Trend of Stunting Weight for Infants and Toddlers Using Decision Tree

Andi Nugroho, Harco Leslie Hendric Spits Warnars, Ford Lumban Gaol, Tokuro Matsuo

Abstract— **This papar is to show the trend of increasing the weight of infants and toddlers is of particular concern to the Indonesian government. This encourages the government to avoid stunting and malnutrition that can affect babies aged 0-59 months. The Indonesian government quickly established the Integrated Health Service Post (POSYANDU) to reduce the number of malnutrition in infants and toddlers. With the existence of POSYANDU, it is hoped that infants and toddlers can avoid malnutrition. The purpose of this study was to see the trend of weight gain for infants and toddlers aged 0-59 months to ensure that the weight of infants and toddlers continues to increase from month to month. To be able to see the trend of weight gain, a decision tree classification algorithm is needed, to see the weight gain of infants and toddlers. In addition, the feature selection process uses the single factor ANOVA method in determining the feature approach possessed by the dataset. This study will also compare the decision tree algorithm with the KNN in determining the classification of weight gain for infants and toddlers. Then the results show that the decision tree algorithm and ANOVA are better at providing accuracy, recall, F1 score, and precision than the KNN classification algorithm. Thus the decision tree + ANOVA algorithm is a classification algorithm used to see the trend of weight gain in infants and toddlers aged 0-59 months.**

Index Terms—**Infants, Children, Classification, POSYANDU**

I. Introduction

Health is a right for all Indonesian people, every person has the right to live in physical and spiritual prosperity, to live, and to live a good and healthy life. environment and the right to health services [1]. Health is the government's main concentration to be able to provide health to all Indonesian people. This health care has been carried out by the government by establishing public and private hospitals, puskesmas, and to a lesser extent, there is Integrated Health Service Post (POSYANDU). Posyandu was established on June 29, 1983, to provide health to the community from infants to the elderly [2]. Along with the times, POSYANDU has become a measure of equal distribution of health for babies aged 0-59 months, pregnant women, nursing mothers, and the elderly. Infants are aged 0 - 59 months are given special attention so that one day they become healthy and intelligent children as the next generation of the Indonesian nation.

POSYANDU will provide vaccines, pay attention to the composition of 4 healthy 5 perfect foods, weigh infants and toddlers aged 0 - 59 months, then ensure that children's nutrition is adequate. In this case looking at bodyweight from month to month until the child is 59 months old. This weight monitoring will be reported to the community health center (Puskesmas) [4]. PUSKESMAS will provide reports from the POSYANDU cadres regarding the health conditions of infants and toddlers aged 0-59 months to PUSKESMAS [5]. The development of infants and toddlers is always monitored through a card provided by Posyandu in the form of a KMS (Health Indicator Card) for infants and toddlers [6].

II. Related Work

Machine learning (ML) is a method used to predict disease, forecast weather, search for the nearest distance, and much more [7]. This study will discuss one of the methods in ML, namely supervised data or commonly referred to as the classification method. Many classification methods have been developed including Support Vector Machine (SVM), Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), Neural network, and many more. The classification methods that will be discussed in this study are the Decision Tree and KNN. In addition, this study will also use the One-way Analysis of Variants (ANOVA) (Single Factor) method which will be used to select features. Feature selection is used to determine features in the dataset in order to increase the accuracy value of the classification algorithm.

A. ANOVA

ANOVA is a statistical technique that can be used to check whether the population or features in a dataset are related or we can fiind mean square and F test to decide the hipotesis [8],[9],[10]. ANOVA is also commonly referred to as the method used for confirmatory and exploratory data analysis. The advantage of using the ANOVA f test is that it can compare the average value of each feature and perform tests on the F Distribution table where the results are compared with the f-test if the hypothesis value of the f-test is greater than or equal to the F distribution, the features contained in The dataset has no relationship, but if it is smaller than the Distribution table F, then the feature has a relationship with other features [11]. ANOVA is also commonly used to find relationships between features in the

Manuscript received December, 2020; revised Sept, 2021. This work is supported by the Research and Technology Transfer Office, Bina Nusantara University as a part of Bina Nusantara University's International Research Grant contract number: No.026/VR.RTT/IV/2020 and contract date: April 6, 2020

Andi Nugroho is a Ph.d Student, Jakarta Indonesia Binus Graduate Program, West Jakarta 11530, Indonesia, (corresponding author to provide e-mail: andi.nugroho002@binus.ac.id)

Harco Leslie Hendric Spits Warnars is a Co Promotor of Doctor Computer Science, Jakarta Indonesia Binus Graduate Program, West Jakarta 11530, Indonesia, (e-mail: spits.hendric@binus.ac.id)

Ford Lumban Gaol is a Head of Doctor Computer Science, Jakarta Indonesia Binus Graduate Program, West Jakarta 11530, Indonesia, (e-mail: fgaol@binus.edu)

Tokuro Matsuo is a Professor Graduate School of Industrial Technology, Tokyo Japan Advanced Institute of Industrial Technology, higashiooi shinagawa-ku Tokyo 140-0011, Japan, (e-mail: matsuo@tokuro.net)

dataset, whether these features are related or not, here are the steps for using the ANOVA analysis method :

H0:Hipotesis

H0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_n$

H0: Not all feature datasets are related

H0 will be rejected if F-test >= F Distribution table

H0 will be accepted if F-test < F Distribution table

Where :

$\overline{X_j}$ = The average value of features

n_j = Amount of data in a feature

$\overline{X}$ = The average score of all features

x = Feature value

k = Amount of features

From the formula above, it can be described as follows :

Calculate the total mean value of each column or feature in the datasets

$$SS_A = \sum_{n_j}(X_j - X)^2 \quad (1)$$

Calculate the average value for each column or feature and then add the average values to get the total average features

$$SS_R = \sum\sum(\mathrm{x} - X_j)^2 \quad (2)$$

Calculates the total value between columns and within columns

$$SS_T = SS_A + SS_R \quad (3)$$

$$\mathrm{k} - 1 \quad (4)$$

$$\mathrm{N} - \mathrm{k} \quad (5)$$

$$\mathrm{N} - 1 \quad (6)$$

Computes the estimated variance of the sum of squares and the Degree of freedom on the value SS_A

$$MSA = \frac{SS_A}{k-1} \quad (7)$$

Computes the estimated variance of the sum of squares and the Degree of freedom on the value SS_R

$$MSR = \frac{SS_R}{\mathrm{N-k}} \quad (8)$$

The F-Test is the quotient of the MSA and MSR values and the F-Test value will be compared with the F distribution value

$$F = \frac{MSA}{MSR} \quad (9)$$

B. Decision Tree

Since the 1950s, artificial intelligence has been discovered, and in the same year machine learning has become the center of world research, a decision tree algorithm has been developed [13]. The decision tree was introduced by Quinlan from the New South Wales Institute of technology. Decision tree is built for knowledge-based systems that can be used for supervised data. The structure of the decision tree algorithm resembles a tree where at the top it is called the root, the branches are the internal nodes, the leaves are the classes in the dataset[14],[15].

N = The total of entire datasets

TABLE 1. FORMULA ANOVA [12]

Source	Sum of Squares	Degree of Freedom	Mean Square	F test
Between Columns (A)	$SS_A = \sum_{n_j}(X_j - X)^2$	k=1	$MSA = \frac{SS_A}{k-1}$	$F = \frac{MSA}{MSR}$
Within Columns (Residual)	$SS_R = \sum\sum(\mathrm{x} - X_j)^2$	N – k	$MSR = \frac{SS_R}{\mathrm{N-k}}$	
Total	$SS_T = SS_A + SS_R$	N – 1		

TABLE 2. DATA SET STUNTING MALNUTISION FROM JANUARI-DES 2019 [13]

NO	Attributes (*in Person*)			
	S	K	D	BGM (Stunting)
1	114	111	86	65
2	78	78	61	46
3	111	111	77	51
4	103	103	68	38
5	74	70	38	14
6	146	146	93	72
7	134	134	96	54
8	118	118	82	58
9	32	32	15	12
10	93	93	71	48
11	110	110	52	41
12	120	120	72	51
13	114	111	86	65
14	78	78	61	46

In the above case, it is known that In the table above it is known that the data is in the form of the total number of stunting infants and toddlers where the value of **S** is the data for infants and toddlers who are registered at the Posyandu, **K** is the data for infants and toddlers who have registered cards at the posyandu, **D** is the data for infants and toddlers who come to the Posyandu. to be weighed and **BGM** is data that is experiencing malnutrition. A decision tree is a classification algorithm designed to determine the class and better accuracy of features that have been determined through the top, branch, and leaf processes [16][17]. A decision tree is a classification algorithm that can be developed into other models such as combining a decision tree with an optimization method. Because the decision tree It is an algorithm that has a high degree of accuracy and interpretability [18].

$$Entropy\ (S) = -\ \sum_{i=1}^{m} p_i \log_2 p_i \quad (10)$$

Where S is the entropy value of the addition of P_i and $\log_2$.

$$Gain(S,A) = Entropy(S) - \sum_{v\ \in Values(A)} \frac{|S_v|}{|S|}\ Entropy\ (S_v) \quad (11)$$

The value of A is the value obtained from the attribute and Sv is the subset of S and v is the value of the subset S that is in attribute A.

$$Gain\ Ratio\ (S,A) = \frac{Gain\ (S,A)}{Split\ Information(S,A)} \quad (12)$$

C. KNN

KNN is a classification algorithm for finding common ground by relying on its closest neighbors to provide information on members of the set or not[19]. Then the KNN algorithm will provide information that the value of x is included in class two [24]. The KNN classification algorithm is included in the top-ten data mining algorithm [25]. The following is the formula for the KNN algorithm, in the following example, the variable gj is the position of the data and X is a sample that is sought or test data that will be tested on the KNN algorithm. Then for a variable i is the iteration of the number of datasets [24][26].

$$gj^{(X)} = \min gi^{(X)} \quad i = 1, 2, \ldots, C \quad (13)$$

In the above formula, it is stated that the rule of the nearest neighbor has two aspects, the first is convergence and the other is generalization error. For the variable value x and the nearest neighbor x' formula $P(e|x, x')$ is the nearest neighbor condition in an error state.

$$P(e|x) = \int P(e|x, x')p(x'|x)dx' \quad (14)$$

Then for $p(x'|x)$ Is a probability density function condition. The variable s is the x-centered hypersphere s position:

$$P_s = \int p(x')dx' \quad (15)$$

From Figure 1, it is known that the methodology has been made from the process of inputting data on infants and toddlers aged 0 - 59 months, after which preprocessing is carried out to obtain patterns in the dataset. Then the next step is to normalize the data to equalize the lower and upper limit values in the dataset using one of the min-max normalization methods. The next step is a dataset that has been normalized in its feature selection using the ANOVA method, in determining the features of the ANOVA method, the hypothesis value will be obtained from the comparison of the F-Test with the F Distribution to get the value of whether F-Test> = F Distribution if yes then the relationship is H0, H1, ..., Hn have nothing to do with but on the contrary, then determination of neighbors in the KNN algorithm is represented by the variable k. The accuracy value of the KNN algorithm affects the K value which is initialized at the start of the KNN algorithm in execution [20][21][22]. The value of k is determined based on training data and test data, to get good accuracy results it is necessary to experiment with giving a value of 3 - 10 to make predictions of the accuracy value. Several studies have been developed on how to increase the value of k using a combination of tree classification techniques, some have done a combination with fuzzy [23]. The following is an overview of the KNN classification algorithm.

$$P(error) = 1 - \sum P(c|x)P(c|x') \quad (16)$$

The following is the assumption that each sample stands independently where the variable C^* is the Bayesian classifier to be used in the variable P (error) :

$$c^* = \arg\max c_{c\epsilon Y}\ P^2(c^*|x) \quad (17)$$

$$P\ (error) \leq 1 - \sum_{c\epsilon Y} P^2\ (c^*|x) \quad (18)$$

III. METHODOLOGY

The methodology used in this research is the stage to get the best accuracy value from the KNN classification algorithm and decision tree with ANOVA. To get the best accuracy value, it must be obtained by weighing babies and toddlers aged 0 - 59 months. Besides, it is also influenced by the amount of data that has been inputted by posyandu cadres. However, from the data obtained, it is known that not all data registered with posyandu cadres attend posyandu events, so the calculation of stunting data in infants and toddlers is influenced by the presence of infants and toddlers. the relationship H0, H1, ..., Hn has nothing to do. From these results obtained feature selection which will be used to be reprocessed using classification algorithms such as decision tree and KNN. However, if the existing dataset is directly processed by a classification algorithm, the results of the accuracy value will be seen. From these Two steps, we will check which value of accuracy is the best between using an ANOVA analysis method or not using the ANOVA analysis method. In this research, an experiment will be carried out using the KNN classification algorithm and the Decision tree using ANOVA with the KNN algorithm and the Decision tree without ANOVA. Which of these experiments yields the best accuracy value.

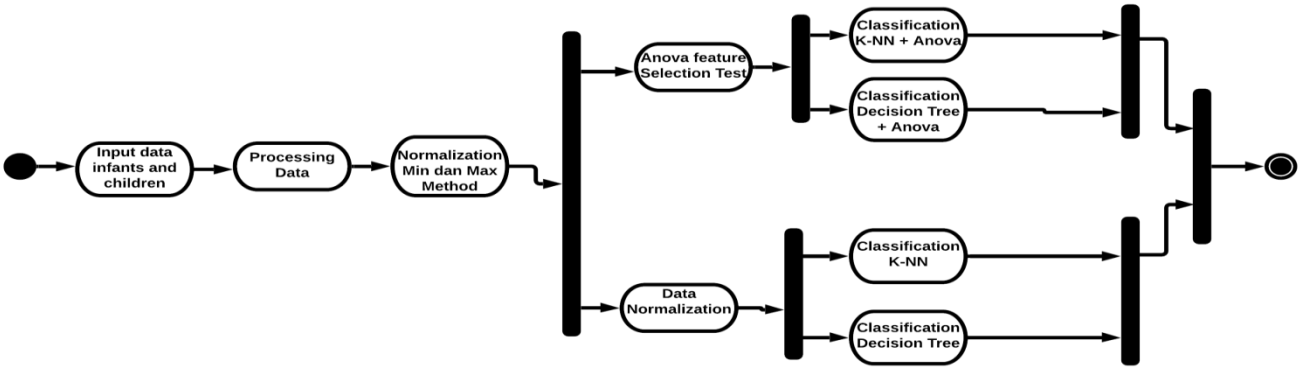

Fig. 1 Research Methodology ANOVA, KNN, and Decision tree

IV. Result

The first step taken was inputting weight data for infants and toddlers by posyandu cadres. The results of the data input process are processed first by making a dataset in the form of columns and records. The following is an example of data input obtained from the posyandu cadre data input process.

A. Infants and toddlers weight input data

The data inputted by posyandu cadres is the weight of infants and toddlers aged 0-59 months, the input process is carried out when the mother of the infant and toddler comes to the posyandu with the baby or toddler and then weighs his body by the posyandu cadre. The data obtained is preprocessed to produce new data in the form of table data containing attributes of registered infant and toddler data, data of infants and toddlers who have just registered, data of infants and toddlers whose weight has increased, and data of infants and toddlers. Who lost weight. And in the end, 5 features will be processed using an ANOVA analysis method.

B. ANOVA analysis Method

The ANOVA analysis method is used to check whether the 5 attributes have a relationship or not, then the 5 attributes are processed and get the results as shown in Figure 5. The results obtained are the results of several experiments. Because from the results of the experiment, it turns out that the relationship between features was not found so that in the end the decreasing features of infants and toddlers were removed and changed with the presentation value for the presence of infants and toddlers by looking for the results of the distribution of the data for infants and toddlers who were presently divided by the total infants and toddlers. Registered and the percentage is on average above 50% and below 50%. With this new feature, the maximum accuracy result is obtained.

TABLE 3. Result ANOVA Feature Selection

Source of variation	SS	df	MS	F	P-value	F crit
Between Groups	1.743	4	0.435	7.352	7.65E-06	2.380
Within Groups	64.314	1085	0.059			
Total	66.058	1089				

From the results of the ANOVA feature selection, as it is known, when compared with F-crit or F-test, the value is 2.380 and for the F-distribution table for df = 4, the F-distribution is 2.371, based on the hypothesis of ANOVA H0> = H1 so that the results here are not There is a relationship between the features that were made, but the value of F-test and F-distribution is not too much difference and this is what makes the accuracy value increase.

C. Classification algorithm

The next step is to enter the feature selection generated from the ANOVA analysis method into the KNN classification algorithm and decision tree. And here is a plotter from the dataset of ANOVA feature selection results as shown in Figure 3.

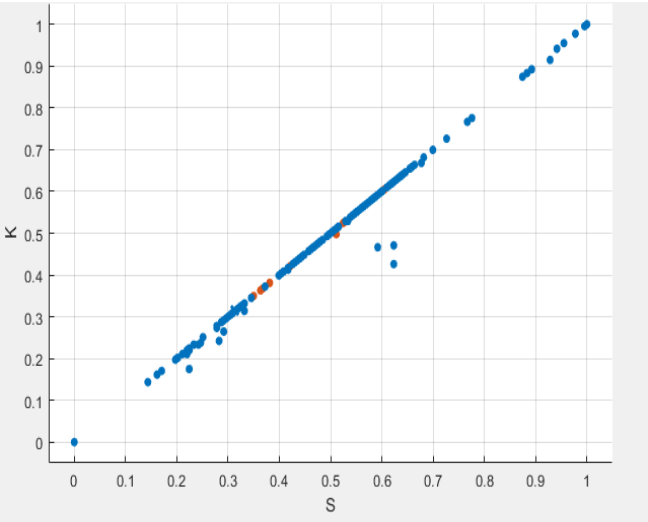

Fig. 3 Infants and Toddler Weight Dataset aged 0-59 months

From this dataset, the results of the KNN algorithm and the Decision Tree process with or without ANOVA are obtained, for more details, see table 4.

TABLE 4. Comparison between KNN+ANOVA and Decision Tree+ANOVA and without ANOVA

No	Dataset weight baby and children age 0 – 59 months				
	Algorithm	Accuracy	Precision	Recall	F1
1	KNN	0.61	0.44	0.30	0.36
2	DECISION TREE	0.57	0.35	0.26	0.29
3	KNN + ANOVA	0.87	1	0.5	0.66
4	**DECISION TREE + ANOVA**	0.96	0.97	0.89	0.93

In table 4 it can be seen that the decision tree + ANOVA algorithm has the best level of accuracy compared to the KNN + ANOVA algorithm or without ANOVA. Thus, it can be seen that the weight gain of infants and toddlers aged 0-59 months can be seen from the arrival of mothers from infants and toddlers to the posyandu. This is a concentration to increase interest in the arrival of mothers who have babies and toddlers. The decision tree algorithm will then increase the accuracy value by using the ANOVA selection feature. Because before the decision tree algorithm is classified, the ANOVA method has a very significant difference in accuracy values without using ANOVA. It can be seen that the value with the decision tree algorithm + ANOVA is 0.96 compared to the KNN + ANOVA which is only 0.87 from the accuracy value.

V. Summary

From the results above, it can be seen that the use of the decision tree + ANOVA algorithm can increase the value of accuracy because the feature selection process is carried out by the ANOVA method. Then for datasets that are linear regression, it would be better to use a decision tree classification algorithm and the feature selection method should be used. As future research, a combination of the decision tree algorithm with the Dragonfly Algorithm can be carried out, the decision tree with the PSO, and other optimization algorithms.

ACKNOWLEDGMENT

The authors thank the university's binus lab for borrowing lab space and computers.

References

[1] R. Dewi and R. Anisa, "*The Influence of Posyandu Cadres Credibility on Community Participation in Health Program*," *J. Messenger*, vol. 10, no. 1, p. 83, 2018.

[2] I. Idris, I. Tandi, P. M. Malen, A. Saade, S. Syamsuddin, and M. I. Aryawiguna, "*The Innovation System on Health Public Service: A Case Study of Posyandu in Makassar, Indonesia*," *Int. J. Bus. Manag.*, vol. 11, no. 4, p. 102, 2016.

[3] O. Soedirham, "*Integrated Services Post (Posyandu) as Sociocultural Approach for Primary Health Care Issue*," *Kesmas Natl. Public Heal. J.*, vol. 7, no. 5, p. 195, 2012.

[4] S. M. Rahmawati, S. Madanijah, F. Anwar, and R. Kolopaking, "*The effectiveness education of counseling of infant and young child feeding as intensive to improve counseling performance of Posyandu cadres in Bogor, Indonesia*," *Int. J. Community Med. Public Heal.*, vol. 6, no. 6, p. 2280, 2019.

[5] A. Nugroho, S. Rahayu, and R. M. Manikam, "*E-POSYANDU for POSYANDU cadres report*," *Int. J. Open Inf. Technol.*, vol. 8, no. 10, pp. 77–83, 2020.

[6] B. Lareno, L. Swastina, and H. M. Junaidi, "*IT Application to Mapping the Potential of Malnutrition Problems*," *Proceeding - 2018 Int. Conf. ICT Smart Soc. Innov. Towar. Smart Soc. Soc. 5.0, ICISS 2018*, pp. 1–4, 2018.

[7] U. Narayanan, A. Unnikrishnan, V. Paul, and S. Joseph, "*A survey on various supervised classification algorithms*," *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput. ICECDS 2017*, pp. 2118–2124, 2018.

[8] G. Hesamian, "*One-way ANOVA based on interval information*," *Int. J. Syst. Sci.*, vol. 47, no. 11, pp. 2682–2690, 2016.

[9] H. K. Edwards, "*System utilization and changes in implemented information systems: A case study*," *Lect. Notes Eng. Comput. Sci.*, no. February, pp. 833–838, 2006.

[10] J. L. Tseng, "*Raising the Learning Effects for Learners with Low Entrance Scores using Project-Based Learning in Virtual Reality Practice*," *IAENG International Journal of Computer Science*, vol. 47, no. 3, pp. 516–521, 2020.

[11] T. M. Chinh Chu, H. J. Zepernick, and V. Sundstedt, "*Analysis of Variance of Opinion Scores for MPEG-4 Scalable and Advanced Video Coding*," *2018, 12th Int. Conf. Signal Process. Commun. Syst. ICSPCS 2018 - Proc.*, 2019.

[12] A. Danila, D. Ungureanu, S. A. Moraru, and N. Voicescu, "*An implementation of the variance analysis (ANOVA) for the power factor optimization at distribution level in smart grid*," *Proc. - 2017 Int. Conf. Optim. Electr. Electron. Equipment, OPTIM 2017 2017 Intl Aegean Conf. Electr. Mach. Power Electron. ACEMP 2017*, no. 1, pp. 48–53, 2017.

[13] J. R. Quinlan, "*Induction Fo Decion Trees*," *Mach. Learn.*, vol. 1, pp. 81–106, 1986.

[14] S. A. Bashir, D. C. Doolan, and A. Petrovski, "*The effect of window length on accuracy of smartphone-based activity recognition*," *IAENG International Journal of Computer Science*, vol. 43, no. 1, pp. 126–136, 2016.

[15] M. M. Nwe and K. T. Lynn, "*Effective resampling approach for skewed distribution on imbalanced data set*," *IAENG International Journal of Computer Science*, vol. 47, no. 2, pp. 234–249, 2020.

[16] V. Vadim, F. Marina, M. Oleg, and A. Sergey, "*Temporal Decision Trees in Diagnostics Systems*," *2018 Int. Conf. Adv. Big Data, Comput. Data Commun. Syst. icABCD 2018*, pp. 1–10, 2018.

[17] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy, "*A Decision Tree Algorithm Combined with Linear Regression for Data Classification*," *2018 Int. Conf. Comput. Control. Electr. Electron. Eng. ICCCEEE 2018*, pp. 1–5, 2018.

[18] S. Ruggieri, "*Complete search for feature selection in decision trees*," *J. Mach. Learn. Res.*, vol. 20, pp. 1–34, 2019.

[19] Y. Li, Y. Yang, J. Che, and L. Zhang, "*Predicting the number of nearest neighbor for kNN classifier*," *IAENG International Journal of Computer Science*, vol. 46, no. 4, pp. 662–669, 2019.

[20] B. Priambodo, A. Ahmad, and R. A. Kadir, "*Prediction of average speed based on relationships between neighbouring roads using K-NN and neural network*," *Int. J. online Biomed. Eng.*, vol. 16, no. 1, pp. 18–33, 2020.

[21] K. U. Syaliman, E. B. Nababan, and O. S. Sitompul, "*Improving the accuracy of k-nearest neighbor using local mean based and distance weight*," *J. Phys. Conf. Ser.*, vol. 978, no. 1, 2018.

[22] T. M. Cover and P. E. Hart, "*Nearest Neighbor Pattern Classification*," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[23] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "*Efficient kNN classification with different numbers of nearest neighbors*," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, 2018.

[24] W. Xing and Y. Bei, "*Medical Health Big Data Classification Based on KNN Classification Algorithm*," *IEEE Access*, vol. 8, pp. 28808–28819, 2020.

[25] S. Zhang, "*Cost-sensitive KNN classification*," *Neurocomputing*, vol. 391, no. xxxx, pp. 234–242, 2020.

[26] D. Wettschereck and T. G. Dietterich, "*An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms*," *Mach. Learn.*, vol. 19, no. 1, pp. 5–27, 1995.