



## INDONESIAN JOURNAL ON GEOSCIENCE

Geological Agency  
Ministry of Energy and Mineral Resources

Journal homepage: <http://ijog.geologi.esdm.go.id>  
ISSN 2355-9314, e-ISSN 2355-9306



### The Application of Parametric and Nonparametric Regression to Predict The Missing Well Log Data

MORDEKHAI MORDEKHAI<sup>1</sup>, IZZUL QUDSI<sup>2</sup>, and JOHN PAPILON STEVEN GUNTORO<sup>3</sup>

<sup>1</sup>Geophysical Engineering, Faculty of Mining and Petroleum Engineering, Institut Teknologi Bandung, Indonesia

<sup>2</sup>P.T. Horizon Perdana Internasional, Indonesia

<sup>3</sup>Pusat Kerja Sama Minyak dan Gas Bumi Universitas Trisakti, Indonesia

Corresponding author: [dekha.68010@gmail.com](mailto:dekha.68010@gmail.com)

Manuscript received: July, 17, 2019; revised: December, 18, 2019;

approved: April, 14, 2021; available online: August, 22, 2021

**Abstract** - Incomplete well log data are very commonly encountered problems in petroleum exploration activity. The development of artificial intelligence technology offers a new possible way to predict the required logs using limited information available. Optimizing conventional statistical theory, machine learning is proven to be a reliable tool for any prediction task in many fields of study. Regression is one of the basic methods that has rapid development and evolved many techniques with different approaches and purposes. In this study, parametric and nonparametric regressions {linear regression, Support Vector Machine (SVM), and Gaussian Process Regression (GPR)} are compared to predict the missing log using the available nearby data. Feature selection was done by performing Principal Component Analysis (PCA) on predictor variables. Different profile of PCA is observed between Cibulakan and Parigi Formations, which is the basis of conducting separate models based on the formation. Among all the selected methods, GPR is consistently making slightly better results. The correlation between the predicted and actual porosity of GPR is observed to be up to 0.19 higher compared to the other methods. Similar observation is also found on the Root Mean Squared Error (RMSE) value comparison. In practice, the GPR method has an inherent advantage compared to other methods, as it provides uncertainty to the prediction based on the standard deviation of each estimation result. The standard deviation of the GPR prediction ranges from 0.006 in high confidence cases and up to 0.077 where uncertainty is high. The models are considered robust and stable according to the RMSE evaluation from cross validation which is consistently giving the value below 0.04. In conclusion, the reliability of regression techniques for predicting the missing well log is exposed in this study, which results demonstrate steady and good accuracy in every formation which are tested on any well logs.

**Keywords:** well log, log prediction, regression, artificial intelligence, machine learning

© IJOG - 2021

#### How to cite this article:

Mordekhai, M., Qudsi, I., and Guntoro, J.P.S., 2021. The Application of Parametric And Nonparametric Regression to Predict The Missing Well Log Data. *Indonesian Journal on Geoscience*, 8 (3), p.385-399. DOI: [10.17014/ijog.8.3.385-399](https://doi.org/10.17014/ijog.8.3.385-399)

## INTRODUCTION

### Background

Ideally, complete wireline log data contain resistivity log (deep, medium, and shallow), porosity log (density, neutron, and sonic) and lithology log (gamma ray and spontaneous

potential). Sometimes miscellaneous logs such as caliper and spectral noise are also available. Unfortunately, this is a rare privilege especially for older operating fields.

It is not uncommon for some logs to be absent, either partially (specific interval) or completely unrecorded. This particular problem could ham-

per the petrophysicist task in estimating the physical properties of the rocks. For instance, the lack of neutron log of the target interval would disrupt the porosity calculation of the reservoir rock.

To overcome this issue, researchers are trying to estimate the value of missing logs using various methods. For example, Bader *et al.* (2018) estimates the missing log using statistical approach by correlating multiple surrounding well logs using local similarity (LSIM). In recent years, the development of artificial intelligence (AI) technology provides. Advanced machine learning (ML) technique was used by many studies in generating the synthetic well logs (Rolon *et al.*, 2009; Parapuram *et al.*, 2015; Salehi *et al.*, 2017). Lately, the implementation of ML on log prediction goes beyond the other log data. Kanfar *et al.* (2020) predicted the real-time well log by creating a model from the drilling parameters.

While most of the mentioned papers above applied neural network techniques on their research, and this paper utilized the combination of the conventional statistical methods with a simple ML approach to determine the missing logs. To be precise, the reliability of the estimation result was assessed from the regression method using the existing logs from multiple wells nearby. Principal Component Analysis (PCA) was done earlier as the exploratory data analysis of the predictor variables (available logs) to select the features that were used later in the regression prediction. This paper is intended to provide a straightforward yet trustworthy prediction technique.

### Regional Geology and Stratigraphy

Geologically, this research is situated on the Northwest Java Basin. According to Noble *et al.* (1997) (Figure 1), this basin comprises two

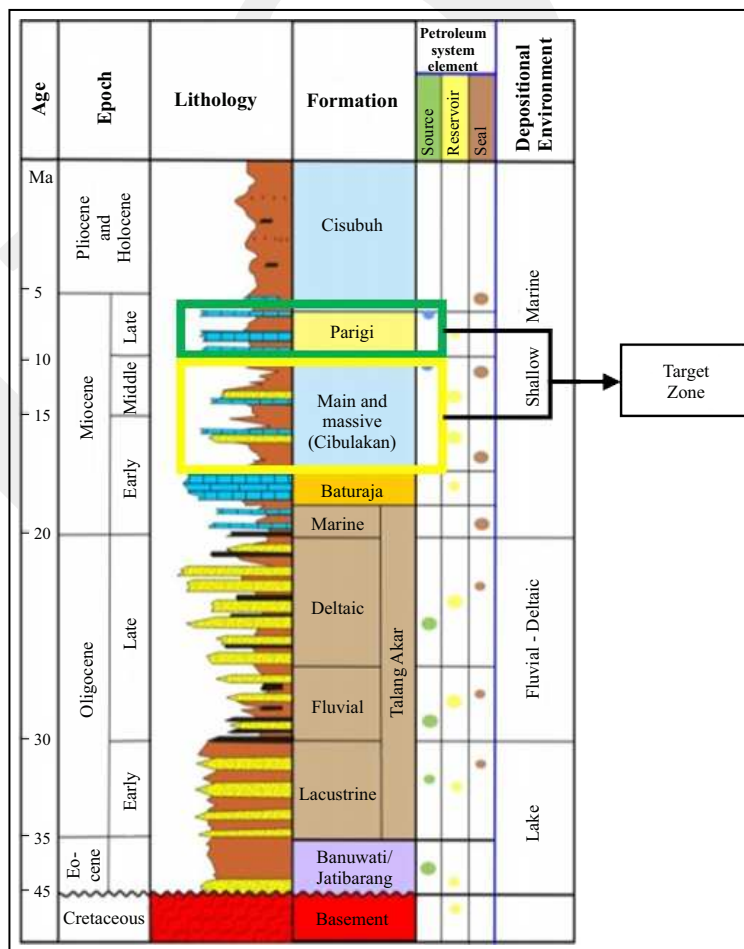


Figure 1. Northwest Java Basin stratigraphic column (Noble *et al.*, 1997).

main half-grabens, Ardjuna Basin, and Jatibarang Basin. The wells used in this research are part of Jatibarang/Talang Akar petroleum system which is categorized as Ardjuna Assessment Unit (Bishop, 2000).

Among many reservoir formations that occurred in this system, the wells cover the Cibulakan Formation and Parigi Formation. The first mentioned formation, Cibulakan, was deposited in Early to Middle Miocene. In this well, two members of this formation are exposed which are the main Cibulakan that consists of interbedded shales, sandstones, siltstones, and limestones (Butterworth *et al.*, 1995) and pre-Parigi that contains localized carbonate, the dolomited wackestone to grainstone (Pertamina BPPKA, 1996). While the second formation, Parigi, comprises carbonate platform and regressive clastics developing in the late post-rift phase (Doust and Noble, 2008).

## METHODS AND MATERIALS

In general, the aim of this research is to predict the missing log (neutron log in this case) using other well data nearby. In this research, models were created for each formation interval. These models later tested on a well log called validation well to check which results that produce the highest accuracy. Three regression methods those are Gaussian Process Regression (GPR), Support Vector Machine (SVM), and Linear Regression are the selected approaches performed in this research. Fundamentally, these three methods have different characteristics. Linear regression and Support Vector Machine (SVM) are parametric regression, while Gaussian Process Regression (GPR) is nonparametric regression. Parametric regression is a regression whose curve pattern

$f(x)$  is known, while nonparametric regression is a regression whose curve pattern  $f(x)$  is not known beforehand.

Before applying those regression methods, distinct features need to be selected for the model creation. A proper feature selection is one of the deciding factors in precisely estimating the value of the missing log, and PCA was executed to get the finest possible result. This was also done to understand the contribution (variance) of each parameter against the variable of the predicted log. In this study, all the data processing was done in a Python programming environment. To summarize, this workflow below explains the steps of this study (Figure 2).

### Linear Regression

A simple linear regression model is a model with a single regressor  $x$  that has a relationship with a response  $y$  that is a straight line (Douglas *et al.*, 2012), as in the Equation 1.

$$y = \beta_0 + \beta_1 x + \varepsilon \dots\dots\dots (1)$$

where:

$y$  is the predicted value,

$\beta_0$  is the intercept,

$\beta_1$  is the slope, and

the difference between the observed value of  $y$  and the straight line  $(\beta_0 + \beta_1 x)$  is the error,  $\varepsilon$ .

This study used more than one predictor variable ( $x$ ), so it uses multiple linear regression models as in the Equation 2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n + \varepsilon \dots\dots\dots (1)$$

where:

$x_1$ ,  $x_2$ , and  $x_n$  are the selected well logs with unknown parameters  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_n$ .

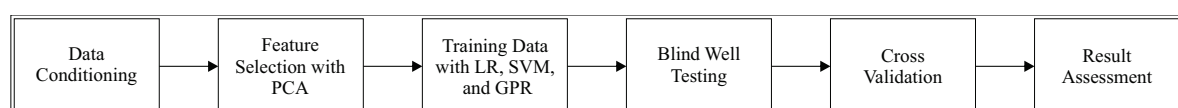


Figure 2. Research workflow.

Assuming that one wants to make a 3-dimensional regression model, hence they are supposed to use  $n=2$  ( $x_1$  and  $x_2$ ) as presented in Figure 3.

### Support Vector Machine (SVM)

Kecman (2005) states that the learning problem setting for SVMs is as follows: there is some unknown and nonlinear dependency (mapping, function)  $y = f(x)$  between some high-dimensional input vector  $x$  and scalar output  $y$  (or the vector output  $y$  as in the case of multiclass SVMs). There is no information about the underlying joint probability functions. Thus, one must perform a distribution-free learning. The only information available is a training data set

$D = \{(x_i, y_i) \in X \times Y\}$ ,  $i = [1, \dots, l]$ , where  $l$  stands for the number of the training data pairs, and is therefore equal to the size of the training data set  $D$ . Often,  $y_i$  is denoted as  $d_i$ , where  $d$  stands for a desired (target) value. Hence, SVMs belong to the supervised learning techniques.

SVM for regression basically works based on a hyperplane and large margin classifier. For the reason of visualization, assume that there is a 2 classifier/predictor as shown in Figure 4.

During the learning stages, our SVM model will find parameter  $w = [w_1 w_2 \dots w_n]^T$ , is the predictor variable from well data, and  $w_i$  is the weight of each variable. The equation  $d(x, w, b)$  is given as in Equation 3.

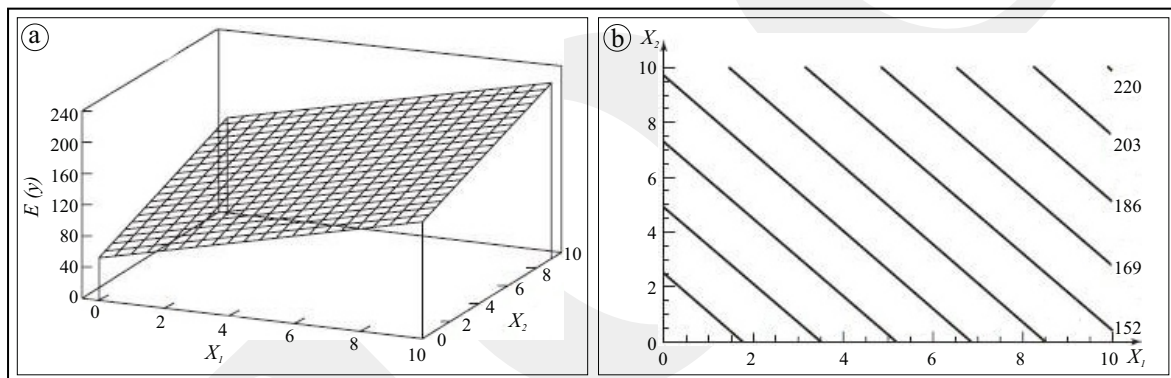


Figure 3. (a). Regression plane for the model  $E(y) = 50 + 10x_1 + 7x_2$ , (b). Contour plot (Douglas *et al.*, 2012).

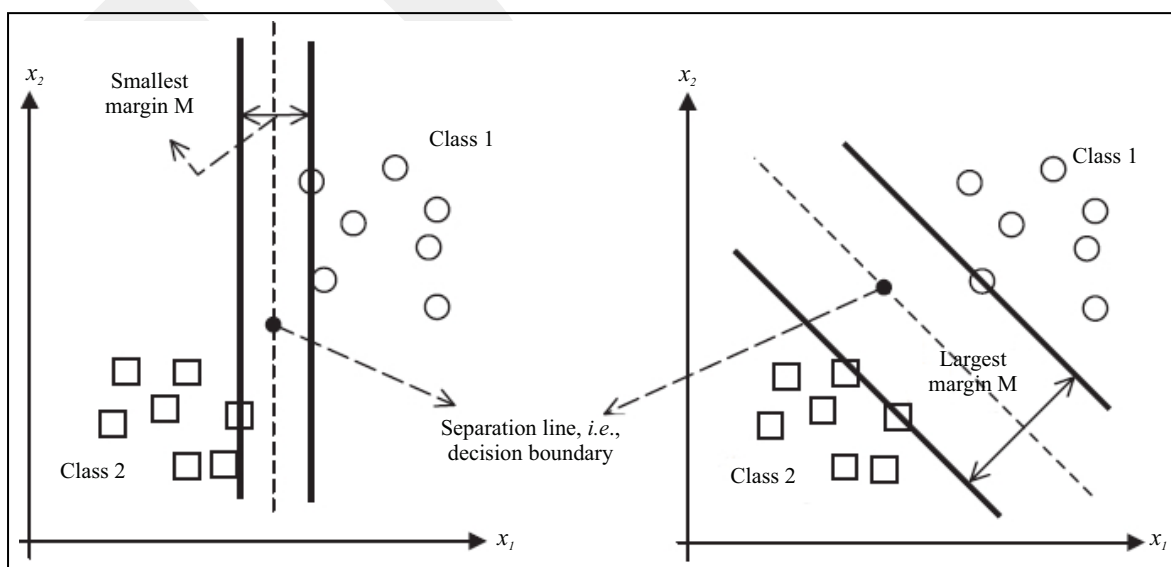


Figure 4. Two-out-of-many separating lines: a good one with a large margin (right) and a less acceptable separating line with a small margin (left) (Kecman, 2005).

$$d(x, w, b) = W^T x + b = \sum_{i=1}^n w_i x_i + b \dots\dots\dots (3)$$

### Gaussian Process Regression

A Gaussian process is a generalization of the Gaussian probability distribution; whereas a probability distribution describes random variables which are scalars or vectors (for multivariate distributions), a stochastic process governs the properties of functions.

Unlike the other supervised machine learning method, Gaussian process is a nonparametric model. There is no worry whether it is possible for the model to fit the data since Gaussian Process infers a probability distribution over all possible values using Bayesian approach. The combination of the prior model and the training data leads to a posterior distribution model.

The mean prediction is shown as a solid line and four samples from the posterior are shown as dashed lines (Figure 5). In both plots, the shaded region denotes twice the standard deviation at each input value  $x$  (Rasmussen and Williams, 2006). Figure 5 also shows that  $f(x)$  is the variable

that is wanted to be predicted (NPHI), while input  $x$  are the well predictor variables.

### Principal Component Analysis (PCA)

PCA is an unsupervised machine learning procedure which could find the patterns of variation from much information without reference to prior knowledge about the data itself. This method allows the dimensionality reduction without losing much important information.

The operation was done by using the linear combination of the original dataset and transforming into a new dimensional space using the eigenvector of each data, which could act as a good summary of the data (Lever, 2017). By this advantage, which parameters (log data) that could be left were assessed without reducing the quality of the result and possibly increased the accuracy in the other ways.

Eigenvalue and eigenvector is a pair of special scalar in a linear equation (*i.e.* matrix equation). In matrix transformation, this information could be the guidance to restore the information from the original matrix. While the eigenvector keeps the direction of the transformation, the eigenvalue indicates the original information that was retained. In this operation (PCA), the direction of the new coordinate axis is the eigenvector and the variations of each parameter displayed by the eigenvalue or the axis, higher eigenvalue indicate the higher variation (Wallisch, 2014).

Abdelaziz *et al.* (2017) illustrate the equation of the full component transformation of  $X$  as below (Figure 6):

$$T_{n \times p} = X_{n \times p} W_{n \times p} \dots\dots\dots (4)$$

where  $T$  is the score matrix with each column representing the value of each principal component at  $(n)$  observations. This matrix was generated from the multiplication of  $X$  which is the original data set matrix consisting of  $(p)$  as each the value of variables at  $(n)$  observations and  $W$  as the weight (loading matrix) of the transformation to the new dimension.

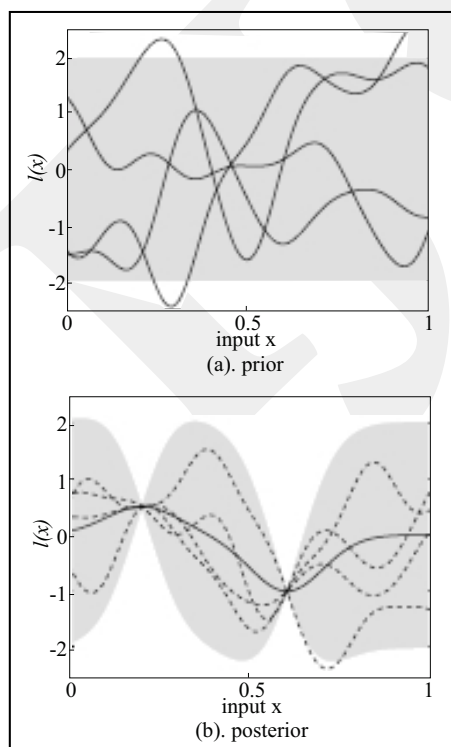


Figure 5. (a). Four samples drawn from the prior distribution; (b). A situation after two data points have been observed (Rasmussen and Williams, 2006).



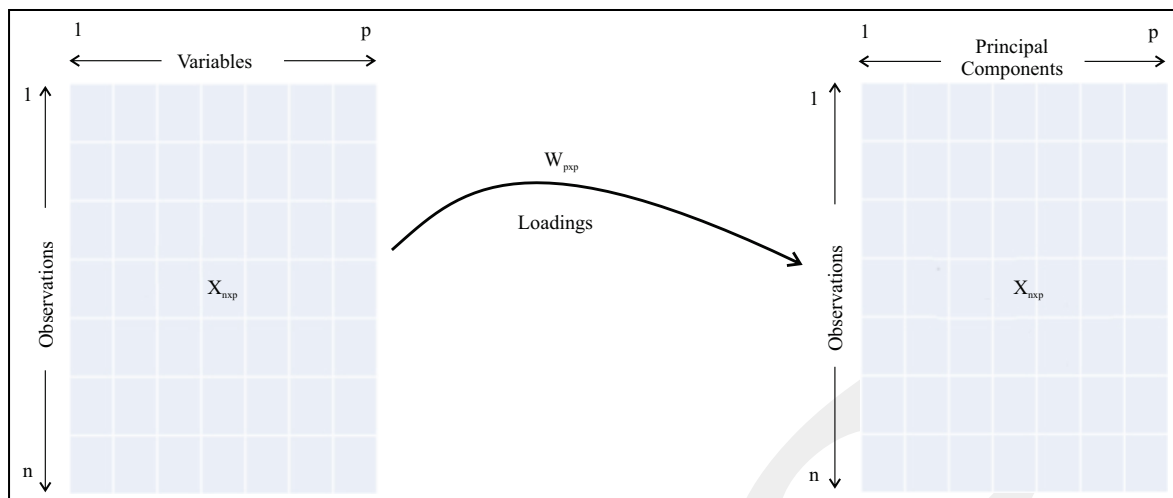


Figure 6. Principal component analysis illustration (Abdelaziz *et al.*, 2017).

In this research, the contribution of the parameters (log informations) is assessed against the target response through the eigenvector of each parameter and decided which features are used for the regression predictions. This method is analogous to the study by Roden *et al.* (2015) for feature selection analysis, where it is inherently assumed that the variation that is contained for each feature is directly proportional to the feature efficacy.

### Grid Search

Grid search is an algorithm that can choose the best parameters for a model based on the given parameter options. This process can automate the “trial and error” method of selecting the best parameters in a regression model. Grid search is then applied to SVM and GPR methods (Figure 7), since these regression methods have hyperparameters that are hard to optimize manually.

### Cross Validation

Cross validation (CV) is a statistical method that can be used to evaluate the performance of a model or algorithm where data are separated into two subsets, *i.e.* training data and validation data. First, merging all the well data is needed (seven wells for Cibulakan Formation and five wells for Parigi and pre-Parigi Formation). Cross validation is done by using K-fold CV. K-Fold CV will separate the dataset (well data) into K-subset. In

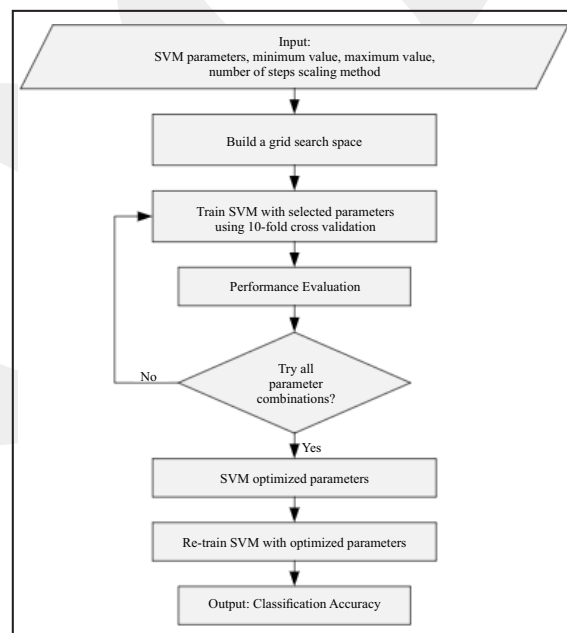


Figure 7. SVM hyperparameter optimization workflow using grid search (Syarif *et al.*, 2016).

this study, ten-fold CV was used as shown at the illustration Figure 8.

For each of the ten subsets of data, CV will use nine folds for training and one fold for testing. This ten-fold CV is then applied to the three regression methods used. The purpose of applying the CV method is to present the model overfitting which is more prone if only one validation set is used. After testing is done, the RMSE calculation was performed to see the accuracy of the model obtained from all of the three regression methods.

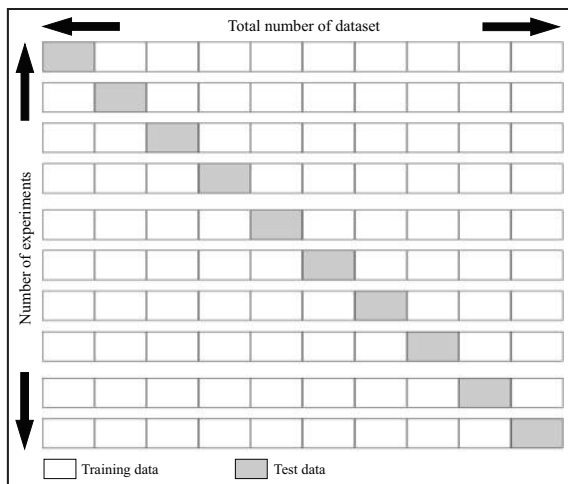


Figure 8. Examples of ten-fold cross validation (Talpur, 2017).

### Materials

This research used seven well-log dataset (Table 1 and see the map at Figure 9 for the distribution of the well locations). For the model creation, only five wells (1 - 5) were used as the training dataset. Well #6 and Well #7 were prepared as the implementation/test dataset of the prediction models (Figures 10, 11, and 12). Ten-fold cross validation (Figure 8) was performed to assess the reliability (avoid the possibility of overfitting) of the models on the results from the prediction of Well #6 and Well #7.

Since the neutron log is the one that is predicted in this research, the other logs (CALI, ILD, MSFL, RHOB, VP, P\_IMP) were used as the predictor parameters.

Table 1. Well Data Descriptions

Well	Remark
Well-1	
Well-2	
Well-3	
Well-4	
Well-5	Not used for Parigi and Pre-Parigi Fm.
Well-6	
Well-7	Not used for Parigi and Pre-Parigi Fm.

Training well  
Testing well

## RESULTS AND DISCUSSION

### Feature Selection

Data normalization was performed to the combined training wells to reduce the possibility of data redundancy and prevent the anomalous information caused by the value range difference of each log. PCA was performed to all the predictor variables on the training wells. This process changes the predictor variable to the principal component (eigenvector and eigenvalue). Eigenvectors were used as the guidance to select a certain feature that was used later in the regression prediction. The first two components of the PCA, namely PC1 and PC2, are shown in Figure 13.

In this study, the eigenvector of PC1 is analyzed which is the main component, and has a very large contribution to the variance of the



Figure 9. Basemap showing distribution of well locations.

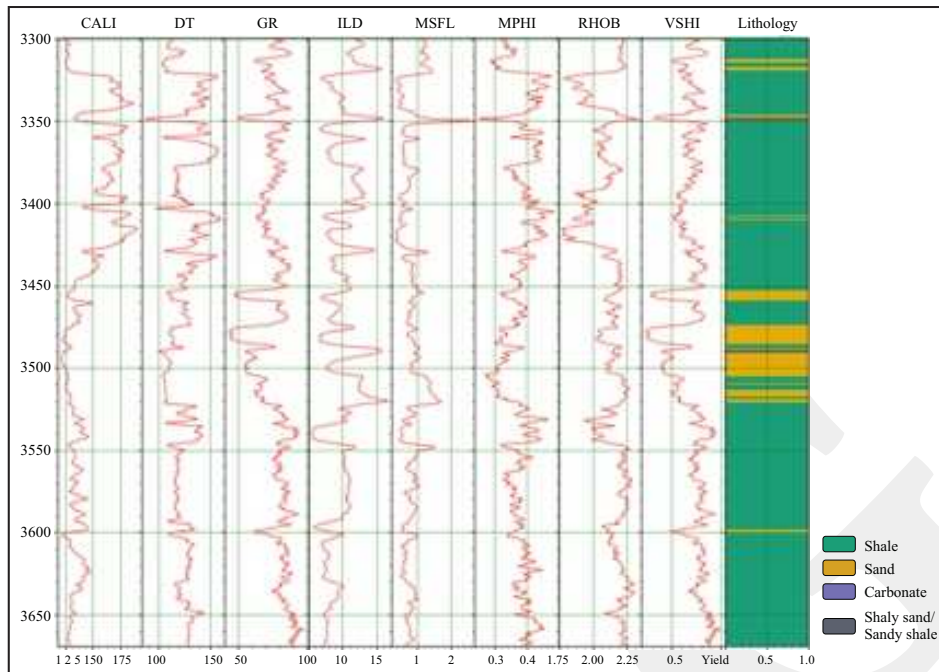


Figure 10. Preview of the log from Well #6 at interval of Cibulakan Formation.

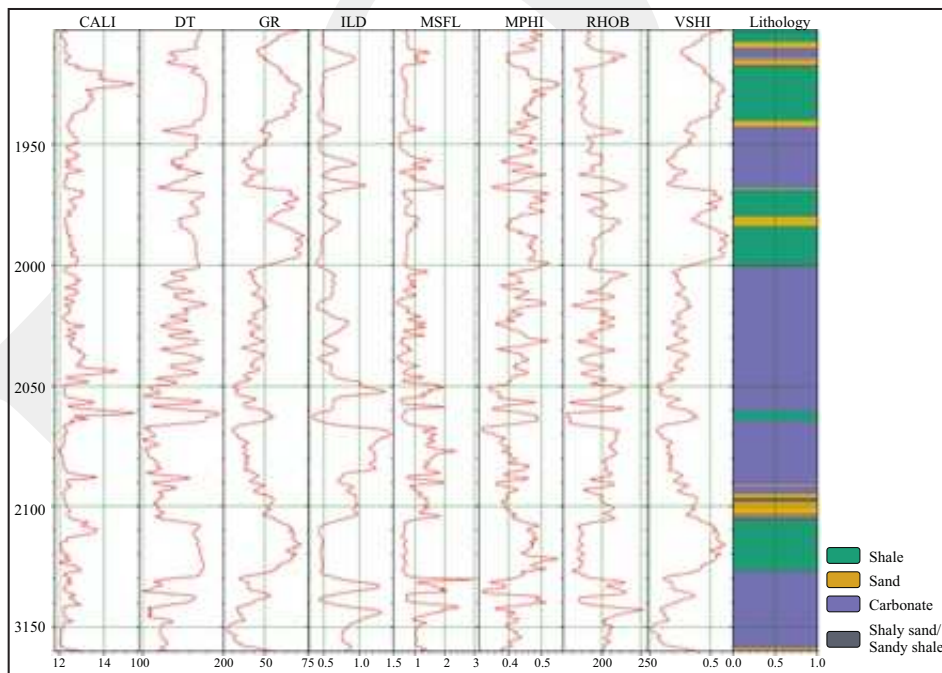


Figure 11. Preview of the log from Well #6 at interval of Parigi Formation.

entire dataset (around 80%). The absolute value of each PC1 coefficients is shown on Figure 14.

The PCA eigenvector results from each formation indicates that parameters with higher correlation to the prediction target (NPHI log) are different. While there are no dominant parameters

correlated with NPHI log at pre-Parigi and Parigi Formations, the result from Cibulakan Formation shows four dominant parameters (ILD, VSh, VP, and P Imp) which have high correlations with NPHI log. Due to this, all the logs were used at pre-Parigi and Parigi Formations as the training



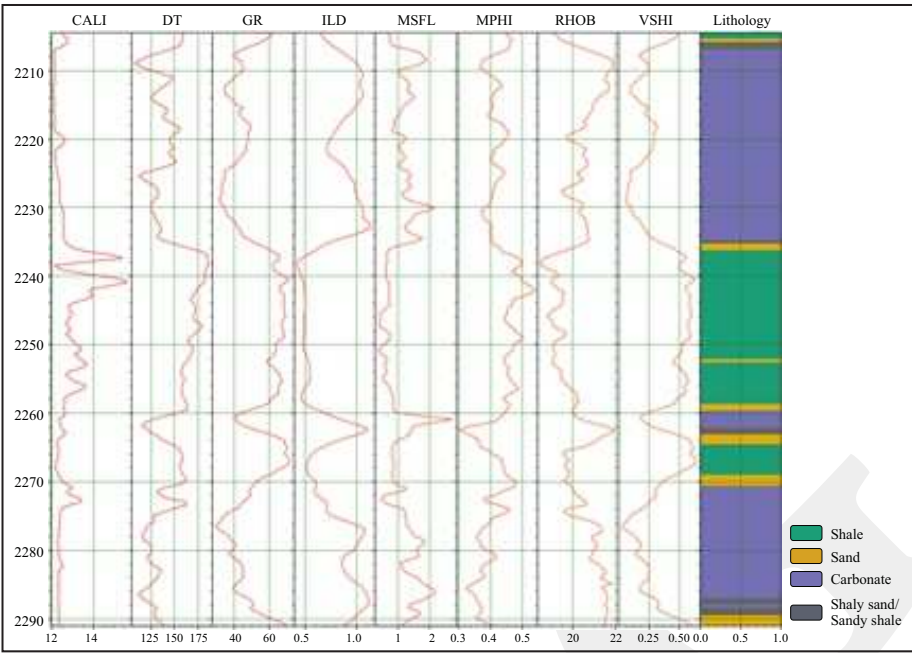


Figure 12. Preview of the log from Well #6 at the interval of pre-Parigi Formation.

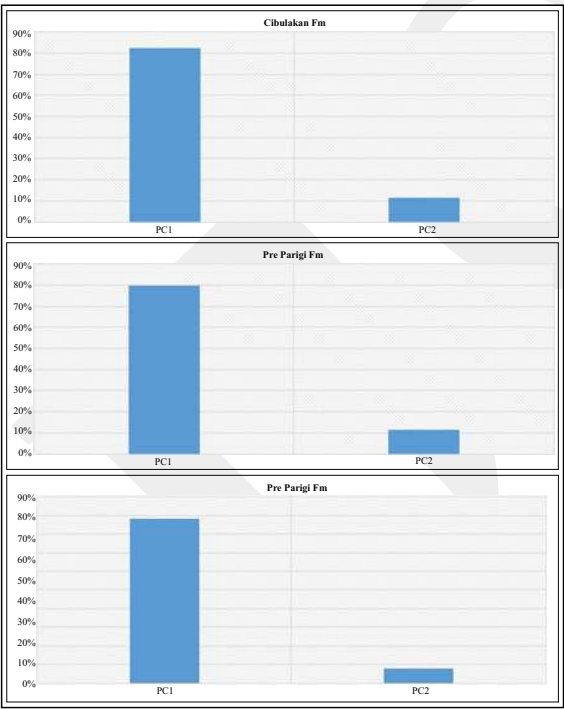


Figure 13. Principal components of each formation.

set and only four logs were used to create the model at Cibulakan Formation interval.

Through comparison of the PCA results, the lithology variation is hypothesized, and it is highly influencing the eigenvalue of the predictor variable on the feature selection. The pre-Parigi

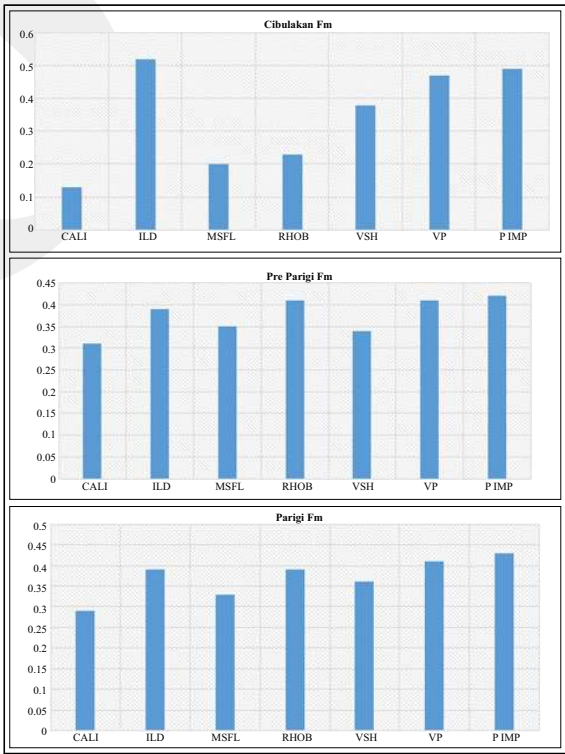


Figure 14. The absolute eigenvectors from each parameter on three different parameters.

and Parigi Formations are discovered to have more complex lithology composition than the Cibulakan Formation which shows different eigenvector results. Due to the variation of eigen-

vectors from each formation, different models for each interval were created to get the best possible prediction result.

### Prediction Comparison

As mentioned earlier, the predictions were done in three sets of regression methods. Each of the models from Cibulakan Formation was implemented on two test wells (Well #6 and Well #7). Comparison of the result from both wells is presented at Figures 15, 16, and 17.

Grid search has been carried out to determine the most optimum hyperparameters in SVM and GPR methods. In this formation, the SVM approach produces the lowest quality of prediction result with correlation value at 0.63 and 0.64 compared to the linear (0.85 and 0.76) and GPR method (0.82 and 0.77). The RMSE of SVM is also relatively higher than the other two on both wells (see the details at Tables 2 and 3).

Figures 18 and 19 show the results of the three regression methods in Parigi and pre-Parigi

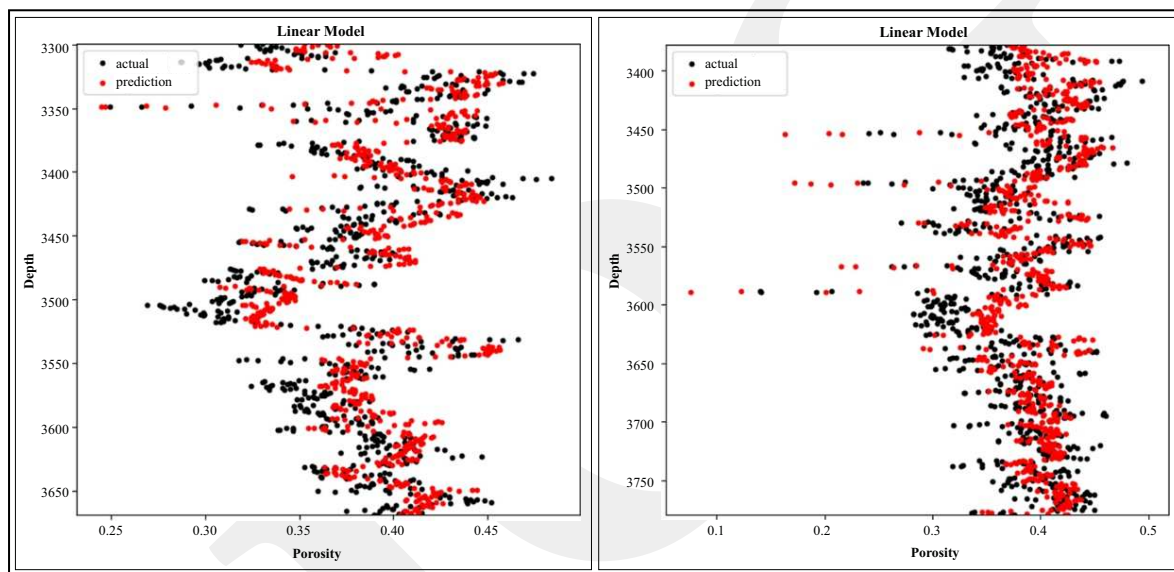


Figure 15. Result of linear regressions on Well #6 and Well #7 at Cibulakan Formation.

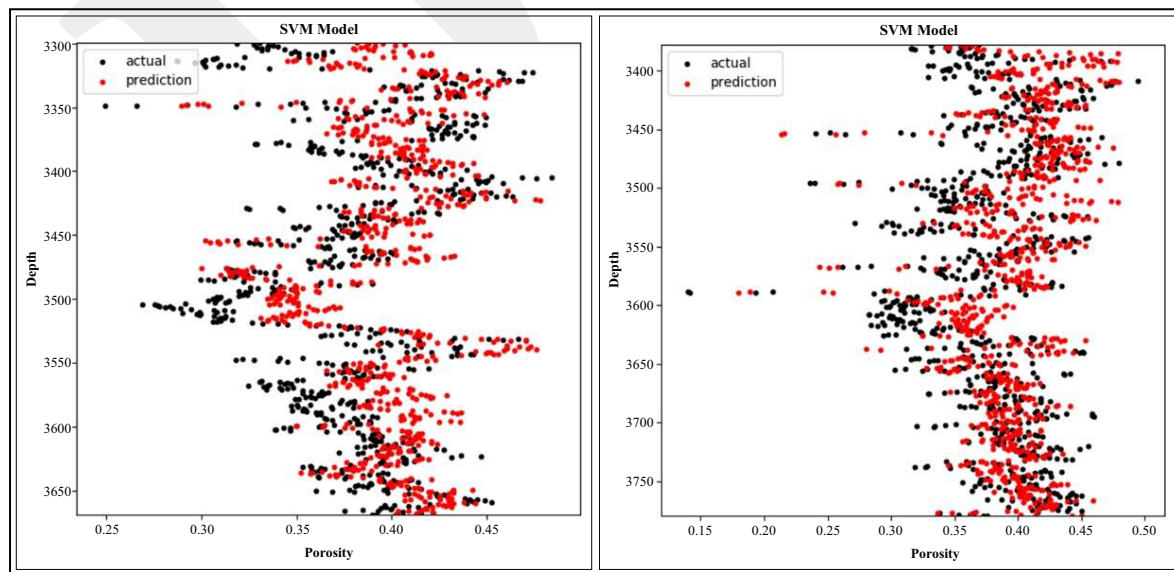


Figure 16. Result of Support Vector Machine on Well #6 and Well #7 at Cibulakan Formation.

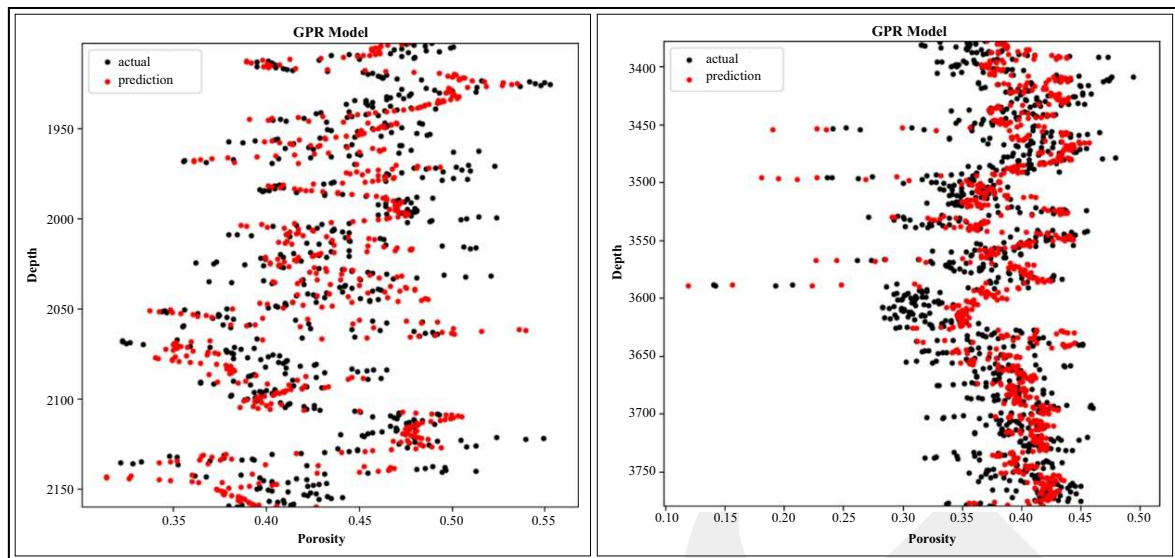


Figure 17. Result of Gaussian Process Regression on Well #6 and Well #7 at Cibulakan Formation.

Table 2. Correlation Between the Actual and Predicted Value for Each Regression Method in Every Formation

Formation	Correlation		
	GPR	SVM	Linear
Cibulakan well #6	0.829787	0.63817	0.850902
Cibulakan well #7	0.779752	0.641468	0.764006
Parigi well #6	0.81016	0.807826	0.786406
Pre Parigi well #6	0.817512	0.815208	0.814958

Table 3. RMSE Between the Actual and Predicted Value for Each Regression Method in Every Formation

Formation	RMSE		
	GPR	SVM	Linear
Cibulakan well #6	0.025895	0.036686	0.024804
Cibulakan well #7	0.031178	0.04002	0.032386
Parigi well #6	0.028199	0.028072	0.032635
Pre Parigi well #6	0.04152	0.026366	0.046271

Formation. Similar to the results obtained from Cibulakan Formation (Figures 15, 16, and 17), the results from pre-Parigi and Parigi Formations using GPR are slightly better than the other two methods, both in the correlation (0.817 and 0.810) and RMSE value (0.0017 and 0.0008).

In general, GPR consistently produces considerably better performance compared to the other methods in any formation interval, both qualitatively and quantitatively. It is assumed that this was caused by the ability of GPR to calculate the distribution of each single data for

the value estimation. Nonparametric characteristics of GPR also have a big role in adjusting to the optimum trend of the target variable, since this method does not have an attachment to a particular form of function. The complete recapitulation of the quantitative comparison of each method in every formation is summarized in Tables 2 and 3.

Another useful property of the GPR method is the intrinsic ability to predict the corresponding uncertainties, as can be seen in Figures 20 and 21.

Based on the confidence interval plotting result, it can be seen that each target variable has its own standard deviation. Table 4 shows the result of minimum and maximum standard deviations.

From Table 4, it can be seen that there is a difference between the standard deviation results of each formation. The Cibulakan Formation tends to have a smaller standard deviation compared to Parigi and pre-Parigi Formations. This indicates that the uncertainty in the Cibulakan Formation is lower than the other formations.

### Cross Validation

After getting the correlation and RMSE results from each method, then the cross validation process was executed. Cross validation is performed to validate whether the predicted results of the three regression methods vary significantly



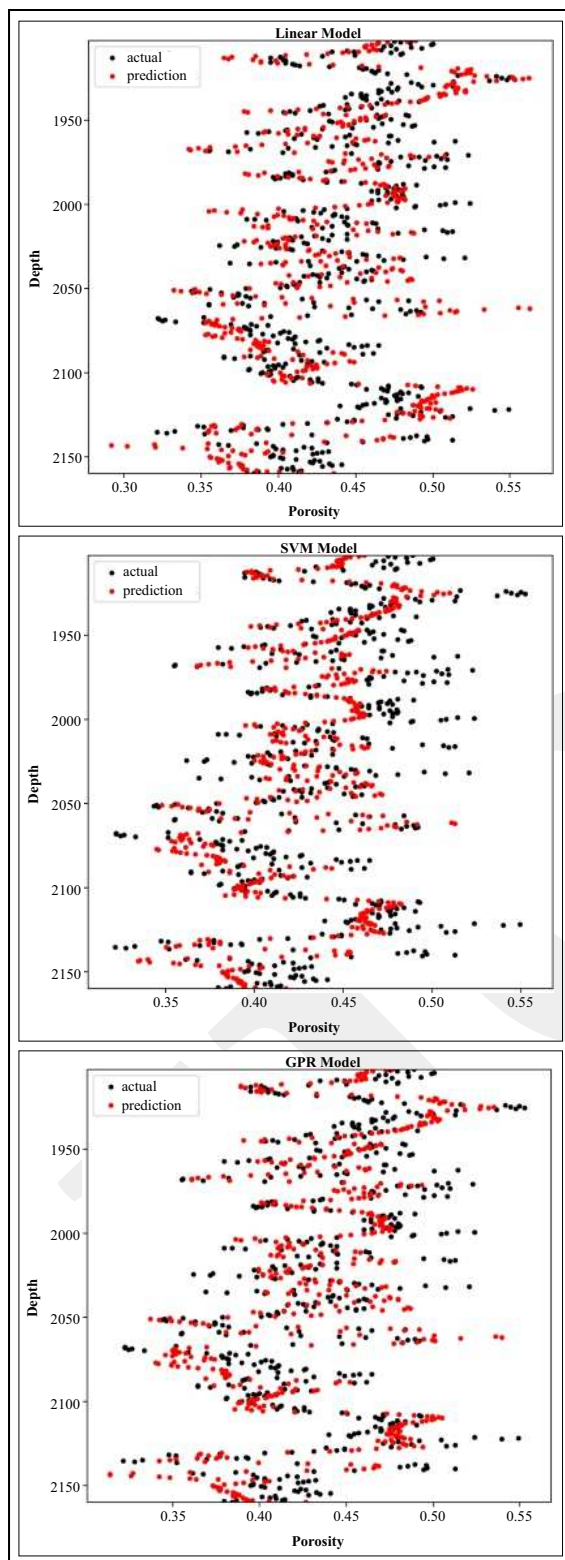


Figure 18. Result of linear regression, SVM, and GPR at interval Parigi Formation on Well #6.

or not when applied to different wells. In this case, a cross validation of all well data was conducted

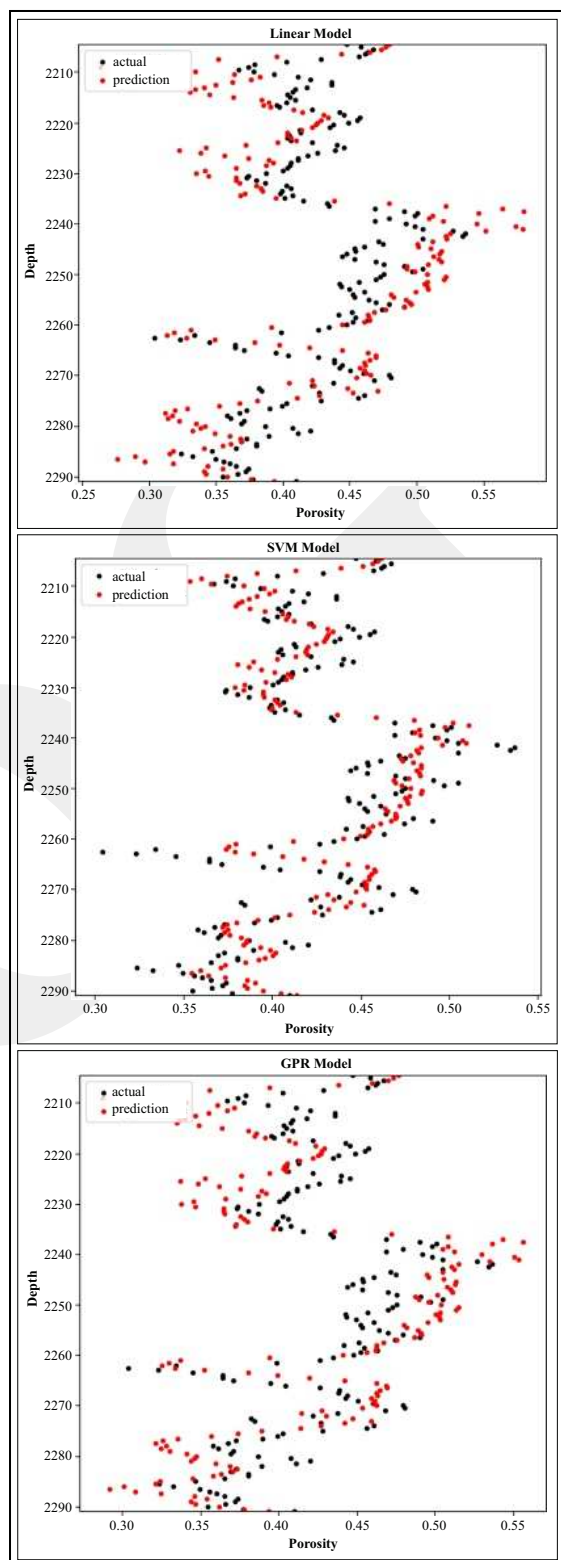


Figure 19. Result of linear regression, SVM, and GPR at interval pre-Parigi Formation on Well #6.

with ten fold validations. Tables 4 and 5 below are the complete cross validation recapitulation

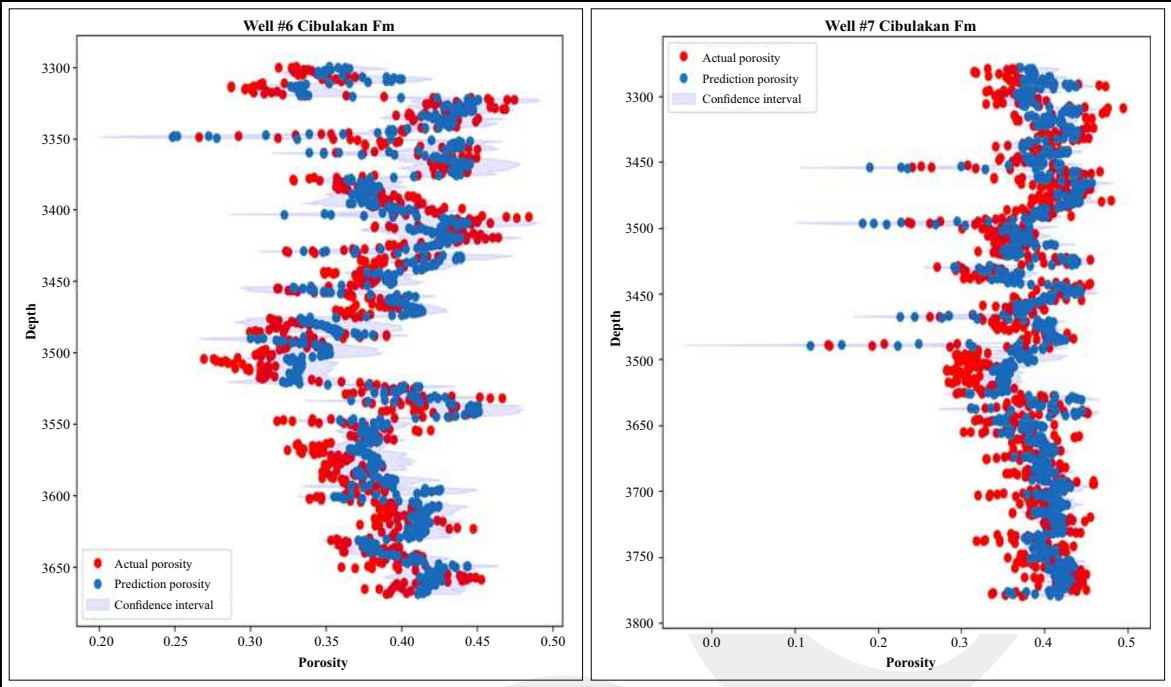


Figure 20. GPR confidence interval plotting overlayed by actual and prediction porosity in Well #6 (left) and Well #7 (right) Cibulakan Formation.

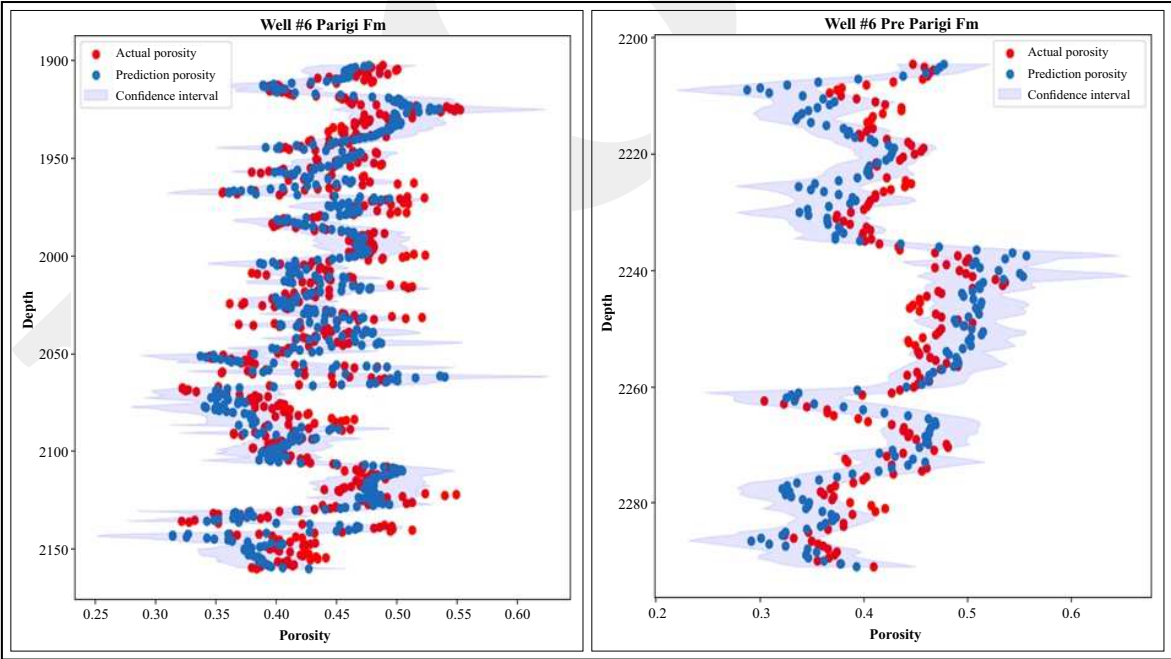


Figure 21. GPR confidence interval plotting overlayed by actual and prediction porosity in Parigi Formation (left) and pre-Parigi Formation (right).

of the quantitative comparison of each method in every formation.

Comparing the results of RMSE between the initial model (Table 3) and cross validation result

(Table 5), it appears that the two results have an identical value. It can be concluded that the model is robust for predicting the target variable in every well.



Table 4. Minimum and Maximum Value of Standard Deviation Results

Standard Deviation		
Formation	Min	Max
Cibulakan Well #6	0.006058	0.02593
Cibulakan Well #7	0.006088	0.077096
Parigi Well #6	0.011651	0.047845
Pre-Parigi Well #6	0.014178	0.05199

Table 5. Cross Validation RMSE Results

RMSE (Cross Validation)			
Formation	GPR	SVM	Linear
Cibulakan	0.029061	0.034201	0.029326
Parigi	0.031962	0.038243	0.030526
Pre-Parigi	0.025868	0.041596	0.025998

## CONCLUSIONS

The eigenvector from PCA of the feature predictor is highly dependent on the lithology variation in each formation, since the PC1 eigenvector value has a significant difference between Cibulakan Formation and Parigi/pre-Parigi Formation.

Regression methods could be a practical option in predicting the missing well log issue faced in the industry. According to the results of this study, high correlation prediction results from two test wells (Well #6 and Well #7) were produced by implementing these three regression algorithms.

GPR is consistently producing better results compared to the other regression methods. GPR also provides the uncertainty of each target variable.

The models created in this study are considerably robust and reliable on predicting the missing log at any well. This is also proven after applying cross validation technique on the models.

## ACKNOWLEDGMENT

The authors would like to express a special gratitude to Sonny Winardhi, Ph.D. from Institut Teknologi Bandung for granting the permission to access the data for this research paper.

## REFERENCES

- Abdelaziz, M., Lastra, R., and Xiao, J.J., 2017. ESP Data Analytics: Predicting Failures for Improved Production Performance. *Society of Petroleum Engineers*. DOI: 10.2118/188513-MS
- Bader, S., Wu, X., and Fomel, S., 2018. *Missing Well Log Estimation by Multiple Well-log Correlation*. DOI: 10.3997/2214-4609.201800989
- Bishop, M.G., 2000. Petroleum systems of the northwest Java Province, Java and offshore southeast Sumatra, Indonesia. *USGS Open-file report 99-50R*.
- Butterworth, P.J., Purantoro, R., and Kaldi, J.G., 1995. Sequence stratigraphic interpretations based on conventional core data: an example from the Miocene upper Cibulakan Formation, offshore Northwest Java. In: Caughey, C.A., Carter, D.C., Clure, J., Gresko, M.J., Lowry, P., Park, R.K., and Wonders, A., (eds.), *Proceedings of International Symposium on Sequence Stratigraphy in Southeast Asia, Indonesia Petroleum Association*, p.311-326.
- Douglas, C.M., Peck, E.A., and Vining, G.G., 2012. *Introduction to Linear Regression Analysis Fifth Edition*. John Wiley & Sons Inc.
- Doust, H. and Noble, R., 2008. *Petroleum systems of Indonesia. Marine and Petroleum Geology*, 25, p.103-129. DOI: 10.1016/j.marpetgeo.2007.05.007.
- Kanfar, R., Shaikh, O., Yousefzadeh, M., and Mukerji, T., 2020. *Real-Time Well Log Prediction From Drilling Data Using Deep Learning*. DOI:10.2523/IPTC-19693-MS.
- Kecman, V., 2005. *Support Vector Machines – An Introduction*. DOI: 10.1007/10984697\_1.
- Lever, J., Krzywinski, M., and Altman, N., 2017. Principal component analysis. *Nature Methods*, 14, p.641-642. DOI: 10.1038/nmeth.4346.
- Noble, R.A., Pratomo, K.H., Nugrahanto, K., Ibrahim, A.M.T., Prasetya, I., Mujahidin, N., Wu, C.H., and Howes, J.V.C., 1997. Petroleum systems of Northwest Java, Indonesia. In:

- Howes, J.V.C. and Noble, R.A. (eds.), *Proceedings of an International Conference on Petroleum Systems of SE Asia & Australasia: Indonesian Petroleum Association*, p.585-600.
- Parapuram, G., Mokhtari, M., and Ben Hmida, J., 2018. An Artificially Intelligent Technique to Generate Synthetic Geomechanical Well Logs for the Bakken Formation. *Energies*, 11 (3), 680pp. DOI: 10.3390/en11030680.
- Pertamina BPPKA, 1996. *Petroleum Geology of Indonesian Basins; Principles, Methods and Application, Volume III*, West Java Sea Basins.
- Rasmussen, C.E. and Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. The MIT Press. ISBN 026218253X.
- Roden, R., Smith, T., and Sacrey, D., 2015. Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation*, 3 (4). DOI: 10.1190/INT-2015-0037.1.
- Rolon, L., Mohaghegh, S.D., Ameri, S., Gaskari, R., and McDaniel, B., 2009. Using artificial neural networks to generate synthetic well logs. *Journal of Natural Gas Science and Engineering*, 1 (4-5), p.118-133. DOI: 10.1016/j.jngse.2009.08.003
- Salehi, M.M., Rahmati, M., Karimnezhad, M., and Omidvar, P., 2017. Estimation of the nonrecords logs from existing logs using artificial neural networks. *Egyptian Journal of Petroleum*, 26 (4), p.957-968. DOI: 10.1016/j.ejpe.2016.11.002
- Syarif, I., Prugel-Bennett, A., and Wills, G., 2016. SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance. *TELKOMNIKA*, 14 (4), p.1502-1509. DOI: 10.12928/telkomnika.v14i4.3956.
- Talpur, A., 2017. *Congestion Detection in Software Defined Networks using Machine Learning of Ali Murad Talpur*. DOI: 10.13140/RG.2.2.14985.85600.
- Wallisch, P., 2014. *Principal Components Analysis*. MATLAB for Neuroscientists, p.305-315. DOI: 10.1016/b978-0-12-383836-0.00017-5