Employing Best Input SVR Robust Lost Function with Nature-Inspired Metaheuristics in Wind Speed Energy Forecasting

Rezzy Eko Caraka, *Member IAENG*, Rung Ching Chen[*], *Member IAENG*, Sakhinah Abu Bakar, Muhammad Tahmid, Toni Toharudin, Bens Pardamean, Su-Wen Huang[*]

***Abstract*—Wind power has been experiencing a quick improvement. Without a doubt, wind is a variable asset that is hard to forecast. For instance, traditionally time series, extra holds are distributed to deal with this uncertainty. This paper presents a comparison of the performance of various Support Vector Regression (SVR) applied to short-term wind power forecasting. The analogy with BORUTA and multivariate adaptive regression splines (MARS) as judge best input and employ genetic algorithm and particle swarm optimization to find best parameter in Support Vector Regression with robust lost function. We measure the accuracy of this models by Symmetric means absolute percentage error (sMAPE) and we get the best model BORUTA-SVR-PSO with sMAPE 2.07155%. Moreover, we measure the energy conversion using Feedback Linearization Control (FLC).**

***Index Terms*—Wind Speed, Feature Importance, Boruta, SVR, Genetic Algorithm, Particle Swarm Optimization**

I. Introduction

Energy is a thing that can not be separated by human life; almost all activities require energy as a driver of growth. Electricity demand is a huge problem significant in the era of the Industrial Revolution 4.0 and Society 5.0. This problem has quite a large number of linkages business[1], industry activities [2], sustainable development[3]. The government needs to take attention to find other energy resources[4]. One of them adds alternative sources like the wind turbine. Before the government determines a potential location for wind turbine development, research needs to be done to see whether the place is indeed feasible or not.

Therefore, Utilization of renewable energy becomes fascinating because the energy is quite cheap and will never run out. Renewable energy, one of which is to utilize wind energy available wherever and whenever. Research on wind turbines must be investigated first before being applied to real-life, so that wind turbines have high efficiency.

Wind energy as a source of power generation is exciting to study. Wind energy is related to the increasing energy needs and demands for energy production that is environmentally friendly and clean. Archer's [5] conducted a study to look at the global energy potential at an altitude of 80m. Wind speed at this height is calculated by using Least Square extrapolation technique from observation and sounding station data. The results showed that the average wind speed of 80m globally was 4.59 m/s while the 10-meter wind was 3.31 m/s. The average wind speed at the offshore station at an altitude of 80 meters has a speed of around 90% higher than on land in general. Wind speed at an altitude of 80 meters globally during the day is slightly higher than at night

The government has proclaimed the policy of conservation and diversification of energy is the right plan to be implemented in Indonesia [4]. Therefore the development of new renewable energy as a complementary energy fossil-based, it is absolute to continue to be implemented. To support this policy, essential things that need to be continually pursued are consistency, commitment, and new mindset from the government and all people that energy sources are not just coal [6], or gas only[7], but also water[8], waves[9], wind [10], sun, geothermal,biofuels, nuclear and others[11]. Various studies have been conducted on renewable energy resources in Indonesia, especially those sourced from wind energy. Mahmudin[12] conducted a study of the potential of wind energy in Sulawesi and Maluku.

The obtained power density is used to draw monthly wind power maps for the sea areas around the islands. Talking about the potential for strong winds in North Sulawesi Province [13] , Because of not only the location of Indonesian

Manuscript received September 31, 2019; Revised June, 20 2020. This research is supported in part by the Ministry of Science and Technology, Taiwan. The Nos are MOST-107-2221-E-324 -018 -MY2 and MOST-106-2218-E-324-002, Taiwan and GUP 2017-118 Universiti Kebangsaan Malaysia and Bina Nusantara University.

R.E.Caraka. College of Informatics, Chaoyang University of Technology, Taichung City, 41349, Taiwan (R.O.C.), and School of Mathematical Sciences, The National University of Malaysia, Bandar Baru Bangi, Selangor, Malaysia, 43600, and Bioinformatics and Data Science Research Center Bina Nusantara University, Indonesia 11480. (e-mail:rezzyekocaraka@gmail.com)

*R.C. Chen. College of Informatics, Chaoyang University of Technology, Taichung City, 41349, Taiwan (R.O.C.); (e-mail: crching@cyut.edu.tw)

S.A.Bakar, School of Mathematical Sciences, The National University of Malaysia, Bandar Baru Bangi, Selangor, Malaysia, 43600.

M.Tahmid. Indonesian Agency for Meteorological, Climatological and Geophysics (BMKG), Manado, North Sulawesi, 91594, Indonesia

T.Toharudin. Department of Statistics, Padjadjaran University, Jatinagor, west java, Indonesia, 45363.

B.Pardamean. Bioinformatics and Data Science Research Center Bina Nusantara University, and Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia 11480

*Su-Wen Huang. Taichung Veterans General Hospital, Taiwan (ROC). (e-mail: dale33663366@vghtc.gov.tw)

region surrounding the equator was a Hadley, walker, and local circulation meeting but also tropical depressive and tropical cyclone activity in the Pacific Ocean east of the Philippines [14] **,** [15]**,** [16] in every year caused the potential for strong winds in parts of North Sulawesi. Classification of wind speeds can be divided into several categories which including calm, gentle breeze, moderate, strong, and gale wind. For information on each group will be explained as follows: First, Wind calm means a wind speed of less than 0.5 m/s (1 kt). Second, Gentle breeze category of wind speed ranges from a maximum wind speed of 5.5 m/s, this category is included in the type of weak wind. In this category, all types of aircraft are safe and are permitted to conduct flight activities [17]. Third, The category of moderate wind speed ranges from a maximum wind speed of 10.7 m/s; this category is included in the moderate and fresh wind. In this category, the types of small aircraft are not obtained to do flight activities[15]. The types of aircraft are not allowed to do flight activities are the types of Cessna, Beechcraft, Grumman, and others. And forth, Wind speed strong category ranges from a maximum wind speed of 20.7 m/s; this category is included in the type of strong wind. In the category, the aircrafts can allow to carry out flight activities are only broad aircraft types such as Airbus A350, Airbus A340, Airbus A380, Boeing 747, Boeing 767, and Boeing 777. 5. The category of gale wind speed ranges from 20.8 m/s and above, this category is included. In this category, all types of aircraft are prohibited from conducting flight activities, both landing and flight because they are very dangerous.

Research on forecasting is very widely used [18], starting from parametric and also non-parametric [19] difference motivation is about the accuracy generated [20] and also what information researcher wants to get from the model [21]. However, for SVR parameter search widely using grid search [22], genetic algorithms [23], particle swarm optimization [24], firefly [25], other metaheuristic techniques [26]. Also, SVR can be used by a single kernel [27] or multi-kernel [28].

In general, genetic algorithms (GA) are inspired by Darwin's theory of evolution [29]. GA is a search algorithm that is based on natural system mechanisms, namely genetics and natural selection. In a GA application [30], the solution variable is coded into a string structure that represents a sequence of genes, which is characteristic of the problem solution. In contrast to conventional searching techniques, GA departs from the set of solutions produced randomly. Chromosomes evolve in a continuous iteration process called generation. Each generation, chromosomes are evaluated based on an evaluation function. After several generations, GA will converge on the best chromosome, which is expected to be the optimal solution. The Main point in GA is the function of evaluation, selection, and genetic operators [31].

Particle Swarm Optimization (PSO) [32], [33] is used to determine the optimum global value [34] and has been widely used in the field of artificial intelligence in recent years. PSO has adaptive properties, global optimization, parallel algorithms, and problem-solving abilities [35].

The remainder of the paper is organized as follows. Section II explains research methods; Section III presents the analysis. Finally, conclusions and future research directions are indicated in Section IV.

II. RESEARCH METHOD

A. Feature Selection

Feature selection is the process of identifying and removing features that are irrelevant and excessive. Features are considered relevant if their values vary systematically with category membership.

This process is essential in the heart of machine learning because many machine learning algorithms experience decreases inaccuracy when the number of variables is significant but not optimal [36], [37]. Other than that, the number of variables slows down algorithm performance and takes too many resources. In this paper, we only use Boruta and multivariate adaptive regression splines (MARS).

Boruta is one of new feature selection [38] and judges the information from all of the variable in our data set. Judge and elimination are based on distribution importance to the best shadow in Boruta. Also, Boruta similar to Random Forest.

The steps of Boruta's algorithm in finding all the relevant variables are as follows [39]:

1. Add shadow attributes, i.e., random data from copies of all attributes.
2. Run a random forest classifier and get *Z* scores.
3. Get the maximum *Z* scores among the shadow attributes (MZSA).
4. Mark the original attributes that have a lower importance than MZSA as 'Rejected' and delete it from the system.
5. Mark the original attributes that have higher importance than MZSA as 'Confirmed.'
6. Remove all shadow attributes.
7. Repeat the procedure until only the Confirmed attribute is left, or when reaching the predetermined iteration limit.

Problems in the real world, in general, are rarely separable linear, most are non-linear. Data mining and machine learning techniques are mostly developed with linear assumptions, so the resulting algorithm is limited for linear cases. This problem can be solved by transforming data into a higher dimension of space so that it can be separated linearly in the new feature space.

In spline regression analysis [40], if given high dimensional data that is the data with predictor variables (X_p) where $3 \leq p \leq 20$ with a sample size of $50 \leq n \leq 1000$ using the Multivariate Adaptive Regression Splines (MARS). MARS is one of the flexible methods for regression modeling with high-dimensional data.

There are third point to consider in using the MARS model. First, Knots are the end of a regression line and the beginning of another regression. Second, Base Function is a function used to explain the relationship between response variables and predictor variables. The number of basis functions (BF) is *2-4* times the predictor variable. Third, Interaction is the correlation between variables with the maximum number of interactions (MI) is *1, 2,* and *3.*

MARS modeling is determined based on trial and error for the combination of BF, MI, and MO to get the value of the minimum smoothing parameter [41]. MO is the minimum distance between knots or the minimum observation between knots (MO) of *0, 1, 2,* and *3,* respectively. Defined the response variable Y_1 moreover, the predictor variables

X_1, X_2, X_3 the MARS model estimator can be written in eq(1):

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km} . \left(x_{v_n(k,m)} - Z_{km} \right) \right] \tag{1}$$

With:
a_0 = base parent function
a_m = coefficient of the m-th basis of function
M = non constant function
K_m = degree of interaction to m
s_{km} = the value is 1 or -1 if the data is to the right or left the knot point
$x_{p_n(k,m)}$ = predictor variable of *p* with observation to *n*
Z_{km} = knot value of the predictor variable $x_{p_n(k,m)}$

So the MARS model is stated in the following eq(2):

$$y_i = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km} . \left(x_{p_n(k,m)} - Z_{km} \right) \right] + \varepsilon_i \tag{2}$$

Then, we can rewrite eq(2) in eq(3) and the complete function in eq(4) as follows:

$$\begin{aligned} \hat{f}(x) &= a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km} . \left(x_{p_n(k,m)} - Z_{km} \right) \right] \\ \hat{f}(x) &= a_0 + \sum_{m=1}^{M} a_m \left[s_{1m} . \left(x_{p_n(1,m)} - Z_{1m} \right) \right] \\ &\quad + \sum_{m=1}^{M} a_m \left[s_{1m} . \left(x_{p_n(1,m)} - Z_{1m} \right) \right] \left[s_{2m} . \left(x_{p_n(2,m)} - Z_{2m} \right) \right] \\ &\quad + \sum_{m=1}^{M} a_m \left[s_{1m} . \left(x_{p_n(1,m)} - Z_{1m} \right) \right] \left[s_{2m} . \left(x_{p_n(2,m)} - Z_{2m} \right) \right] \left[s_{3m} . \left(x_{p_n(3,m)} - Z_{3m} \right) \right] \\ &\quad + \cdots \end{aligned} \tag{3}$$

$$\hat{f}(x) = a_0 + f_i(x_i) + f_{ij}(x_i, x_j) + f_{ijk}(x_i, x_j, x_k) + \cdots \tag{4}$$

Determination of knots on MARS uses forward stepwise and backward stepwise steps. Model selection using forward stepwise is made to get the number of base functions. To fulfill the concept of parsimony (a simple model) backward stepwise is carried out by removing the base function that has a small contribution to the response of the forward stepwise that needs to be considered in selecting the best MARS model is if the generalized cross-validation (GCV) value of the model has the most value low (minimum) among other models in eq (5).

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^{N} \left[y_i - \hat{f}_M(x_i) \right]^2}{\left[1 - \frac{\hat{C}(M)}{N} \right]^2} \tag{5}$$

With y_i as the response variable, x_i is a predictor variable. N is the number of observations and $\hat{f}_M(x_i)$ is defined as the estimated value of the response variable on *M* base function at x_i. Also, *M* is a maximum number of base functions $\hat{C}(M) = C(M) + d.M$. Besides, C(M) is *Trace* $[B\,(B^T B)^{-1} B^T] +$. However, $\boldsymbol{B}$ is a matrix of *M* basis functions. Moreover, *d* is value when each base function reaches optimization *(2≤d ≤4).*

Base function (BF) used to explain the relationship between response variables and predictor variables. This basis function is a parametric function defined in each region. In general, the selected base function is polynomial with a continuous derivative at each knot point. Friedman suggests that the maximum number of base functions (BF) is 2-4 times the number of predictor variables. The maximum number of interactions (MI) used is *1,2*, or *3*. If MI> 3 will produce increasingly sophisticated models and the model will be challenging to interpret. The minimum distance between knots or minimum observation (MO) used is *0, 1, 2,* and *3.*

B. Robust Lost Function in SVR

Support vector regression, or generally known as SVR, is an extended version of the support vector machine [42] [43]. The different is, SVR only used for regression and time series. The concept of ε-insensitive loss in SVM functions can be generalized to SVR [44].

Moreover, SVM combines computational methods such as hyperplane, kernel, and others that can minimize their classification errors. Also, SVM as supervised learning. Initially, SVM was only used in categorical data with classes *+1* and *-1*.

SVR uses the concept of structural risk minimization, which is to estimate a function by minimizing the limits of generalization error so that it can overcome overfitting. The function in SVR can be expressed in eq(6) and the condition in eq(7) and eq(8) as follows:

$$R(f(x)) = \frac{c}{T} \sum_{t=1}^{T} L_t(y_t, f(x_t)) + \frac{1}{2} ||w||^2 \tag{6}$$

Where

$$L_t(y_t, f(x_t)) = \begin{cases} 0 & , y_t = f(y_t) \\ |y_t, f(x_t)| - \varepsilon & , others \end{cases} \tag{7}$$

With L_t

$$\mathcal{L}^{\varepsilon}(\xi) = \begin{cases} 0, & |\xi| \leq \varepsilon \\ \frac{1}{2\gamma} (|\xi| - \varepsilon)^2 & , \varepsilon \leq |\xi| \leq e_C \\ C(|\xi| - \varepsilon) - \frac{1}{2} \gamma C^2 & , e_C \leq |\xi| \end{cases} \tag{8}$$

$e_C = \varepsilon + \gamma C$ and ε are the insensitivity parameter, and γ and C control the tradeoff between regularization and losses. Three different regions yield to distinguish different types of noise [45]. First, $\varepsilon -$ insensitivity zone: ignores errors less thanε. Zone with a quadratic cost: uses the *2*-dimension norm of the errors, which is suitable for the Gaussian noise. Also,

the last. Zone with a linear cost: limits the effects of the outliers. Note that eq (9) is the Vapnik ε −insensitive cost function when γ is small enough. The parameter γ plays the role of numeric regulation in the quadratic problem. The primal problem corresponding to eq (8) is as follows:

$$minimize_{w,b,\xi,\xi^*}\ L_p = \frac{1}{2}||w||^2 + \frac{1}{2\gamma}\sum_{i\epsilon I_1}(\xi_i^2 + \xi_i^{*2}) + C\ i\epsilon I1\xi i + \xi i* - \gamma C2$$

$$s.t \quad -y_i + (\langle w \cdot x_i \rangle + b) \le \varepsilon + \xi_i \qquad -y_i - (\langle w \cdot x_i \rangle + b) \le \varepsilon + \xi_i^* \qquad 0 \le \xi_i, \xi_i^* \tag{9}$$

I_1 and I_2 Are the sets of samples for which the losses have quadratic and linear cost, respectively. The corresponding Lagrangian is then it is eq(10).

$$\begin{aligned} L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*) &= \frac{1}{2}||w||^2 + \frac{1}{2\gamma}\sum_{i\epsilon I_1}(\xi_i^2 + \xi_i^{*2}) + C\sum_{i\epsilon I_1}\left(\xi_i + \xi_i^* - \frac{\gamma C}{2}\right) \\ &+ \sum_{i=1}^{N}\alpha_i(w^T x_i + b - y_i - \xi_i - \varepsilon \\ &+ \sum_{i=1}^{N}\alpha_i^*(y_i - w^T x_i - b - \xi_i^* - \varepsilon \\ &- \beta_i\xi_i - \sum_{i=1}^{N}\beta_i^*\xi_i^* \end{aligned} \tag{10}$$

Then Karush–Kuhn–Tucker (KKT) condition represent in eq(11) and eq(12).

$$\frac{\delta L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*)}{\delta w} = w + \sum_{i=1}^{N}\alpha_i x_i - \sum_{i=1}^{N}\alpha_i^* x_i = 0, \tag{11}$$

and

$$\begin{aligned} &\frac{\delta L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*)}{\delta b} = \sum_{i=1}^{N}\alpha_i - \sum_{i=1}^{N}\alpha_i^* = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0 \\ &\frac{\delta L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*)}{\delta b} = \frac{1}{\gamma}\xi_i - \alpha_i - \beta_i = 0, \qquad i\epsilon I_1 \\ &\frac{\delta L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*)}{\delta b} = C - \alpha_i - \beta_i = 0, \qquad i\epsilon I_2 \\ &\frac{\delta L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*)}{\delta b} = \frac{1}{\gamma}\xi_i^* - \alpha_i^* - \beta_i^* = 0, \qquad i\epsilon I_1 \\ &\frac{\delta L_{pd}(w,b,\xi,\xi^*,\alpha,\alpha^*)}{\delta b} = C - \alpha_i^* - \beta_i^* = 0, \qquad i\epsilon I_2 \end{aligned} \tag{12}$$

Then, The KKT complementarity conditions imply that

$$\alpha_i(w^T x_i + b - y_i - \varepsilon - \xi_i) = 0$$

$$\alpha_i^*(y_i w^T x_i - b - y_i - \varepsilon - \xi_i^*) = 0$$

$$\beta_i\xi_i = 0\,, \beta_i^*\,\xi_i^* = 0$$

The result represents the solution for the vector of parameters w. Besides, βi is null in the intervals I_1, and I_2 since $\xi_i > 0$ in these intervals. From previous expressions, we obtain the following equality of Lagrange multipliers αi (equivalently for αi∗) in eq (13).

$$\alpha_i = \begin{cases} C, & \xi_i \ge \varepsilon + \gamma C \\ \frac{\xi_i}{\gamma}, & \varepsilon \le \xi_i \le \varepsilon + \gamma C \\ 0, & 0 \le \xi_i \le \varepsilon \end{cases} \tag{13}$$

Therefore, we can apply the lagrangian to obtain dual in eq(14) :

$$\begin{aligned} L_d(\alpha,\alpha^*) &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)x_i^T x_j \\ &+ \sum_{i=1}^{N}((\alpha_i^* - \alpha_i)y_i - (\alpha_i^* + \alpha_i)\varepsilon) \\ &- \sum_{i\epsilon I_1}\frac{\gamma}{2}(\alpha_i^2 + \alpha_i^{*2}) - \sum_{i\epsilon I_2}\frac{\gamma C^2}{2} \end{aligned} \tag{14}$$

At the same time, we have taken into account in the interval I_2; then we have α_i and$\alpha_i^* = C$. We may rewrite the last term by $\sum_{i\epsilon I_2}\frac{\gamma}{2}(\alpha_i^2 + \alpha_i^{*2})$. So the equation can be written in eq(15):

$$\begin{aligned} L_d(\alpha,\alpha^*) &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)x_i^T x_j \\ &+ \sum_{i=1}^{N}\left((\alpha_i^* - \alpha_i)y_i - (\alpha_i^* + \alpha_i)\varepsilon\right) - \frac{\gamma}{2}(\alpha_i^2 + \alpha_i^{*2})) \end{aligned} \tag{15}$$

III. ANALYSIS

A. Feature Importance

This research was conducted at the climate station north minahasa, North Sulawesi Province and dataset consist of monthly wind speed from 1980 until 2016. This study will analyze and found which years are influential for forecasting in the period 2016- 2017. Then in our dataset will be grouping consisting of Group 1 (1980-1985), Group 2 (1986-1990), Group 3 (1991-1995), Group 4 (1996-2000), Group 5 (2001-2005), Group 6 (2005-2010), and Group 7 (2011-2016).

In the pre-modeling stage, group 7 will be used as a Dependent variable (*Y*), and each group 1 through group 6 will be an independent variable. Then the feature selection will be selected using MARS, and Boruta. In AI and statistics, variable selection, otherwise called feature selection, is the way toward choosing a subset of essential highlights for use in model development.

Based on Figure 1, the heat map can be interpreted that the density of wind velocity distribution around a range of 2 m/s. The data for analysis is monthly data so that the label on the heat map is numbered 1 to 60 which means that, 1 describes every first month in the group classification, e.g., 1 (January 1980), 2 (February 1980), 13 (January 1981). Next, we find the best features using Boruta. In this analysis, we perform

Boruta 10 iterations in 0.174123 secs. Six attributes confirmed important: group1, group2, group3, group4, group5 and one more; No attributes deemed unimportant.

In Figure 2 maxRuns explains the maximal number of random forest runs. Moroever, we can consider increasing this parameter if tentative attributes are left. With default is 100. Then, we do trace and it refers to verbosity level. At the same time, 0 means no tracing and 1 means are reporting attribute decision as soon as it is cleared. 2 means all of 1 plus additionally indicating each iteration and the default is 0.

Moreover, *hold history* describe the full history of importance runs is stored if set to TRUE (Default). Gives a plot of Classifier run vs. importance when the *plotImpHistory* function is called. Then, we find the feature selection using MARS the coefficient can be seen in Table I. In this technique, selected 6 of 16 terms, and 3 of 6 predictors. With termination condition reached 21. Also, we get the importance: group4, group1, group3-unused, group6, group2-unused, group5-unused.

The number of terms at each degree of interaction: 15 (additive model). Based on Figure 3, it can be seen that the performance of this method — the accuracy obtained by GCV 3.125701, RSS 125.0801, GRSq 0.452959, and R^2 0.6226816, respectively.

TABLE I COEFFICIENTS BASED ON MARS

GROUP	COEFFICIENTS
(Intercept)	4.713016
h (group1-3.8)	4.829473
h (group1-4.4)	-5.036424
h (2.5-group4)	-2.503461
h (group6-2.4)	18.929536
h (group6-2.5)	-19.282929

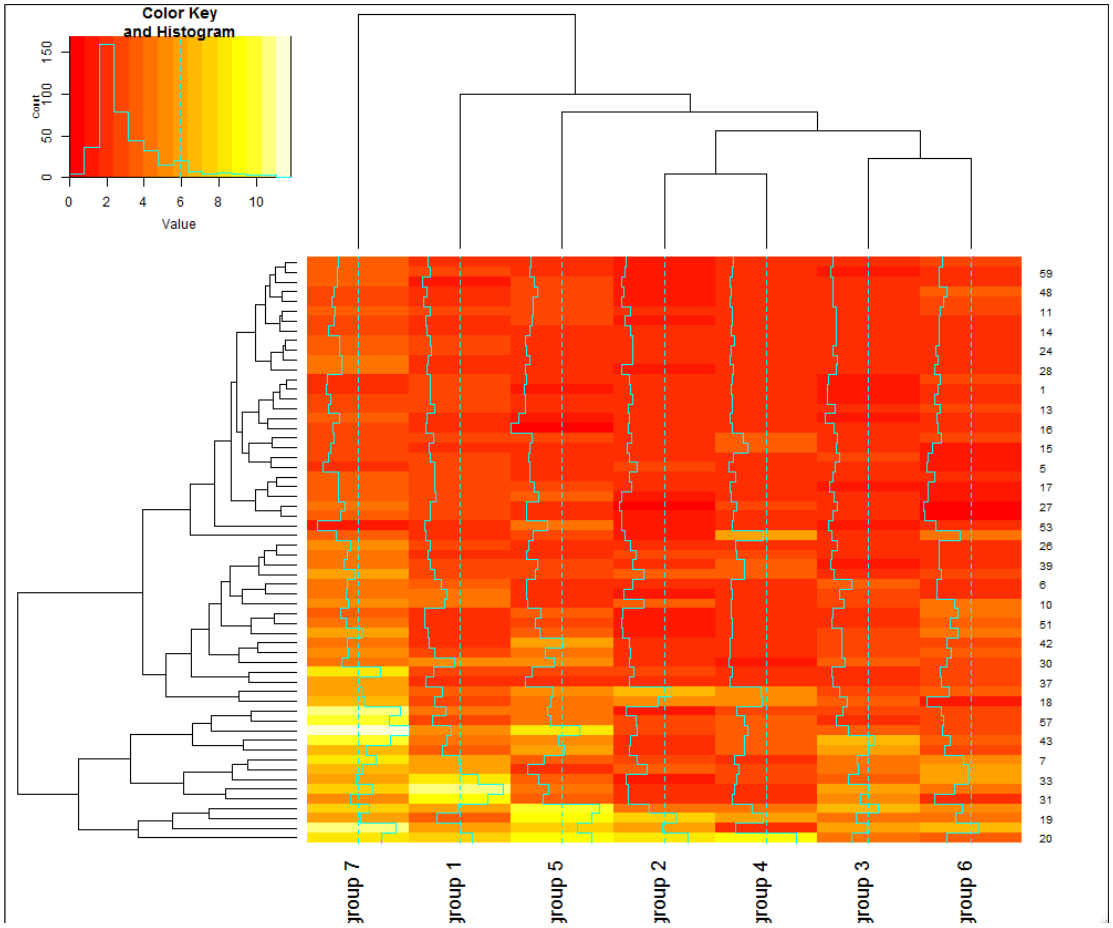

Figure 1. Heat Map

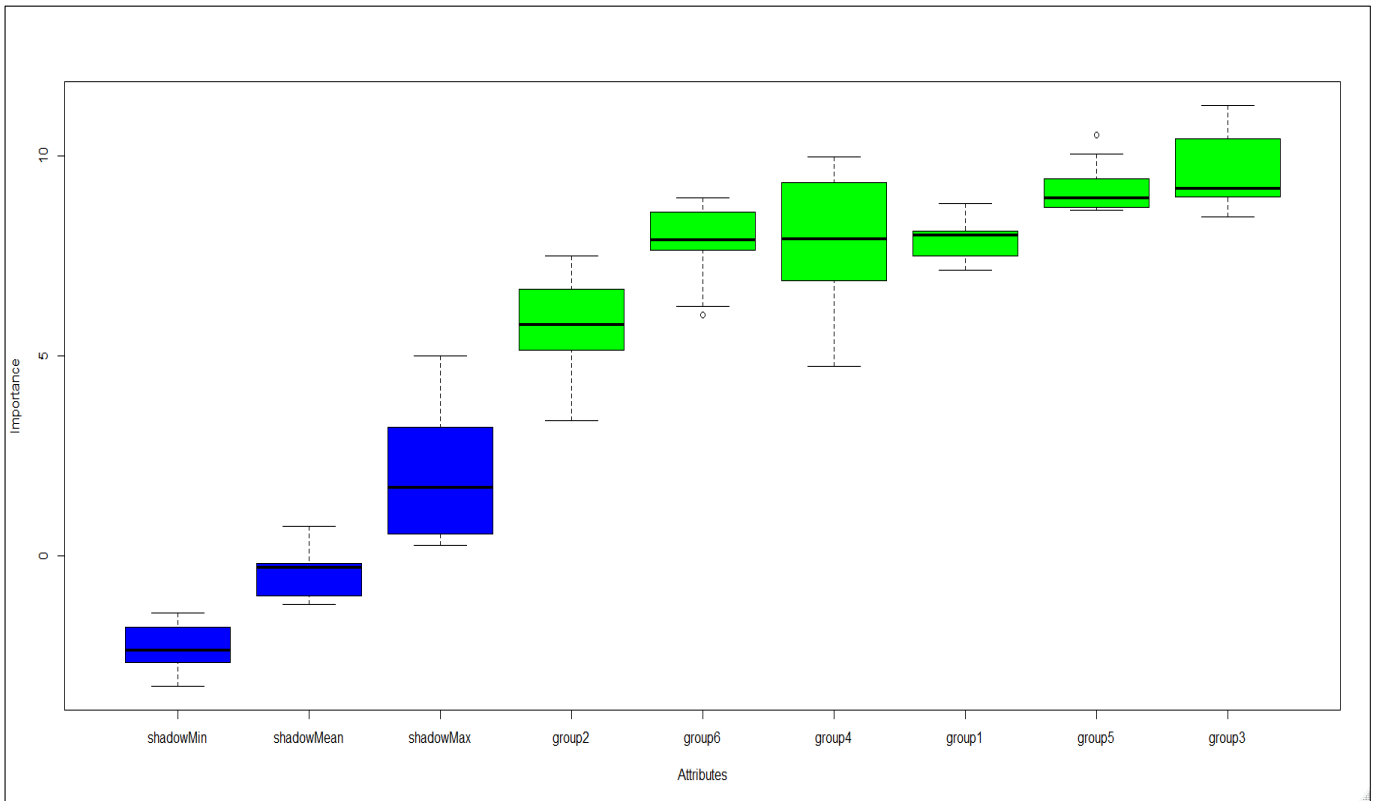

Figure 2. Boruta Feature Selectio

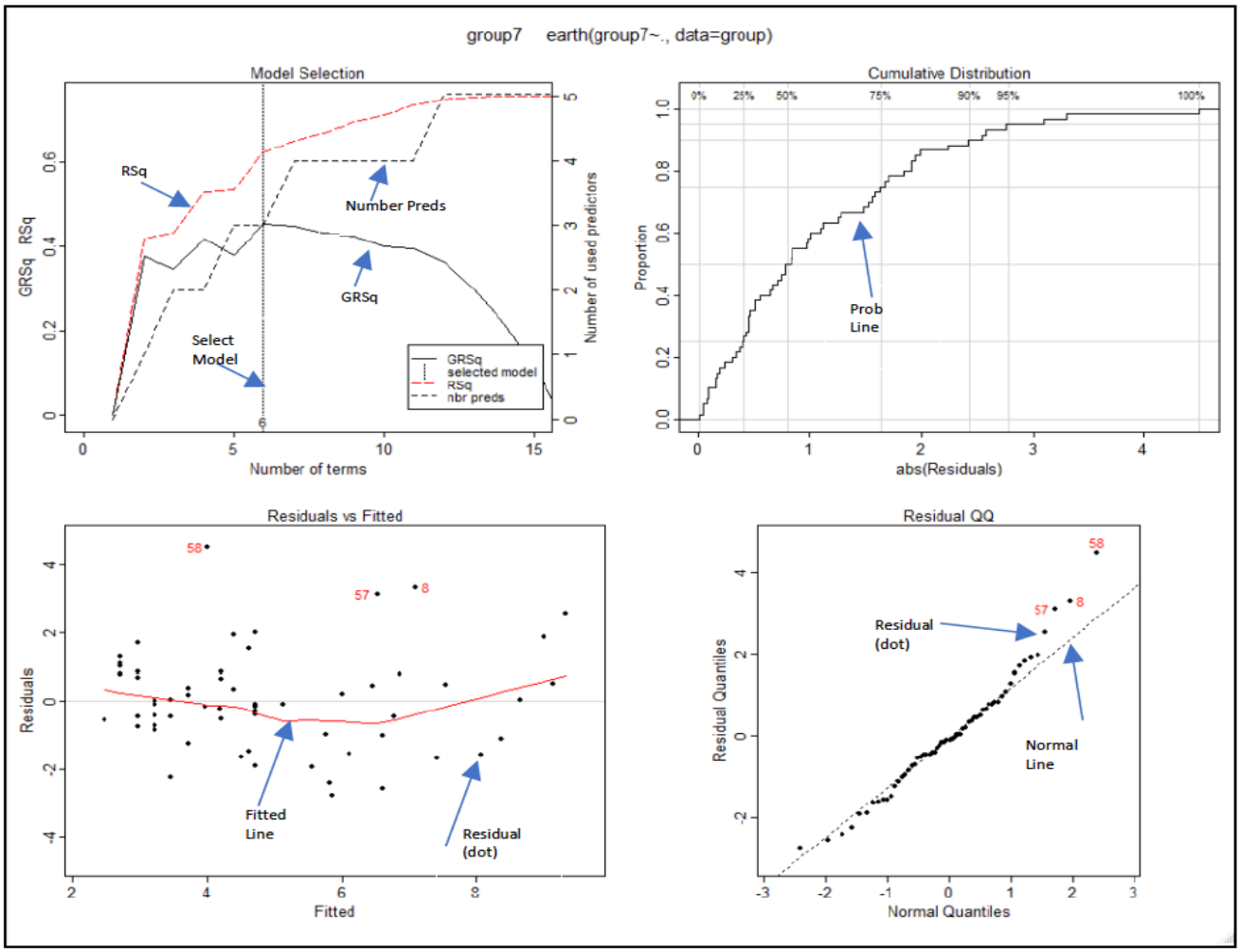

Figure 3. MARS

B. SVR Genetic Algorithm

In this paper, we optimize the SVR model parameters with the Gaussian kernel function (RBF) so that chromosomes are formed from 3x1 vectors. The vector will be filled with parameter values from SVR (cost, gamma, and epsilon). The value for chromosome formation can be filled from the results of previous SVR modeling as Good Initial Value.

A chromosome is good or not based on the value of the fitness function. The fitness function used is the value of SMAPE. Chromosomes can be said to be good chromosomes and can survive if they have a small SMAPE value. Of the 100 chromosomes that have been generated, the SMAPE value of each chromosome is calculated. Each chromosome will be selected to serve as a prospective parent.

Selection is made using a roulette wheel. Fitness values obtained in the previous stage are used to guide prospective parents. The frequency value is cumulative fitness divided by total fitness. This value will be the chosen limit of the chromosome. Chromosomes will be selected if the value of a random number lies in the range of frequency values of the previous chromosome and that chromosome.

Chromosome who has been chosen as a prospective parent is given a random uniform number (0.1). If the value of the number is less than the probability of a cross-over (P_c = 0.8), the chromosome is chosen as a parent, and a cross-over process occurs. Then, sort the fitness values of all the chromosomes. One chromosome with the highest fitness value will be stored and used as a parent in the next generation. Next is generating a new population of 100 chromosomes with the lowest fitness value.

Then iterate until the 10th iteration. The first phase will be carried out modeling with SVR GA with input from Boruta. In this study, we only set genetic algorithms with Population size = 100, Number of generations = 100, Elitism = 5, Crossover probability = 0.8, Mutation probability = 0.1, Search domain lower cost = 0.0001, lower gamma = 0.001, lower epsilon = 0.001, upper cost = 4, upper gamma = 1, and upper epsilon = 1.

Regarding the BORUTA-SVR-GA stage, the chromosome initialization will be carried out. In the modeling with BORUTA-SVR-GA-RBF obtained cost values: 2.094068, gamma: 1.811705, and epsilon: 1.994761. Figure 4 can be seen as a fitness simulation of this model.

Cost	Gamma	Epsilon
2.094068	1.811705	1.994761

At the same time, we will do a simulation also with MARS-SVR-GA, and we get the cost: 1.031778, gamma: 0.6733236, dan epsilon: 0.04152507. Furthermore, Figure 5 can be seen as a fitness simulation of this model.

Cost	Gamma	Epsilon
1.031778	0.6733236	0.04152507

C. SVR Particle Swarm Optimization

PSO algorithm adopts a large number of particles to search the optimum in the space solution, where the velocity of the particle depends on the direction and distance of its flight, and they have their own fitness values [46]. Optimization using PSO has five steps main thing to do. First, Defining the parameters used in PSO, the number of particles, the maximum iteration limit, the weight of inertia, the maximum speed limit, the component coefficient for each particle (C_1) and the component coefficient for all particles (C_2). The number of particles used is 100 particles with a maximum iteration limit of 100 times iteration.

The weight of inertia (w) is 0.9, and the maximum speed limit is 1. C_1 and C_2 are equal to 0.2. Second, initialize the position vector of each particle which can be illustrated in Figure 6. The dimensions of the PSO particle are determined by the number of parameters to be optimized.

In this research, the parameters to be optimized are the parameters of the SVR Model with Gaussian kernel (RBF) functions, which are cost, gamma, and epsilon. So the position vector is formed in size 3x1 because there are three parameters to be optimized. The parameter position will be generated randomly between the ranges of parameters used.

The range of parameters used successively for the cost, gamma, and epsilon parameters are {(10:1)}. Third, determine the fitness function. We justify the Fitness function that by SMAPE. Specifically, the 100 position vectors, the fitness value is calculated and then sorted from the smallest to the largest amount. Fourth, make the particle shift by using the following formula in eq(16).

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{16}$$

The vector position of a particle in eq (17) will be added to a vector speed and become a new position. The velocity vector is calculated using the following formula.

$$v_i(t) = v_i(t-1) + c_1 r_1\big(localbest(t) - x_i(t-1)\big) + c_2 r_2(localbest(t) - x_i(t-1)) \tag{17}$$

Where local best is the best fitness value of a particle and global best is the best fitness value of the whole particle. After the shift is made, the fitness value of the new particle position is then calculated. Finally, then steps 1 to 4 are repeated until it reaches the iteration limit or after all the particles produce convergent fitness. Figure 7 explains the fitness value to get the best SVR-PSO parameter with Boruta input. So we get the value of cost = 0.99, gamma = 0.45, and Epsilon = 0.0002.

Cost	Gamma	Epsilon
0.9998705	0.4568728	0.0002481071

Then, we find the SVR-PSO parameters with the best input from MARS. Based on Figure 8, it can be seen that the best and adjacent mean are different from the input using Boruta. Then we get the value of cost = 0.06, gamma = 0.31, and epsilon = 0.1

Cost	Gamma	Epsilon
0.06037308	0.3170344	0.1179467

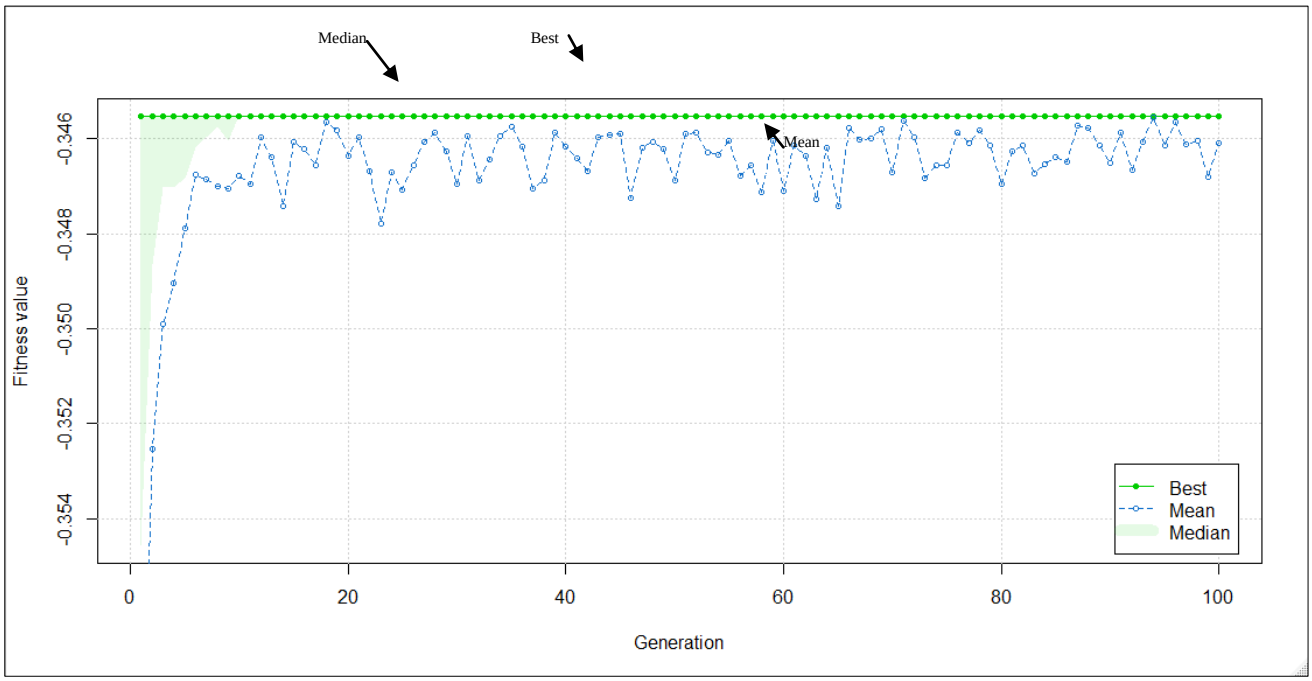

Figure 4. BORUTA-SVR-GA Fitness

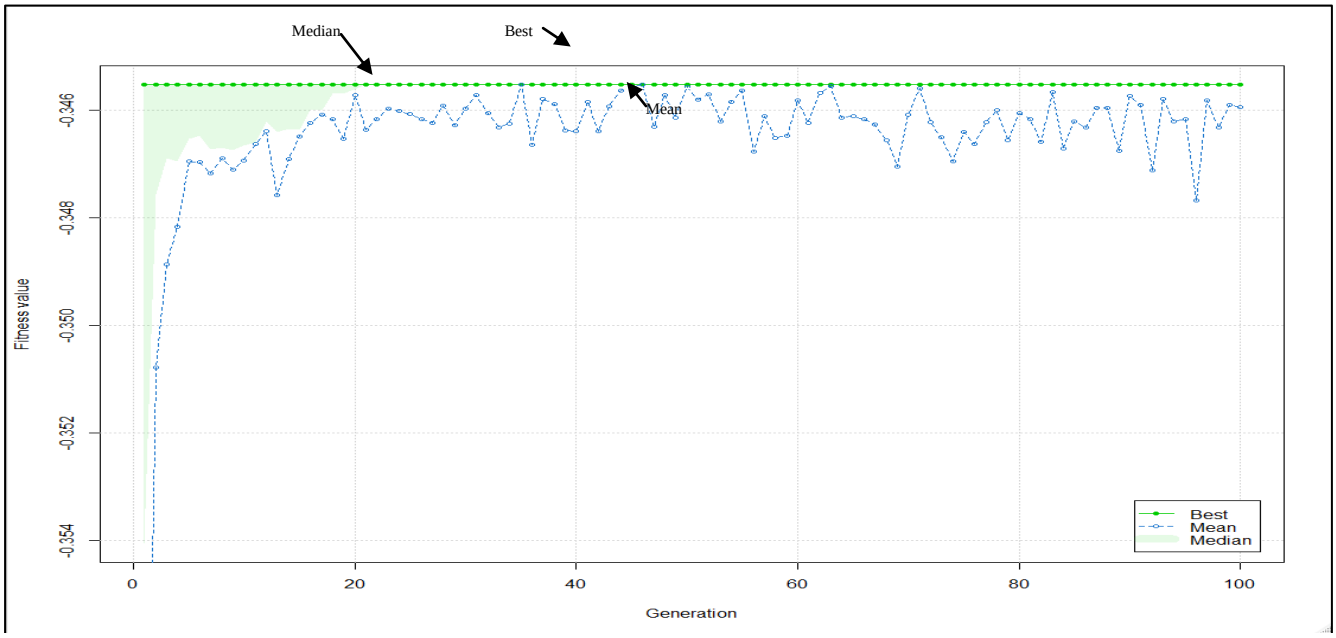

Figure 5. MARS-SVR-GA Fitness

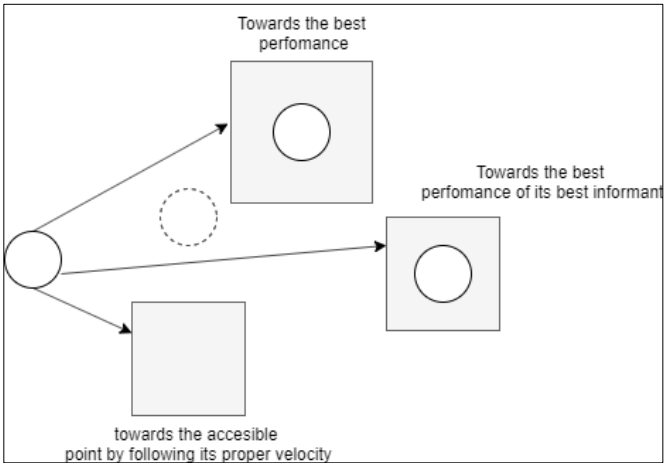

Figure 6. Concept of SVR-PSO

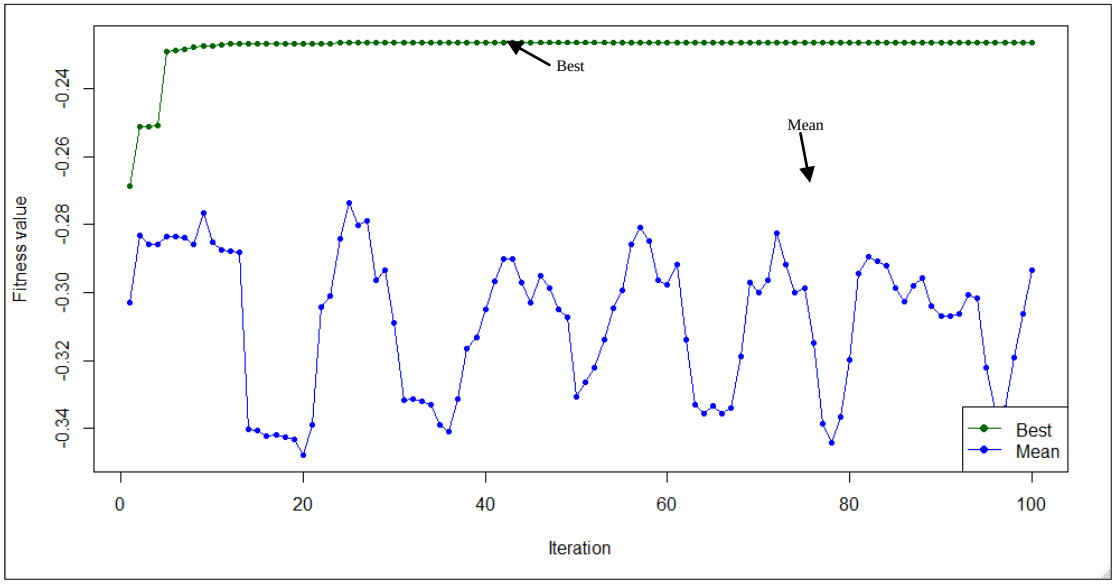

Figure 7. Boruta SVR-PSO Fitness

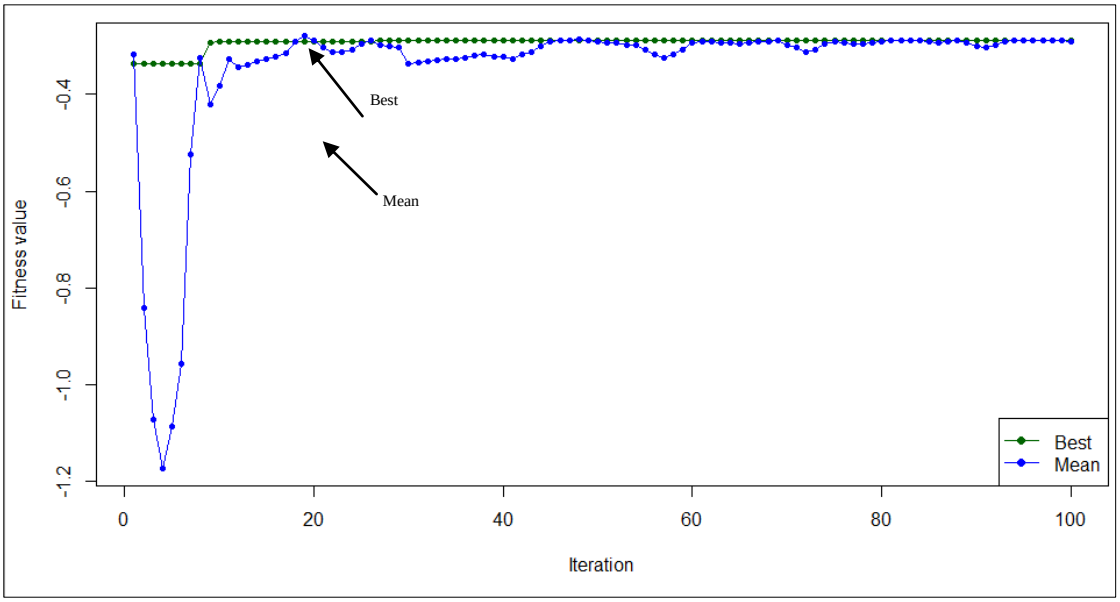

Figure 8. MARS SVR-PSO Fitness

D. Evaluation

To evaluation the model, the results can be evaluated by the value of Root Mean Square Error (RMSE) in eq (18) and symmetric Mean Absolute Percentage Error (sMAPE) in eq (19). The method with the smallest RMSE and sMAPE values is the best method of the other methods. RMSE and sMAPE values are calculated from out-sample data compared to the forecast results for each method.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{M}\sum_{l=1}^{M} e_l^2} \tag{18}$$

$$SMAPE = \frac{100\%}{T}\sum_{t=1}^{T}\frac{|Y_t - \hat{Y}_t}{(|Y_t - \hat{Y}_t|)/2} \tag{19}$$

The accuracy table of the simulation can be seen in Table II as below:

TABLE II ACCURACY

	SMAPE	RMSE
BORUTA-SVR-PSO	2.07155%	0.84678
BORUTA-SVR-GA	2.235%	0.9790
MARS-SVR-PSO	2.3913%	1.1257
MARS-SVR-GA	2.5211%	1.382145

Then, from each model obtained will be forecasted and compared with actual data to see whether the resulting model is accurate. Based on Figure 9 and 10, respectively it is clear that the model with Boruta input has a reasonably good performance compared to MARS. Moreover, PSO is also better than genetic algorithms.

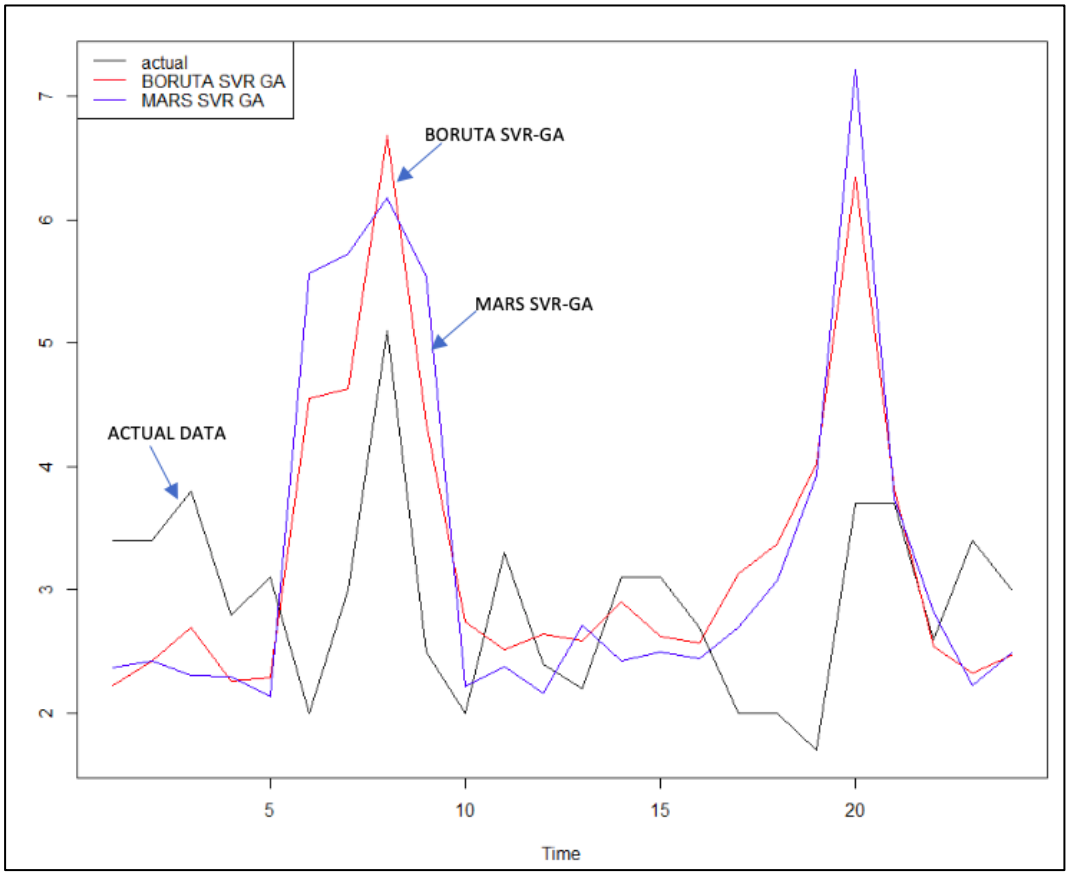

Figure 9. Actual Data VS BORUTA-SVR-GA VS MARS-SVR-GA

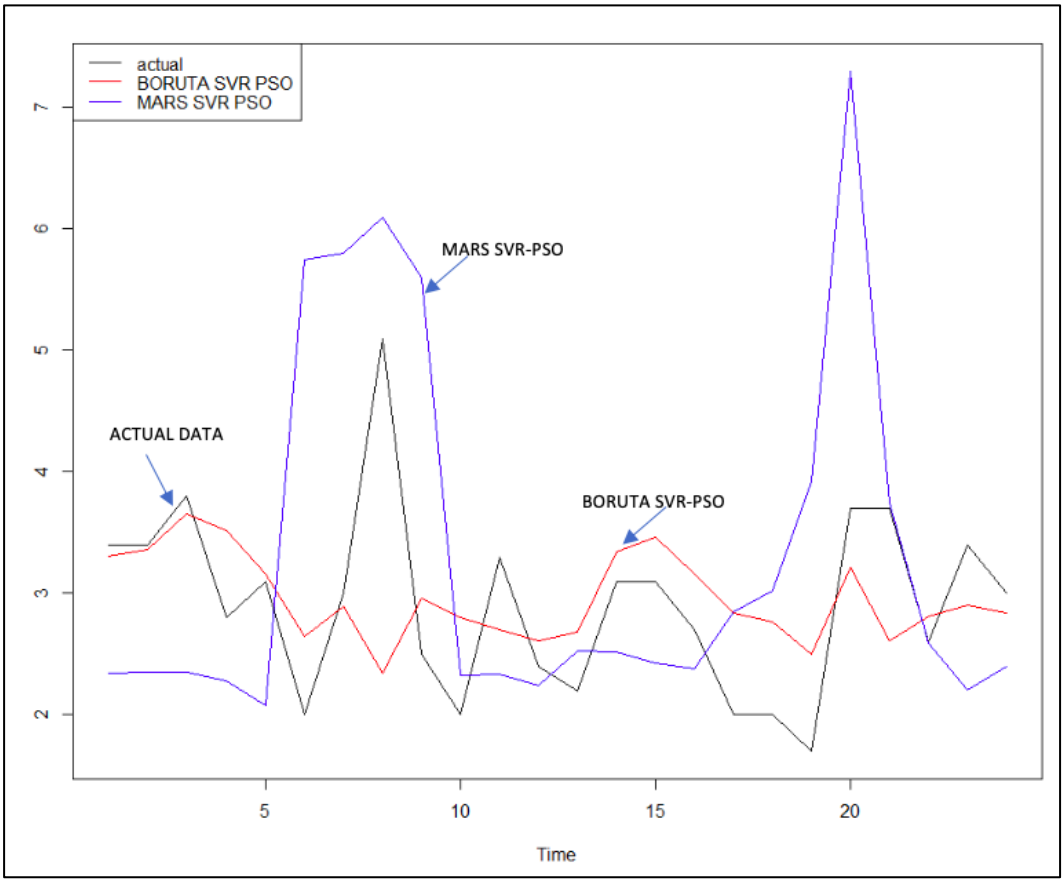

Figure 10. Actual Data VS BORUTA-SVR-PSO VS MARS-SVR- PSO

E. Energy Forecasting

Based on forecasting results, both BORUTA-SVR-GA and MARS-SVR-GA. The average wind speed in the area is around 2 m/s - 3 m/s. The working principle of a wind turbine to produce electrical energy is to convert the kinetic energy that can move the propeller into mechanical energy in the turbine.

The propeller rotates on its axis so that it can convert the kinetic energy of the rotary into electrical energy. Based on the influence of wind speed on the turbine rotation, current and voltage, the work of the wind turbine can be seen from the power, torque and efficiency it generates.

Research on the conversion of wind energy into electrical energy on a laboratory scale can be done to prove it so that this energy can be examined in the future on a larger scale. Moreover, To produce electrical energy from wind speed, it can be converted into mechanical energy by windmills in the form of shaft rotation and subsequently by using a permanent magnet synchronous generator (PMSG)[47].

Moreover,Wind speed conversion system is a non-linear system. So in this paper, the feedback linearization control method is used to overcome this nonlinearity through changes in variables and corresponding control inputs so that maximum power can be identified. Then simulate the results of the analysis of the conversion of wind energy into electrical energy[48].

If we used a constant rotor speed of 30 rad/s and perform Feedback Linearization Control [49] at the wind speed of 3 m/s will produce energy 3.82KW, a wind speed of 4 m/s will produce 6.79KW, and a wind speed of 5 m/s produces 10.60 KW, respectively.

The potential for strong winds in North Sulawesi Province, it's because of not only the location of Indonesian region surrounding the equator was a Hadley, walker, and local circulation meeting but also tropical depressive and tropical cyclone activity in the Pacific Ocean east of the Philippines in every year caused the potential for strong winds in parts of North Sulawesi.

IV. CONCLUSION

The Lagrangian treatment of convex optimization problems leads to a dual description easier to solve than primal issues due to the difficulty of handling the inequality constraints directly. The dual challenge is obtained by introducing Lagrange multipliers, also called dual variables. The dual methods are based on the idea that only the dual variables are the primary unknown of the problem. We can transform the primal problem into a dual problem by making only the derivatives of the Lagrangian concerning primal variables to zero, and substituting the obtained relations into the Lagrangian. Therefore, we eliminate the dependence on primal variables correspond to calculate the function explicitly $\theta(\alpha,\beta) = \inf_{w\in\Omega} L(w,\alpha,\beta)$ resulting function contains only dual variables and should be maximized under more straightforward constraints. This strategy will be adopted later as one of the standard techniques in the theory of support vector machines. The use of dual representations in the theory of SVrs allows us not only to work in spaces of higher dimensions but also to use algorithmic techniques derived from optimization theory. Regarding feature selection, it gives a score for each element of our information, the higher the score increasingly significant or important is the component towards your yield variable. Besides the optimization, GA is sensitive to the initial population used. At the same time, PSO has a weakness when the algorithm is looking for an optimal solution; this method cannot accurately determine the optimal solution. In other words, PSO has reduced adjustments to the search space locally. In the next research, we will try other metaheuristics like bee colony[50], firefly algorithm[51], and Cuckoo Search Algorithm[52].

REFERENCES

[1] L. Martini, J. H. Tjakraatmadja, Y. Anggoro, A. Pritasari, and L. Hutapea, "Triple Helix Collaboration to Develop Economic Corridors as Knowledge Hub in Indonesia," *Procedia - Soc. Behav. Sci.*, vol. 52, pp. 130–139, 2012.

[2] S. Thomas, M. Richter, W. Lestari, S. Prabawaningtyas, Y. Anggoro, and I. Kuntoadji, "Transdisciplinary research methods in community energy development and governance in Indonesia: Insights for sustainability science," *Energy Res. Soc. Sci.*, vol. 45, pp. 184–194, 2018.

[3] Y.-C. Yang, C.-Y. Hong, S. Muzayyanah, and R. Adha, "Decomposition of Growth Factors in High-Tech Industries and CO2 Emissions: After The World Financial Crisis in 2008," *Int. J. Energy Econ. Policy*, vol. 10, no. 4, pp. 500–506, 2020.

[4] R. E. Caraka and P. C. Ekacitta, "Simulasi Kalkulator Energi Baru Terbarukan (EBT) Guna Memenuhi Ketahanan Energi di Indonesia," *Statistika*, vol. 16, no. 2, pp. 77–88, 2016.

[5] C. L. Archer and M. Z. Jacobson, "Evaluation of global wind power," *J. Geophys. Res. D Atmos.*, vol. 110, no. D12, pp. 1–20, 2005.

[6] S. H. Mohr and G. M. Evans, "Forecasting coal production until 2100," *Fuel*, vol. 88, no. 11, pp. 2059–2067, 2009.

[7] R. E. Caraka and H. Yasin, "Prediksi Produksi Gas Bumi Dengan General Regression Neural Network (GRNN)," in *National Seminar Statistics IV " Implementation Statistics in the Exploration Energy,"* 2014.

[8] L. Liu *et al.*, "Water demands for electricity generation in the U.S.: Modeling different scenarios for the water-energy nexus," *Technol. Forecast. Soc. Change*, vol. 94, pp. 318–334, 2015.

[9] F. Fusco and J. V. Ringwood, "Short-term wave forecasting for real-time control of wave energy converters," *IEEE Trans. Sustain. Energy*, vol. 1, no. 2, pp. 99–106, 2010.

[10] European Wind Energy Association, *Wind Energy - The Facts*. 2012.

[11] M. A. Delucchi and M. Z. Jacobson, "Providing all global energy with wind, water, and solar power, Part II: Reliability, system and transmission costs, and policies," *Energy Policy*, vol. 39, no. 3, pp. 1170–1190, 2011.

[12] F. Mahmuddin, M. Idrus, and Hamzah, "Analysis of Ocean Wind Energy Density around Sulawesi and Maluku Islands with Scatterometer Data," in *Energy Procedia*, 2015, pp. 107–115.

[13] T. Jansen, "Review of Tide and Wind Affect to Currents Pattern in Amurang Bay, North Sulawesi, Indonesia," in *Journal of Sustainable Engineering: Proceedings Series*, 2019, p. 68074.

[14] J. R. E. Harger, "ENSO variations and drought occurrence in Indonesia and the Philippines," *Atmos. Environ.*, vol. 29, no. 16, pp. 1943–1955, 1995.

[15] R. E. Caraka, B. D. Supatmanto, J. Soebagyo, M. A. Mauludin, A. Iskandar, and B. Pardamean, "Rainfall Forecasting Using PSPLINE and Rice Production with Ocean-Atmosphere Interaction," *IOP Conf. Ser. Earth Environ. Sci.*, 2018.

[16] Supari, F. Tangang, L. Juneng, and E. Aldrian, "Observed changes in extreme temperature and precipitation over Indonesia," *Int. J. Climatol.*, vol. 37, no. 4, pp. 1979–1997, 2017.

[17] A. Brezoescu, P. Castillo, and R. Lozano, "Wind estimation for accurate airplane path following applications," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2014, pp. 869–874.

[18] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *Int. J. Forecast.*, vol. 22, no. 3, pp. 443–473, 2006.

[19] H. Madsen, *Time series analysis*. 2007.

[20] Suhartono, B. S. S. Ulama, and A. J. Endharta, "Seasonal time series data forecasting by using neural networks multiscale autoregressive model," *Am. J. Appl. Sci.*, vol. 7, no. 10, 2010.

[21] R. E. Caraka, S. A. Bakar, M. Tahmid, H. Yasin, and I. D. Kurniawan, "Neurocomputing Fundamental Climate Analysis," *Telkomnika*, vol. 17, no. 4, pp. 1818–1827, 2019.
[22] H. Yasin, R. E. Caraka, Tarno, and A. Hoyyi, "Prediction of crude oil prices using support vector regression (SVR) with grid search - Cross validation algorithm," *Glob. J. Pure Appl. Math.*, vol. 12, no. 4, 2016.
[23] H. Iba and C. C. Aranha, "Introduction to genetic algorithms," *Adaptation, Learning, and Optimization*, 2012.
[24] K. Premalatha and A. M. Natarajan, "Hybrid PSO and GA for Global Maximization," *Int. J. Open Probl. Compt. Math*, 2009.
[25] X.-S. Yang, "Firefly Algorithms," in *Nature-Inspired Optimization Algorithms*, 2014.
[26] X. Yang, *Nature-Inspired Metaheuristic Algorithms*. 2010.
[27] A. J. Smola, B. Sch, and B. Schölkopf, "A Tutorial on Support Vector Regression," *Stat. Comput.*, vol. 14, no. 3, pp. 199–222, 2004.
[28] R. E. Caraka, S. A. Bakar, and M. Tahmid, "Rainfall Forecasting Multi Kernel Support Vector Regression Seasonal Autoregressive Integrated Moving Average," *AIP Conf. Proc.*, vol. 020014, no. June, 2019.
[29] F. M. Wuketits, "Charles darwin and modern moral philosophy," *Ludus Vitalis*, vol. XVII, no. 32, pp. 395–404, 2009.
[30] D. Whitley, "A genetic algorithm tutorial," *Stat. Comput.*, vol. 4, no. 2, pp. 65–85, 1994.
[31] C. García-Martínez, F. J. Rodriguez, and M. Lozano, "Genetic algorithms," in *Handbook of Heuristics*, 2018.
[32] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.
[33] R. . Caraka, R. C. Chen, H. Yasin, B. Pardamean, T. Toharudin, and S. H. Wu, "Prediction of Status Particulate Matter 2.5 using State Markov Chain Stochastic Process and HYBRID VAR-NN-PSO," *IEEE Access*, vol. 7, pp. 161654–161665, 2019.
[34] D. D. Cox and S. John, "A statistical method for global optimization," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 1992, pp. 1–15.
[35] B. Warsito, H. Yasin, and A. Prahutama, "Particle swarm optimization versus gradient based methods in optimizing neural network," *J. Phys. Conf. Ser.*, vol. 1217, p. 012101, 2019.
[36] M. a. Hall and L. a. Smith, "Practical feature subset selection for machine learning," 1998.
[37] C. Dewi and R.-C. Chen, "Random Forest And Support Vector Machine On Features Selection For Regression Analysis," *Int. J. Innov. Comput. Inf. Control*, vol. 15, no. 6, 2019.
[38] M. B. Kursa, A. Jankowski, and W. R. Rudnicki, "Boruta - A system for feature selection," *Fundam. Informaticae*, vol. 101, no. 4, pp. 271–285, 2010.
[39] M. B. Kursa and W. R. Rudnicki, " Feature Selection with the Boruta Package ," *J. Stat. Softw.*, 2015.
[40] B. S. Yandell and R. L. Eubank, "Spline Smoothing and Nonparametric Regression," *Technometrics*, vol. 31, no. 3, pp. 379–380, 1989.
[41] J. H. Friedman, "Rejoinder: Multivariate Adaptive Regression Splines," *Ann. Stat.*, vol. 19, no. 1, pp. 123–141, 1991.
[42] R. E. Caraka and S. A. Bakar, "Evaluation Performance of Hybrid Localized Multi Kernel SVR (LMKSVR) In Electrical Load Data Using 4 Different Optimizations," *J. Eng. Appl. Sci.*, vol. 13, no. 17, 2018.
[43] G. Sanchez-Torres and I. Díaz Bolaño, "Support Vector Regression for PM10 Concentration Modeling in Santa Marta Urban Area," *Eng. Lett.*, vol. 27, no. 3, pp. 432–440, 2019.
[44] D. D. Prastyo, F. S. Nabila, Suhartono, M. H. Lee, N. Suhermi, and S. F. Fam, "VAR and GSTAR-based feature selection in support vector regression for multivariate spatio-temporal forecasting," in *Communications in Computer and Information Science*, 2019, pp. 46–57.
[45] H. Yasin, R. E. Caraka, A. Hoyyi, and Sugito, "Stock Price Modelling Using Localized Multi Kernel Learning Support Vector Regression," *ICIC Express Lett. Part B Appl.*, vol. 11, no. 4, pp. 1–7, 2020.
[46] J.-S. Wang and J.-D. Song, "A Hybrid Algorithm Based on Gravitational Search and Particle Swarm Optimization Algorithm to Solve Function Optimization Problems," *Eng. Lett.*, vol. 25, no. 1, pp. 22–29, 2017.
[47] T. K. A. Brekken, A. Yokochi, A. Von Jouanne, Z. Z. Yen, H. M. Hapke, and D. A. Halamay, "Optimal energy storage sizing and control for wind power applications," *IEEE Trans. Sustain. Energy*, 2011.
[48] W. J. Jemai, H. Jerbi, and M. N. Abdelkrim, "Synthesis of an approximate feedback nonlinear control based on optimization methods," *WSEAS Trans. Syst. Control*, 2010.
[49] I. Rizkiani, Kamiran, and Subchan, "Analysis and Simulation of Wind Energy Conversion to Electrical Energy Using Methods Feedback Linearization Control," *Sains dan Seni*, vol. 1, no. 1, pp. 12–17, 2012.
[50] W. C. Hong, "Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm," *Energy*, vol. 36, no. 9, pp. 5568–5578, 2011.
[51] R. Moazenzadeh, B. Mohammadi, S. Shamshirband, and K. Chau, "Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 584–597, 2018.
[52] S. Wen, J. S. Wang, and J. Gao, "Fault diagnosis strategy of polymerization kettle equipment based on support vector machine and cuckoo search algorithm," *Eng. Lett.*, vol. 25, no. 4, pp. 474–482, 2017.

REZZY EKO CARAKA, received The B.S. degree (S.Si) from the Department of Statistics Diponegoro University and Master of Science by research (MSc-Res) School of Mathematical Sciences the National University of Malaysia. Moreover, in 2019 he starts Ph.D. in College of Informatics, Chaoyang University of Technology, Taiwan. He acts as a researcher in Bioinformatics & Data Science Research Center University of Bina Nusantara (BINUS). He also fellow researcher in lab GLM-H Department of Statistics, Seoul National University, South Korea under supervised Prof.Youngjo Lee. At the same time, He was co-founder Statistical Calculator (STATCAL). He already published 10 statistics book including the theoretical and application in wide area. His research interests include Statistical Climatology, Climate Modeling, Ecological Modelling

RUNG CHING CHEN, received a B.S. from the Department of Electrical Engineering in 1987, and an M. S. from the Institute of Computer Engineering in 1990, both from National Taiwan University of Science and Technology, Taipei, Taiwan. In 1998, he received his Ph.D. from the Department of Applied Mathematics in computer science, National Chung Hsing University. He is now a distinguished professor in the Department of Information Management, Taichung, Taiwan. His research interests include network technology, pattern recognition, and knowledge engineering, IoT and data analysis, and applications of Artificial Intelligence. Network technology, pattern recognition, and knowledge engineering, IoT and data analysis, and applications of Artificial Intelligence. crching@cyut.edu.tw

SAKHINAH ABU BAKAR, currently works at School of Mathematical Sciences, The National University of Malaysia. She earned Master of Science in Mathematics from Universiti Teknologi Malaysia (UTM) and her PhD in Computer Science from The University of Sydney, Australia. Her research interest including Computational Mathematics, Intelligent Systems, Machine learning and Network Science. sakhinah@ukm.edu.my

MUHAMMAD TAHMID, currently forecaster of climate in Meteorology, Climatology, and Geophysical Agency, Indonesia. His research interests include climate Change, Climatology, Climate Modelling. muhammad.tahmid@bmkg.go.id

TONI TOHARUDIN, currently Associate Professor at the Department of Statistics, Universitas Padjadjaran. Toni does researches in Statistics. He received the Master of Science University of Leuven Belgium (2004-2005) and Ph.D. Spatial Sciences University of Groningen (2007–2010). Moreover, Toni act as head of the research group in time series and regression. toni.toharudin@unpad.ac.id

BENS PARDAMEAN, has over thirty years of global experience in information technology, bioinformatics, and education, including a strong background in database systems, computer networks, and quantitative research. His professional experience includes being a practitioner, researcher, consultant, entrepreneur, and lecturer. After successfully leading the Bioinformatics Research Interest Group, he currently holds a dual appointment as the Director of Bioinformatics & Data Science Research Center (BDSRC) and as an Associate Professor of Computer Science at the University of Bina Nusantara (BINUS) in Jakarta, Indonesia. He earned a doctoral degree in informative research from the University of Southern California (USC), as well as a master's degree in computer education and a bachelor's degree in computer science from California State University, Los Angeles. bpardamean@binus.ac.id

SU-WEN HUANG, received the degree in China Medical University The Graduate Institute of Health Care Administration College of Public Health in 2005 and the Ph.D. of the Department of Information Management, the Chaoyang University of Technology in 2014. Currently, She is General Affairs Office Director in the Taichung Veterans General Hospital, Taichung, Taiwan. Her research interest in medical information, health insurance. Currently, She is head of the General Affairs Department of Taichung Veterans General Hospital, Taiwan. dale33663366@vghtc.gov.tw