, doi: 10.26555/ijain.v8i1.707.

[9] M. Jamil, A. Phatak, S. Mehta, M. Beato, D. Memmert, and M. Connor, "Using multiple machine learning algorithms to classify elite and sub-elite goalkeepers in professional men's football," *Sci. Rep.*, vol. 11, no. 1, pp. 1–7, 2021, doi: 10.1038/s41598-021-01187-5.

[10] A. Çetinkaya and Ö. K. Baykan, "Prediction of middle school students' programming talent using artificial neural networks," *Eng. Sci. Technol. an Int. J.*, vol. 23, no. 6, pp. 1301–1307, Dec. 2020, doi: 10.1016/j.jestch.2020.07.005.

[11] M. Siener, I. Faber, and A. Hohmann, "Prognostic Validity of Statistical Prediction Methods Used for Talent Identification in Youth Tennis Players Based on Motor Abilities," *Appl. Sci.*, vol. 11, no. 15, p. 7051, Jul. 2021, doi: 10.3390/app11157051.

[12] T. Purwoningsih, H. B. Santoso, K. A. Puspitasari, and Z. A. Hasibuan, "Early Prediction of Students' Academic Achievement : Categorical Data from Fully Online Learning on Machine-Learning Classification Algorithms," *J. Hunan Univ. Sci.*, vol. 48, no. 9, pp. 131–141, 2021, [Online]. Available: http://jonuns.com/index.php/journal/article/view/713/710.

[13] D. Noviana, Y. Susanti, and I. Susanto, "Analysis Of Scholarship Recipient Recommendations Using The K-Nearest Neighbor (K-Nn) Algorithm And C4.5 Algorithm," 2019. [Online]. Available: http://jurnal.umt.ac.id/index.php/cpu/article/view/1685/1078.

[14] D. W. Triscowati, W. P. Buana, and A. H. Marsuhandi, "Mapping Corn Land Potential Using Satellite Imagery and Random Forest on Cloud computing Google Earth Engine," *Semin. Nas. Off. Stat.*, vol. 2021, no. 1, pp. 1001–1011, Nov. 2021, doi: 10.34123/semnasoffstat.v2021i1.889.

[15] P. Alkhairi and Z. Situmorang, "Application of Data Mining to Analyze Employee Satisfaction with HR Services with the C4.5 Algorithm," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 7, no. 1, p. 40, Feb. 2022, doi: 10.30645/jurasik.v7i1.414.

[16] J. Kusuma, A. Jinan, and Z. Situmorang, "Application of Decision Tree Algorithm C4.5 in Determining Shiploading License," *MEANS (Media Inf. Anal. dan Sist.*, vol. 7, no. 1, pp. 72–76, May 2022, doi: 10.54367/means.v7i1.1632.

[17] P. P. P. A. N. W. F. I. R.H. Zer, M. Wahyuni, A. Rangga, and Z. Situmorang, "Model Analysis of Student Satisfaction with Lecturer Teaching Methods Using the C4.5 Algorithm," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 1, p. 58, Feb. 2022, doi: 10.26798/jiko.v6i1.520.

[18] K. Dalimunthe and Z. Situmorang, "Study of C45 Algorithm In Predicting New Employee Acception," *Int. J. Econ. Technol. Soc. Sci.*, vol. 2, no. 2, pp. 518–524, 2021.

[19] I. Firmansyah, J. T. Samudra, D. Pardede, and Z. Situmorang, "Comparison Of Random Forest And Logistic Regression In The Classification Of Covid-19 Sufferers Based On Symptoms," *J. Sci. Soc. Res.*, vol. 5, no. 3, p. 595, Oct. 2022, doi: 10.54314/jssr.v5i3.994.

[20] M. Azhari, Z. Situmorang, and R. Rosnelly, "Comparison of Classification Accuracy, Recall, and Precision on C4.5, Random Forest, SVM, and Naive Bayes Algorithms," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.

[21] A. S. Wincy Pon Annal, R. Manonmani, and C. Booma, "Emotion Recognition from Speech using SVM and Random Forest Classifier," *J. Soft Comput. Paradig.*, vol. 4, no. 1, pp. 39–47, 2022, doi: 10.36548/jscp.2022.1.005.

[22] R. B. Bhardwaj and S. R. Chaurasia, "Use of ANN , C4.5 and Random Forest Algorithm in the Evaluation of Seismic Soil Liquefaction," *J. Soft Comput. Civ. Eng.*, vol. 6, no. 2, pp. 92–106, 2022.

[23] A. Rampurawala, "Kids Hobby Prediction Dataset," 2020. [Online]. Available: https://www.kaggle.com/datasets/abtabm/hobby-prediction-basic.

[24] D. Pardede, I. Firmansyah, M. Handayani, M. Riandini, and R. Rosnelly, "Comparison Of Multilayer Perceptron's Activation And Op-Timization Functions In Classification Of Covid-19 Patients," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 8, no. 3, pp. 271–278, Aug. 2022, doi: 10.33330/jurteksi.v8i3.1482.