Hybrid Support Vector Machine to Preterm Birth Prediction

Noviyanti Santoso*[1], Sri Pingit Wulandari[2]
[1,2]Department of Business Statistics, Institut Teknologi Sepuluh Nopember
Kampus ITS Sukolilo-Surabaya, 60111, Indonesia
e-mail: ***[1]noviyanti_s@statistika.its.ac.id**, [2]sri_pingit@statistika.its.ac.id

Abstrak

Kelahiran prematur merupakan salah satu kontributor kematian bayi yang utama. Masalah tersebut menjadi penting pada bidang kesehatan terutama reproduksi manusia baik di negara maju maupun negara berkembang. Pada tahun 2015 Indonesia menduduki ranking kelima untuk jumlah bayi lahir prematur tertinggi di dunia. Salah satu upaya untuk menekan angka bayi premature adalah dengan mengurangi factor resiko yang berkaitan dengannya. Pada penelitian kali ini akan dibuat model prediksi untuk kelahiran prematur dengan menggunakan metode hybrid multivariate adaptive regression spline (MARS) dan support vector machine (SVM). MARS digunakan untuk menyeleksi variabel yang diduga mempengaruhi bayi prematur. Hasil penelitian menunjukkan bahwa model hybrid MARS-SVM memberikan performa prediksi terbaik dibandingkan model lainnya.

Kata kunci—*kelahiran prematur, support vector machine, MRS, hybrid, klasifikasi*

Abstract

Preterm birth is one of the major contributors to perinatal and neonatal mortality. This issue became important in health research area especially human reproduction both in developed and developing country. In 2015 Indonesia rank fifth as the country with the highest number of premature babies in the world. The ability to reduce the number of preterm birth is to reduce risk factors associated with it. This research will be made the prediction model of preterm birth using hybrid multivariate adaptive regression splines (MARS) and Support Vector Machine (SVM). MARS used to select the attributes which suspected to affect premature babies. The result of this research is prediction model based on hybrid MARS-SVM obtains better performance than the other models.

Keywords— *preterm birth prediction, support vector machine, MARS, hybrid, classification*

1. INTRODUCTION

Preterm is abnormal birth according to gestational age and baby born called of premature infants. Premature birth means the birth of a live baby before 37 weeks of gestation from the normal 40 weeks or their weight are less than 2500 grams. The preterm birth is a risky thing because it has the potential to enhance perinatal mortality as much as 65%-75%. This issue became important in health research area especially human reproduction both in developed and developing country.

Preterm is relating to morbidity and mortality infants. It is one of the major contributors to perinatal and neonatal mortality, both short and long-term. Prematurity is the second leading

cause of death in infants after pneumonia and the leading cause of neonatal death. Thirty-five percent of the world's neonatal deaths are caused by complications of premature birth [1].

The World Health Organization (WHO) said that Indonesia ranks fifth as the country with the highest number of premature babies in the world. Based on data from the Central Bureau of Statistics (BPS) 2015, infant mortality (IMR) reached 25 deaths per 1,000 babies born. East Java is one of the provinces where preterm birth rates are high at 11.5 percent in 2014, which is above the national average of 10.2 percent. One of the regencies in East Java is Sumenep which preterm birth rate reaches 2.3 percent in 2014 [2].

Prematurity is a multifactor problem. Various studies have been conducted to look for risk factors for preterm birth. However, the presence of these risk factors does not necessarily lead to premature birth. Some premature births that occur spontaneously do not have a clear risk factor. There are no clear factors that can cause prematurity, so prevention through one or more factors may not work. Therefore, if you want to reduce the number of premature births, then the first step to preventing premature birth is to reduce risk factors associated with premature birth [3].

In previous studies, the researcher used Pearson correlation test [4] or binary logistics regression [5] [6] to identify factors causing preterm birth. The results showed that mother's condition became majority factors, such as age, education, activity, premature rupture of membranes, history of miscarriage, history of preterm, diabetes mellitus, and preeclampsia. In addition, birth order of babies was indicated as a factor causing preterm. Pearson correlation is a statistical method to measure and identify a relationship between two variables. The weakness of the correlation coefficient are that it only determine linear relationships among two variables. If the relationship is non-linear then the result is invalid. In addition to this, the correlation is useless if it is about categorical data.

Nowadays, data mining technique becomes popular because it showed better performance that the traditional one. In the development of data mining in the era of 1990 was emerging variety of classification methods, such as a decision tree (DT), multivariate adaptive regression splines (MARS), artificial neural networks (ANNs), and a new technique of support vector machine (SVM).

MARS performs the form of the development in splines basis functions, where the number of basis functions, as well as the parameters correlated with each one are regularly determined by the data [7]. There were various studies using MARS in health area to classification or prediction diseases. [8] have applied multivariate adaptive regression spline to predict hypertension in Indonesia which gave some variables important affecting blood pressure. [9] have used MARS and smooth support vector machine (SSVM) to diagnose breast cancer.

A number of effective prediction diseases model in hybrid technique have also been proposed in recent years. [10] proposed hybrid random forest and MARS, [11] integrated random forest and MARS to predict HIV patients in Surabaya Indonesia. However, MARS-SVM hybrid method in health research has not been used.

SVM is a new method in the data mining, which is a new procedure to overcome machine-learning problems by development of optimization approach [12]. There is three main concern when applying SVM to treat classification (1) selecting the optimal feature; (2) the choice of kernels; (3) the determination of the kernel's parameter. Feature selection is an important issue in the classification model. The reduction of feature is helpful to improve the prediction accuracy and computation time [13]. There is some functions kernel in SVM such as linear, polynomial and RBF. Many researchers suggest using RBF because it performs better than others. But the parameters and variables should be optimized to decrease the incorrect classification. In this paper, we proposed hybrid techniques depend on two steps: (1) using MARS to select input features, (2) using a grid search to optimize model parameters. The goal of this research is to calculate accuracy of prediction preterm birth in Sumenep, East Java, Indonesia using hybrid MARS-SVM.

2. METHODS

2.1 Multivariate Adaptive Regression Spline

Multivariate adaptive regression splines (MARS) was first suggested by Friedman as a adaptable method, which the model is built with enclose interactions between variables [7]. In the MARS algorithm, there are no assumptions about functional relationships between dependent variable and independent variables. Optimal transformations and interactions of variable can founded in the model. Moreover, the MARS model represent the complex data structure that characterizes the high dimensional data, hence can effectively expose the data patterns which is important. MARS model is used to overcome the weakness of RPR is to produce a continuous model on knots. The function of MARS can be described using the equation 1:

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

where a_0 and a_m are parameters, M is the number of basis functions, K_m is the number of knots, S_{km} takes on value of either 1 or -1 and indicates the right/left correlation step function, $v(k,m)$ is the independent variable labelling, and t_{km} define as the knot location.

In MARS modeling, knots are automatically determined from the data and generate a continuous model of knots, while to model selection on MARS using stepwise (forward and backward) methods [14]. Forward stepwise is performed to obtain function with the maximum number of basis functions. The criterion for selection of basis functions on forwarding stepwise is by minimizing the average sum of square residual (ASR). This base function is a parametric function defined in each region. Generally, the selected base function is polynomial with continuous derivative at each knot point. Friedman suggests the maximum number of base functions (BF) is 2 to 4 times the number of predictor variables. As for the maximum number of interactions (MI) are 1, 2 and 3 under consideration if more than 3 will result in an increasingly complex model. The minimum distance between knots or minimum observations between knots of 0, 1, 2, and 3. Fulfillment of the parsimony concept by backward stepwise, which is to select the basis function generated from forwarding stepwise by minimizing the value of generalized cross-validation (GCV) [15]. The decreasing GCV value when the variable is removed from the model can be used as a measure to determine the variable importance level. The minimum GCV function is defined as equation 2:

$$GCV(M) = \frac{ASR}{[1 - \frac{C(\tilde{M})}{n}]^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}[y_i - \hat{f}_M(x_i)]^2}{[1 - \frac{C(\tilde{M})}{n}]^2} \tag{2}$$

where there are n observations, and C(M) is the cost-penalty measures of a model containing M basis function (therefore the numerator measures the lack of fit on the M basis function model $f_M(x_i)$ and the denominator denotes the penalty for model complexity C(M)).

2. 2 Support Vector Machine

Support vector machine (SVM) is a promising technique. It follows the principle of structural risk minimization, which has been successfully used for data classification and regression in nonlinear modeling [12]. SVM uses the linear model to implement nonlinear class boundaries through some nonlinear mapping the input vectors x into the high-dimensional feature space. In the new space, an optimal separating hyperplane is constructed. Thus, SVM is known as the algorithm that needs a special kind of linear model, the maximum margin

hyperplane. The maximum margin hyperplane gives the maximum separation between the decision classes as shown in Figure 1. Support vectors are the training objects that are closest to the maximum margin hyperplane. Consider a training set $G = (x_i, y_i)$ where $i = 1, 2, ..., \lambda$ with $x = \{x_1, x_2, x_3\}... \subseteq \Re^N$ as input variables vector and $y = \{y_t, ..., y_\lambda\}... \subseteq \Re$ are output variables. The SVM function is:

$$f(x) = w \times \varphi(x) + b \tag{3}$$

where $\varphi(x)$ denotes the high-dimensional feature space, which is non-linearity mapped from input space x, w and b are coefficients that estimated by minimizing the regularized risk function.

Feature space usually has a higher dimension of the input space. This results in computing on feature space is very large because there is the possibility of feature space has an unlimited number of features and difficult to know the proper transformation function. To solve the problem, in SVM use kernel trick. By kernel method, a data x in input space is mapping to feature space F with higher dimension through map φ as well $\varphi : x \rightarrow \varphi(x)$. Therefore, data x as input space become $\varphi(x)$ in feature space. Hence, non linear regression function is formulated as follows:

$$f(x) = \sum_{i=1}^{\lambda} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \tag{4}$$

Here α_i and α_i^* are Lagrange multipliers. $K(x_i, x_j)$ is kernel function. Setting the kernel parameters became crucial because it can gain robust results. The most often used kernels are linear, polynomial, and radial basis function.
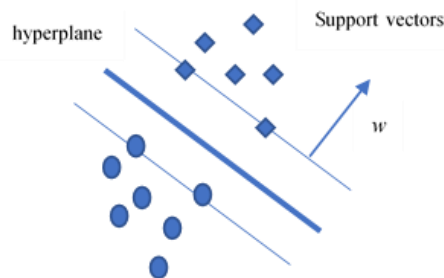

Figure 1 SVM with maximum margin hyperplane

2. 3 Performance Evaluation

To formulate criteria of performance of classification, statisticians work with confusion matrix (Table 1). Simply, *a* is the number of correctly classified class and *c* is the number of misclassified the positive class. Performance criteria is provided by accuracy rate. The higher the accuracy rate, the better the classification model performs.

$$Accuracy\ rate = \frac{a + d}{a + b + c + d} \tag{5}$$

But some researchers prefer to work with sensitivity and specificity with formula as follows:

$$Sensitivity = \frac{d}{d + c} \tag{6}$$

$$Specificity = \frac{a}{a+b} \tag{7}$$

The relation of sensitivity and specificity can be capture by what is called the relative operating characteristics (ROC) curve. The curve shows to what extent accuracy on positive class drop with reduced error rate on negative class and appropriate for unbalanced data [16]. An ideal score of area under curve (AUC) in range $0 - 100$. The larger area below the ROC curve, the higher the classification performs.

Table 1 Confusion matrix

		Predicted	
		Negative	Positive
Observation	Negative	a	b
	Positive	c	d

2. 4 Dataset

The data source is obtained from patient's medical record in one of the hospital in Sumenep, East Java, Indonesia from January 2015 until December 2015. The features in dataset consist of age, education, profession, premature rupture of membranes, history of miscarriage, and preeclampsia which are the categorical type, while the dependent variable is baby birth status preterm or normal. There are 428 patients in the dataset were randomly selected and then used to build the preterm birth models. Among them, 80% of the datasets will be used as the training set and the remaining 20% will be reserved as the testing set. Detail of each attribute shown in Table 2.

Table 2 Detail of attributes

Attributes	Definition	Category	Scale
Y	Baby birth status	0: normal 1: preterm	Nominal
X_1	Mother's age	0: < 20 years old 1: 20-35 years old 2: > 35 years old	Ordinal
X_2	Mother's education	0: not graduate school 1: primary school graduate 2: junior high school graduate 3: senior high school graduate 4: undergraduate	Ordinal
X_3	Mother's activity	0: stay at home 1: working mother	Nominal
X_4	Baby birth order	0: first 1: second 2: >2	Ordinal
X_5	Premature rupture of membranes	0: no 1: yes	Nominal
X_6	Preeclampsia	0: no 1: yes	Nominal

2. 5 Methodology

Babies birth is, in fact, classification, and the status (normal, preterm) is treated as classification label and the mother's condition such as age, education, history of miscarriage,

activity, preeclampsia, number of children, etc as classification attributes. Preterm birth prediction procedure based on SVM is, data collecting and preprocessing, selection of input features, selection of a kernel, determine best pair of parameters, build SVM classifier, and applied the model using testing dataset.

To improve the accuracy of preterm birth prediction, MARS is used to select the input features; grid search algorithm is adopted to obtain the optimal parameters, for each pair of parameters, 5-fold cross-validation is conducted on the training set. We implemented the SVM by LIBSVM [17]. The algorithm of predicting preterm birth is as follows and illustrated in Figure 2.

First, we use MARS to obtain variable importance with best combination of maximum basis function (BF = 18, 24), maximum of interaction (MI = 2, 3) and minimum observasion (MO = 0, 5, 10). Second, removing variables with zero importance, rebuild the model and randomly separate dataset into 80% training and 20% testing. Then, we choose the kernel function, in this study we use RBF. Considering parameter (C,γ) with C = (10, 30, 100) and γ = (0.01, 0.5, 2) for each pair paramaters, conduct 5-fold cross-validation on the training dataset. Furthermore, choosing the best combination of parameter (C,γ) and use to build the classification model of preterm birth prediction. The last is evaluating the accuracy of prediction using AUC score
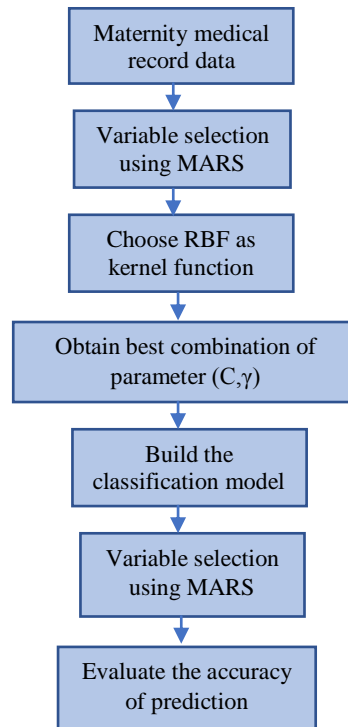
Figure 2 The algorithm of hybrid SVM technique

3. RESULTS AND DISCUSSION

The dataset consists of 428 instances which are categorized into two classes, normal and preterm baby born. Table 3 show distribution of category variable premature rupture of membranes in that classes.

Table 3 Crosstabulation baby birth status vs premature membranes

		premature rupture of membranes		Total
		no	yes	
Status	normal	347	24	371
	preterm	15	42	57
Total		362	66	428

Based on Table 3 preterm baby born was about 13.32% or equal to 57 cases from 428 baby birth in that hospital. As many as 42 mothers who give birth to premature baby were have premature rupture of membranes. While the distribution of category mother's age who give birth was shown by Table 4. The age of mother gives birth at most between 20-35 years old. It is a productive and fertile age that has a minimum risk. But, in fact, 27 mothers which that age are give birth premature baby. Hence, we must analyze more detail about that.

Table 4 Crosstabulation baby birth status vs mother's age

		age			Total
		<20	20-35	>35	
Status	normal	33	303	35	371
	preterm	21	27	9	57
Total		54	330	44	428

The software MARS 2.0 is not only provided the performs classification and regression problem well, but also provided variable importance ranking. In this paper, MARS used to select the feature of the applicant's information. In order to compare the performance of the proposed hybrid SVM technique, the prediction result of MARS is also presented.

We use X_1, X_2, …, X_6 to denote the features mentioned above, the result shows that the premature rupture of membranes is the most important, followed by age, education, activity, and the lowest is preeclampsia. The only variable which did not contribute is baby birth order. For details of the importance of the applicant's attributes, see Figure 3.
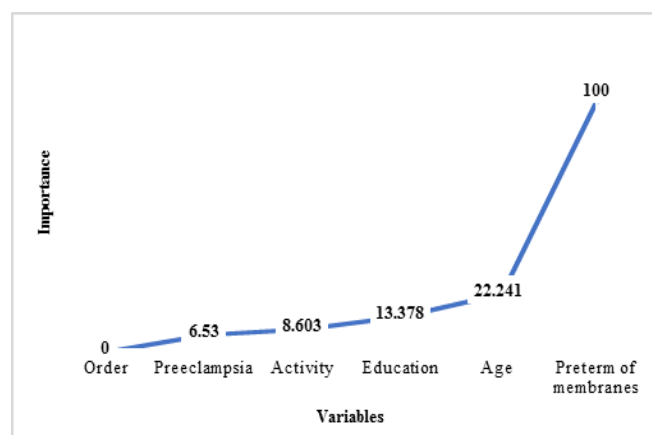

Figure 3 Variables importance

Rank importance of variables is indicator that was used in MARS algorithm to obtain best model. Means, the higher importance score, the higher it decrease GCV score. The best model was choosen based on the lowest GCV score from some results of combination BF, MI, and MO. Summary of GCV score each model was shown by Table 5. The best model was

obtained by best combination BF = 24, MI = 3, and MO = 0 with MARS GCV = 0.070 and accuracy rate = 92.11%. Actually, the highest accuracy occurred when MO = 10, but it followed by increasing GCV score too. Performance of MARS to predict baby birth status was stasifaying. It shown by score of accuracy rate (88.37%) and AUC (0.78) for testing dataset. But, the result will compare with the other models such as SVM and hybrid SVM to determine the best one.

Table 5 GCV score and accuracy rate of MARS models

BF	MI	MO	GCV	Accuracy
18	2	0	0.071	91.81%
18	2	10	0.071	91.81%
18	3	0	0.071	92.11%
24	2	0	0.071	91.81%
24	2	10	0.071	91.81%
24	3	0	0.070	92.11%
24	3	10	0.072	92.46%

To evaluate SVM performance without feature selection technique, we used training dataset which is consist of six variables and tried to build classification model using SVM. This research have been performed using Weka 3.8 which is provide LIBSVM method. RBF kernel function is choosen in SVM as classification model. Using 5-fold cross validation to obtain the best combination of parameters (C,γ) were C = 10 and γ = 0.5 with average of accuracy is 89.53%. Detail of the results shown by Table 6.

In order to improve the accuracy, this study conducted hybrid method, namely MARS-SVM. The main idea is integrate MARS and SVM procedure. As many as five variables have been selected using MARS, then used to build the classification model by SVM. The result was appropriate with our expectation that hybrid model increased the accuracy. The comparison of three methods are summarized in Table 7.

Table 6 Accuracy rate of SVM classification

C	γ	fold					average
		1	2	3	4	5	
10	0.01	90.21	89.1	88.23	90.12	88.01	89.13
	0.5	91.03	88.45	88.77	91.08	88.33	89.53
	2	89.13	88.45	88.78	90.32	87.44	88.82
30	0.01	92.13	86.43	88.57	89.1	90.03	89.25
	0.5	89.46	91.43	90.67	87.32	88.01	89.38
	2	90.78	90.05	88.12	87.79	87.19	88.79
100	0.01	91.44	88.97	88.08	89.21	88.17	89.17
	0.5	90.78	91.28	88.39	88.98	88.02	89.49
	2	88.23	89.48	90.26	88.18	91.1	89.45

Based on Table 7, we analyze that increasing accuracy rate is affect by feature selection. In this research, we used MARS to determine importance rank of variables. Hybrid MARS-SVM shown relative preferably results compare with MARS and SVM. The accuracy rate and specificity score are the highest, while sensitivity and AUC score are lower than SVM model.

Table 7 Classification results using MARS-SVM technique

Methods	Accuracy	Sensitivity	Specificity	AUC
MARS	88.37%	0.64	0.93	0.78
SVM	89.53%	0.93	0.88	0.91
MARS-SVM	93.02%	0.86	0.94	0.90

4. CONCLUSIONS

An accurate of preterm baby birth became crucial issue in the health research area, especially di Indonesia. Reducing the number of premature birth is effective by knowing risk factors associated with premature birth. Constructing the preterm birth prediction models from a patient's medical record database can be taken as a task of data mining. The artificial intelligence techniques do not require the knowledge of the underlying relationships between input and output variables. SVM and MARS are modern data mining techniques which suitable for regression and classification problems.

This research has successfully accomplished the objectives where three classification techniques (MARS, SVM, and hybrid MARS-SVM) were performed for preterm birth prediction. The main objective of this study is to identify the best technique for baby birth prediction. Hence, after applying the three techniques, a comparative analysis has been performed to determine the most appropriate technique. The experimental results showed that hybrid MARS-SVM perform well because of its abilities to predict the higher portion of data with higher accuracy rate and specificity.

For future work, the following suggestions can be considered; Combining other feature selection technique such as stepwise, correlation, etc; Use more features that can generalize or discriminate the classes has a significant impact on the effectiveness; Use more dataset and explore more areas and locations in Indonesia would be a valuable idea.

ACKNOWLEDGEMENT

REFERENCES

[1]   H. Blencowe, S. Cousens, D. Chou, M. Oestergaard, L. Say, A. B. Moller, M. Kinney and J. Lawn, "Born Too Soon: The global epidemiology of 15 million preterm births," Reproductive Health, vol. 10 (Suppl 1): S2, 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3828585/. [Accessed: 14-May-2018]

[2]   Dinas kesehatan Provinsi Jatim 2014. *Buku Profil Kesehatan Jatim 2014*. Jawa Timur.

[3]   D. Sulistiarini and S. M. Berliana, "Faktor-Faktor Yang Memengaruhi Kelahiran Prematur Di Indonesia: Analisis Data Riskesdas 2013," E-Journal WIDYA Kesehatan Dan Lingkungan, vol. 1, no. 2, pp. 109-115, March 2016. [Online]. Available: http://e-journal.jurwidyakop3.com/index.php/kes-ling/article/view/242. [Accessed: 12-April-2018]

[4]   A. A. H. Asl, S. Safari, and M. P. Hamrah, "Epidemiology and Related Risk Factors of Preterm Labor as an obstetrics emergency", *An Academic Emergency Medicine Journal*,

vol 5, no. 1, Jan. 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5325899/. [Accessed: 20-March-2018]

[5] Y. P. Zhang, X. H. Liu, S. H. Gao, J. M. Wang, Y. S. Gu, J. Y. Zhang, X. Zhou, and Q. X. Li, "Risk Factors for Preterm Birth in Five Maternal and Child Health Hospitals in Beijing", *PLoS ONE*, vol 7, no. 12, Dec. 2012. [Online]. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0052780. [Accessed: 20-March-2018]

[6] T. B. Temu, G. Masenga, J. Obure, D. Mosha, and M. J. Mahande, "Maternal and obstetric risk factors associated with preterm delivery at a referral hospital in northern-eastern Tanzania," Asian Pacific Journal of Reproduction, Vol. 5, Issue 5, pp. 365-370, Sept 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2305050016300768. [Accessed: 20-March-2018]

[7] J. H. Friedman, "Multivariate adaptive regression splines (with discussion)", *Annals of Statistics*, vol. 19, pp. 1–141, 1991.

[8] Suroto, B. W. Otok, Suharto, A. Wibowo, "Multivariate Adaptive Regression Spline for Prediction of Hypertension Cases the Measurement of Blood Pressure in Indonesia". *J.Appl. Environ. Biol. Sci.*, vol. 7, no. 5, pp. 41-46, 2017. [Online]. Available: https://www.textroad.com/pdf/JAEBS/J.%20Appl.%20Environ.%20Biol.%20Sci.,%207(5)41-46,%202017.pdf. [Accessed: 4-March-2018]

[9] S. W. Purnami, S. Andari, and Y. D. Pertiwi, "High-Dimensional Data Classification Based on Smooth Support Vector Machines", *Procedia Computer Science*, vol. 72, pp. 477 – 484, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915035905. [Accessed: 4-March-2018]

[10] D. Yao, J. Yang, and X. Zhan, "A Novel Method for Disease Prediction: Hybrid of Random Forest and Multivariate Adaptive Regression Splines", *Journal of Computers*, vol. 8, no. 1, 2013. [Online]. Available: http://www.jcomputers.us/index.php?m=content&c=index&a=show&catid=50&id=523. [Accessed: 20-March-2018]

[11] Nidhomuddin and B. W. Otok, "Random Forest Dan Multivariate Adaptive Regression Spline (MARS) Binary Response Untuk Klasifikasi Penderita HIV/AIDS Di Surabaya", *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 3, no.1, 2015. [Online]. Available: https://jurnal.unimus.ac.id/index.php/statistik/article/view/1439. [Accessed: 10-April-2018]

[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer, 1995.

[13] S. Li and S. Oh, "Improving feature selection performance using pairwise pre-evaluation," BioMed Central Bioinformatics, vol. 17, pp. 312, Aug 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4992252/. [Accessed: 14-May-2018]

[14] D. S. Kumar and S. Sukanya, "Feature Selection Using Multivariate Adaptive Regression Splines," International Journal of Research and Reviews in Applied Sciences And Engineering (IJRRASE), vol. 8, no.1, pp. 17-24, 2016. [Online]. Available: http://www.ijcns.com/pdf/ijrrasevol8no11016-4.pdf

[15] J. H. Friedman and B. W. Silverman, "Flexible Parsimony Smoothing and Additive Modelling", *Technometrics*, vol. 31, 1989.

[16] N. Santoso, W. Wibowo, "Financial Distress Prediction using Linear Discriminant Analysis and Support Vector Machine," in AIP Conf. Series: Journal of Physics: Conf. Series, 2018, vol. 979, pp. 012089. [Online]. Available: http://iopscience.iop.org/article/10.1088/1742-6596/979/1/012089.

[17] C.C. Chang and C. J. Lin, "LIBSVM: A library for support vector machine," 2013. Available: https://www.csie.ntu.edu.tw/~cjlin//papers/libsvm.pdf