Available online: http://journal.uir.ac.id/index.php/JEEE/index

Journal of Earth Energy Engineering

Publisher: Universitas Islam Riau (UIR) Press

Application of PSO-LSSVM in Prediction and Analysis of Slow Drilling (Rate of Penetration)

Geovanny Branchiny Imasuly[1], Wilma Latuny[2*], Robert Hutagalung[3]

[1]Department of Petroleum Engineering, Faculty of Engineering, Pattimura University, Jl. Ir. M. Putuhena, Poka, Kec. Tlk. Ambon, Kota Ambon, Maluku,97233

[2]Department of Industrial Engineering & Department of Petroleum Engineering, Faculty of Engineering, Pattimura University, Jl. Ir. M. Putuhena, Poka, Kec. Tlk. Ambon, Kota Ambon, Maluku,97233

[3]Department of Physics & Department of Geology, Faculty of Mathematics Natural Sciences & Faculty of Engineering, Pattimura University, Jl. Ir. M. Putuhena, Poka, Kec. Tlk. Ambon, Kota Ambon, Maluku,97233

*Corresponding Author: wlatuny@gmail.com

Article History:

Received: August 2, 2023
Receive in Revised Form: September 30, 2023
Accepted: November 11, 2023

Keywords:

ROP Prediction, PSO-LSSVM, Drilling Parameters.

Abstract

The application of artificial intelligence to predict the accuracy of the rate of penetration (ROP) is very important in optimizing drilling parameters and increasing ROP during drilling. Slow drilling refers to a rate of penetration (ROP), not at the desired level. Drilling an oil well is usually expensive, but this cost can be reduced by carrying out optimal operations. In this paper, the model used is PSO-LSSVM to predict penetration rate. This requires drilling data sequentially and predicting ROP continuously and has a higher success rate in predicting ROP, especially Hole Depth, weight on bit (WOB), Bit Rotation per minute (RPM), Torque, Hook Load, and Standpipe Pressure. The trained data was collected from one drilled oil well, and 7,553 data were used to create the model and were divided into 70% trainset and 30% test-set. The results show that the PSO-LSSVM model has a high level of accuracy in predicting drilling penetration rates. The statistical evaluation shows that the developed PSO-LSSVM model has a high level of accuracy (MAE = 20.10, MSE = 757.9, RMSE= 27.53, and R2 = 0.83) and also in this study PSO which is used in conjunction with SVM in the ROP prediction, the optimum value (best pos): -0.0429913, 0.00350291, and the optimum value (best cost): 0.00186. The results show that the LSSVM optimized with PSO has stronger search and convergence capabilities and higher prediction accuracy for ROP prediction in well X.

INTRODUCTION

One of the main goals of drilling is to reduce total time, minimize risks, save costs, and increase efficiency (planning and exploration stages) (Sobhi et al., 2022a). Slow drilling refers to the rate of penetration (ROP), which is not at the desired and attractive level in drilling operations. ROP characterizes the speed at which the drill bit penetrates the underlying rock to deepen the drill hole because it directly controls the speed and efficiency of drilling, ultimately impacting development costs (Hossain., 2018). Also, ROP shows how fast the drill bit drills through the formation and captures the speed or movement of the drill bit when breaking rock, and in field units, is known as ft/hour (Elkatatny et al., 2020).

In the oil and gas industry, it is known that most well costs come from drilling operations. Therefore, drilling carefully to improve the drilling process is very important. However, because most drilling parameters are interdependent and influence ROP, it is not easy to know the impact of each parameter. These parameters are classified into five types: formation properties, drilling fluid properties, hydraulic parameters, mechanical parameters, and rig efficiency (Elkatatny, 2018). These five categories can also be divided into two main factors: environmental factors and controllable factors (Hossain & Al-Majed, 2015). Environmental factors, such as the nature of the formation, and Controllable factors can be changed, such as RPM, WOB, and hydraulics (Elkatatny et al., 2019).

The oil and gas industry are undergoing a revolution in automation and digital transformation (Balaji et al., 2018; Pandey et al., 2020). Artificial Intelligence has become an application in various aspects of the oil and gas industry, from exploration to production and reservoir management (Bizhani & Kuru, 2022). Therefore, much effort has been made to predict and optimize using Machine learning models to improve ROP predictions' efficiency and accuracy (Ji et al., 2023).

Machine Learning (ML) can help solve complex problems with the maximum possible efficiency.ML can be used as a case-based reasoning tool (Shah et al., 2022). The PSO algorithm is used to optimize superparameters in building the LSSVM model to predict ROP, besides using statistical analysis methods.

DOI: 10.25299/jeee.2023.14004
eISSN: 2540-9352 pISSN: 2301-8097

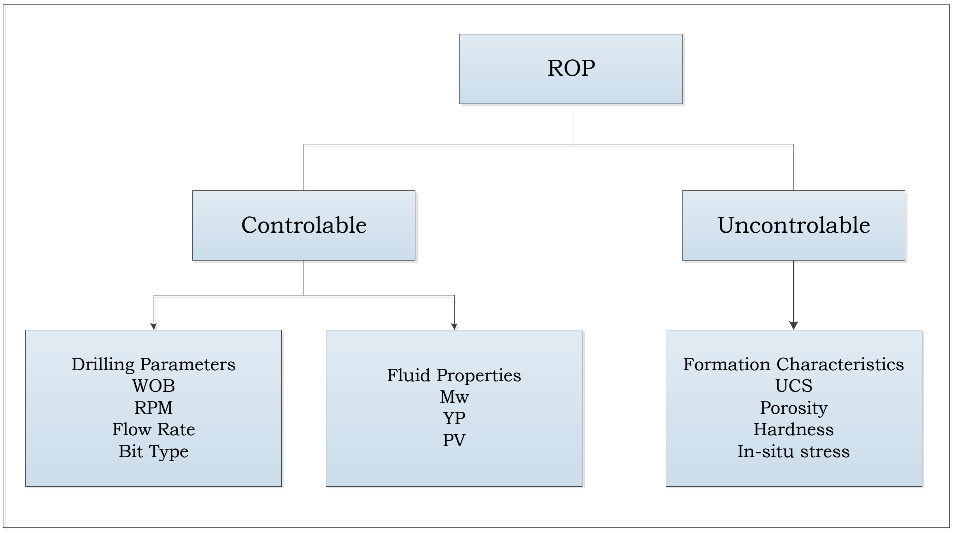

Figure 1. The main factors affecting ROP (Riazi et al., 2022).

Statistical analysis methods are used to evaluate model performance, generally done by comparing model predictions with experimental values. There are several methods in statistical analysis, namely Mean absolute error (MAE), Mean squared error (MSE), Root mean squared error (RMSE), and correlation coefficient (R2) (Sobhi et al., 2022b). The most critical problem in the oil and gas industry, in particular, is drilling time, which can increase drilling costs. (Riazi et al., 2022), Moreover, to estimate drilling rates based on variables such as formation characteristics, operating parameters, and other factors that may influence ROP.

METHODS

This will be done using the input dataset HD, WOB, RPM, Torque, Hook Load, SPP, ROP, and all drilling data provided to maximize the accuracy and reliability of the model, and variables are selected according to their importance to ROP. The predictive model used for ROP prediction must be designed according to the role and function of the model in prediction using the PSO-LSSVM algorithm to find optimal relationships. The simulated sample data is evaluated using the train set and test set to find a match between the predictions and the actual. Then, look at the observed errors to find matching dependencies.

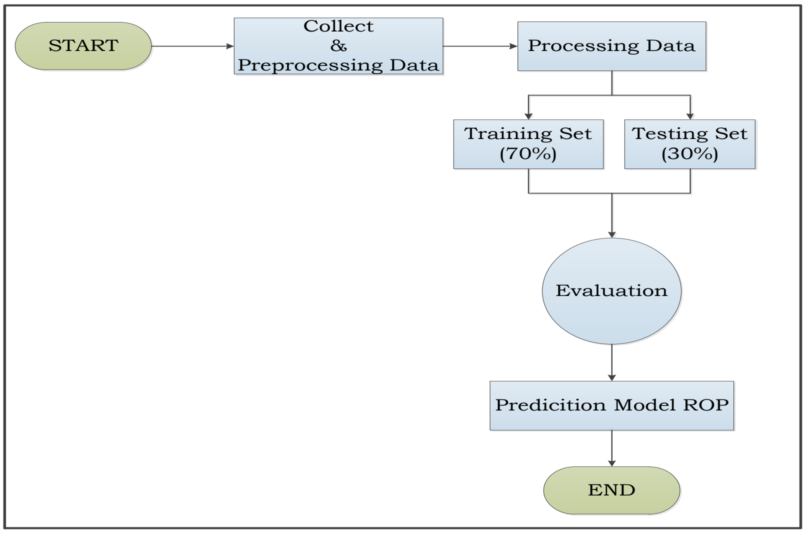

Figure 2. Method Flowchart ROP prediction.

The process for the ROP prediction flow chart is as follows:

1. Collect and explore data sets: raw data from wells covering multiple explored drilling parameters to analyze the properties of attributes of interest when tested for research purposes. Seven drilling parameters were measured, and ROP data was finally selected, as described in the previous section. Then, a data quality check is performed, and simple activity logic is applied to ensure that only drilling data is used. Noise and spurious data in the dataset are manually filtered out by removing data that is out of range. The statistical properties of data in various forms, such as standard deviation, mean, median, etc., as seen in Figures 3 and 4, are taken before training the learning model. Statistical analysis helps to reveal specific characteristics of a data set, and one such important characteristic is standard deviation, which reveals that the data set varies

Journal of Earth Energy Engineering
Vol. 12 No. 3, December 2023, pp 121-128

significantly as a result of different units of measurement of parameters, and thus, data normalization is performed as part of pre-processing. This places various data in the same range to align the distribution and prevent the model from being biased towards large values in the data set (Bodaghi et al., 2015). The statistics also show some low or high values in the predictor, which unconsolidated formations in the drilling interval may cause. Therefore, an outlier removal algorithm is embedded at this stage to remove other spurious data that may be missed during manual data inspection, impacting model training.

2. Data Processing: splitting and implementing algorithms are carried out to ensure uniform data distribution and eliminate the effects of biased sampling; normalized data is then randomized before being used in model development. The data from the two wells was randomly divided into 70% for the training set and 30% for the testing set on which the algorithm was trained, each modified to produce an acceptable model. The LSSVM model uses the RBF kernel and a stochastic optimization algorithm with Particle Swarm Optimization (PSO).
3. Evaluation of models: train and test datasets are used to evaluate models using five criteria, namely, MAE, MSE, RMSE, R^2 and R for ROP prediction.

RESULTS AND DISCUSSION

Data Processing

Several steps need to be taken before the data can be used for processing purposes, in which case the data must be cleaned to remove inconsistencies noisy data and compensate for missing data before use. Data that needs to be cleaned includes data entered into data storage because there were errors or outliers.

This study uses data from one drilled oil well and 7,553 data sets consisting of seven drilling operation parameters, including Hole Depth (ft), Weight on Bit (klbf), Rotation per Minute (RPM), Torque (lbft), Hook Load (lbf), Standpipe Pressure (psig), Rate of Penetration (ft/h). Figure 3 summarizes the basic information of these parameters, including units, average value (*mean*), standard deviation (*std*), minimum value (*min*), and maximum value (*max*). The data is summarized after going through the stage of *processing* that has been done.

	count	mean	std	min	25%	50%	75%	max
HD	7553.0	3854.047134	2204.094281	41.000	1945.00	3853.00	5762.00	7671.00
HL	7553.0	53.121038	24.386298	3.300	32.20	49.20	74.90	120.80
SPP	7553.0	370.377995	62.155159	108.000	341.00	377.00	401.00	721.00
RPM	7553.0	117.741427	477.088629	0.000	55.00	57.00	59.00	4993.00
Torque	7553.0	2582.512330	955.995579	18.338	2147.32	2619.85	3216.26	9793.86
WoB	7553.0	3.050616	1.786359	0.400	2.30	2.90	3.50	98.50
ROP	7553.0	195.789829	66.442764	0.960	148.97	195.26	238.19	631.49

Figure 3. Input and output data parameter statistics.

Next, the data results for each existing parameter are displayed in descriptive analysis via a histogram diagram.

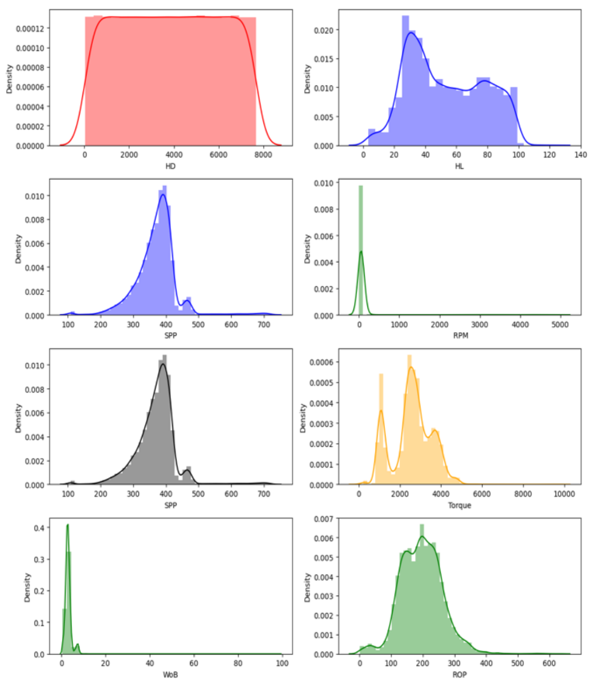

Figure 4. Histogram parameter data input and output.

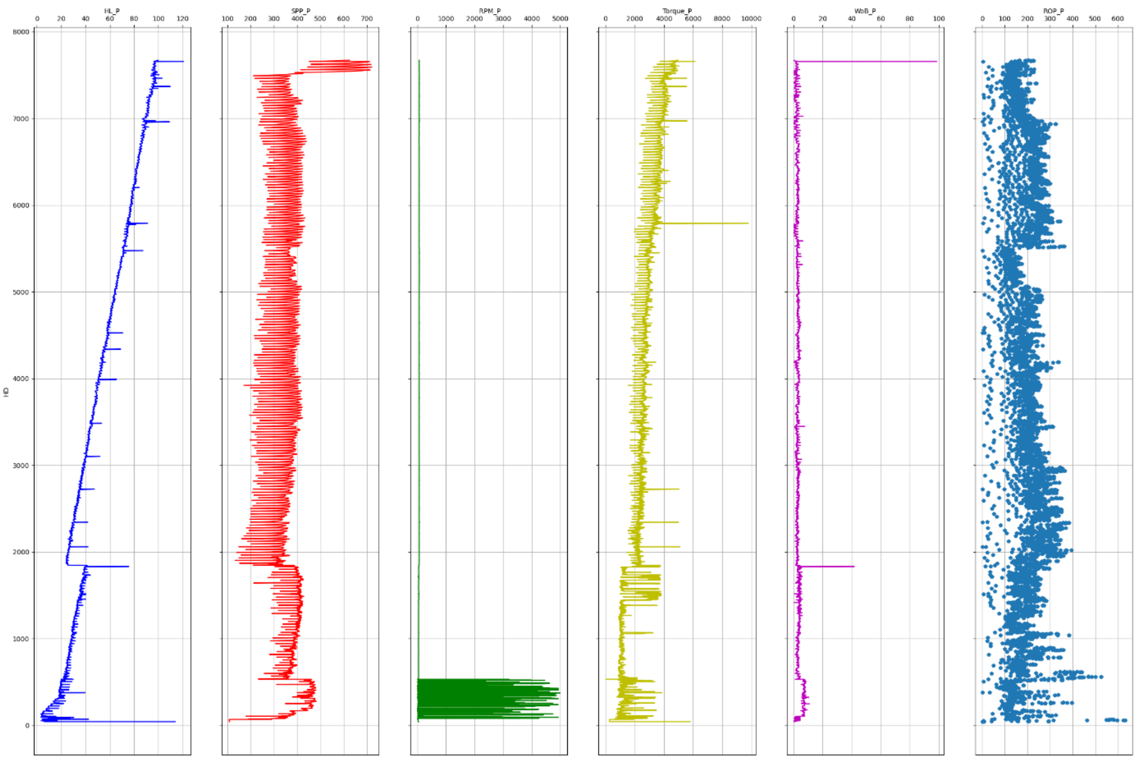

Figure 5. Relation between Hole Depth and HL, SPP, RPM, Torque, WOB, ROP.

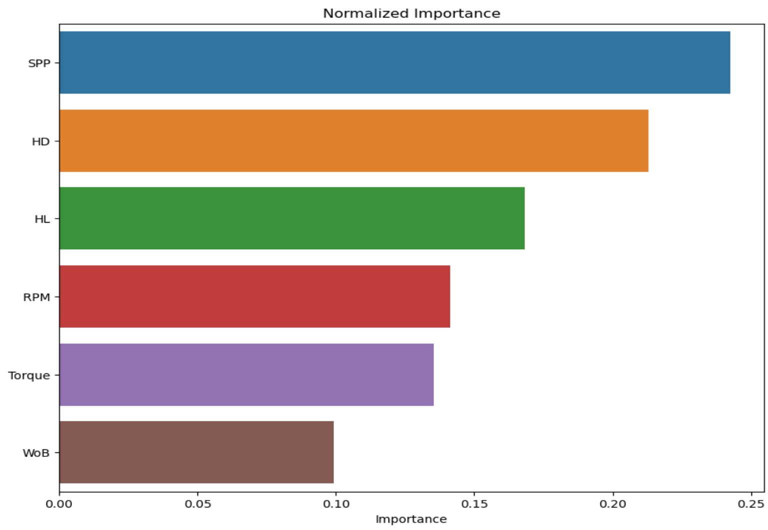

Figure 6. Parameter Importance Normalization.

The selection of variables for all parameters is evaluated by looking for correlation and normalized to determine their importance to ROP. Therefore, SPP, HD, HL, RPM, Torque, and WOB are the most critical factors of ROP. As seen in Figure 6, six parameters are critical and thus significantly contribute to the ROP results. There are significant differences between the six parameters; namely, SPP is the pressure response to changes in ROP. HD measures the depth of the hole that has been drilled, which impacts the total drilling time. HL is the effect of pressure on the bit and can indicate problems such as jamming or obstacles in the formation. RPM is the rotation speed; the higher the RPM, the faster the bit drills and increases drilling efficiency. Torque refers to the rotating force applied to the drill string, causing it to rotate, and WOB is the load applied to the bit by the drill collar for ROP. Its relationship with ROP prediction is to plan drilling operations better and more efficiently. So, in ROP prediction, the application uses PSO-LSSVM.

Least Squares Support Vector Machine (LSSVM) is a machine learning method for predicting the Rate of Penetration (ROP) compared to other methods because LSSVM has advantages when facing problems with limited data. After all, it tends to be more resistant to overfitting. LSSVM tends to generalize well compared to other methods, meaning it can make good predictions on new data that has never been seen before. This is an essential aspect of ROP prediction because geological conditions and drilling operations can vary from one well to another. When using Particle Swarm Optimization (PSO) to optimize parameters. The advantage of PSO over other optimization methods, such as Bayesian optimization, random forest, and grid search, is that PSO generally explores

Journal of Earth Energy Engineering
Vol. 12 No. 3, December 2023, pp 121-128

the parameter space more quickly and tends to find better solutions with fewer iterations. It helps to find the best parameters that fit the training data and improves accuracy. Predictions.

Applications of PSO-LSSVM

Predicting ROP with artificial intelligence is more flexible and can solve complex problems. In this research, we developed a model using PSO-LSSVM. First, the data is divided randomly into two subsets called the train set, namely 70%, where the model learns and tries to find the best and optimal prediction model. The test set, namely 30%, is used to investigate predictions.

Table 1. Optimal value Gaussian RBF kernel in PSO-LSSVM.

Parameters	Test Accuracy
Variansi dari Gaussian RBF kernel	83.19%

Statistical evaluation

Statistical analysis was carried out to evaluate and compare the models developed in this study. The goal is to calculate mean Average Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and coefficient of Determination (R2), and the results are summarized in Table 1 from Equations (1) – (4) which present the formulation used to calculate these parameters:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|ROP_i - ROP_{ipredicted}\right| \qquad (1)$$

Predict ROP using MAE to express the average forecast error in the same units as the target (e.g., *feet* per hour or meters per hour). The lower the MAE value, the higher the prediction quality because it shows that the overall ROP prediction is close to the actual value. Therefore, information about the overall prediction error rate can be obtained when predicting ROP.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left|ROP_i - ROP_{ipredicted}\right|^2 \qquad (2)$$

In ROP prediction, using MSE to describe the average of the squared prediction errors in the same units as the target is the same as MAE. The lower the MSE value, the higher the prediction quality. This shows that the overall ROP prediction is close to the actual value. To acquire information about the overall degree of error in ROP predictions by prioritizing more significant errors

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left|ROP_i - ROP_{ipredicted}\right|^2}{N}} \qquad (3)$$

RMSE measures the root mean square of the difference between the predicted and actual values in the same units as the target. RMSE measures the prediction error relative to the square root of the scaled data. A smaller RMSE value indicates better prediction quality because the ROP prediction is more accurate and closer to the actual value. RMSE provides information about the overall ROP prediction error rate, with higher errors indicating lower accuracy. Additionally, because RMSE shares units with targets, it is easy to understand and compare. Its more straightforward interpretation makes it a simpler metric to understand.

$$R^2 = 1 - \frac{\sum_{l=1}^{Num}(ROP_{exp,l} - ROP_{pred,l})^2}{\sum_{l=1}^{Num}(ROP_{pred,l} - \overline{ROP})^2} \qquad (4)$$

R^2 (Coefficient of Determination) is an evaluation metric used in ROP prediction to benchmark the ROP prediction model against actual data. R-value between 0 and 1, where 0 indicates that the model provides no explanation for the variability in the data, and one indicates that the model provides a perfect explanation(perfect) for data variability. R^2 measures the proportion of variation in the target variable that the model can explain.

The predicted value results show that the greater the R^2 (approaching 1), the greater RMSE, MAE, and MSE, so the prediction model's performance improves. The model performance evaluation index before and after optimization is also shown in Table 2.

Table 2. Comparison of Model Performance Evaluation Metrics.

	MAE	MSE	RMSE	R^2	R
ROP Train	20.10	757.9	27.53	0.83	0.91
ROP Test	25.35	1329.3	36.46	0.70	0.84

R^2, a higher value, indicates that the model can better explain variations in the data. However, remember that R2 does not provide information about the overall quality of predictions but focuses on how well the model can explain variations in the data. Therefore, R^2 should be used with metrics such as MAE, MSE, and RMSE to understand ROP prediction performance better.

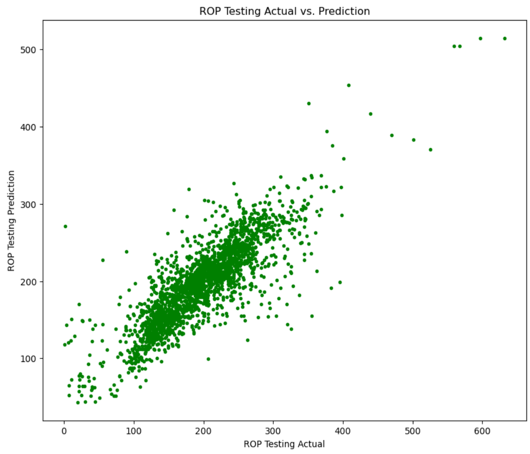

Figure 7. Plot ROP testing actual vs prediction.

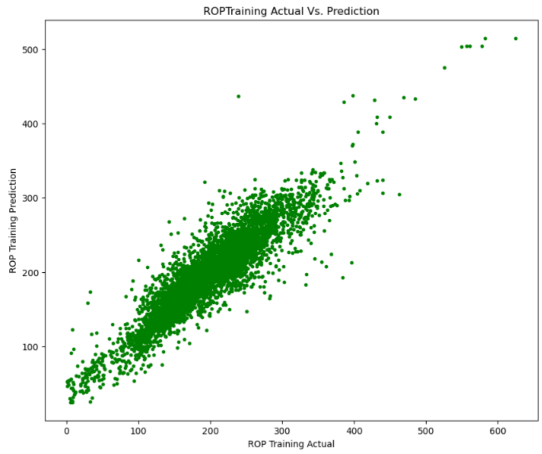

Figure 8. Plot ROP Training actual vs prediction.

The PSO process is initialized with a population of random particles, and the algorithm searches for optimal solutions by updating generation. c1 and c2 are acceleration constant, and inertia weight w controls the effect of the previous particle's speed on the current particle. The range for the number of particles (population size) is 20-40, but 10 particles will get better results. The inertial weight scale decreases linearly from 0.9 to 0.4 during the search process to effectively balance the swarm's local and global search capabilities.

Table 3. Parameters in PSO-LSSVM.

Parameter	Value
Population size	10
Acceleration c_1	2
Acceleration c_2	2
Inertia weight w	0.9
Number of iteration	100

Examples of trajectories and performance criteria during optimization are shown in Figures 9 and 10, each parameter obtained for 100 iterations.

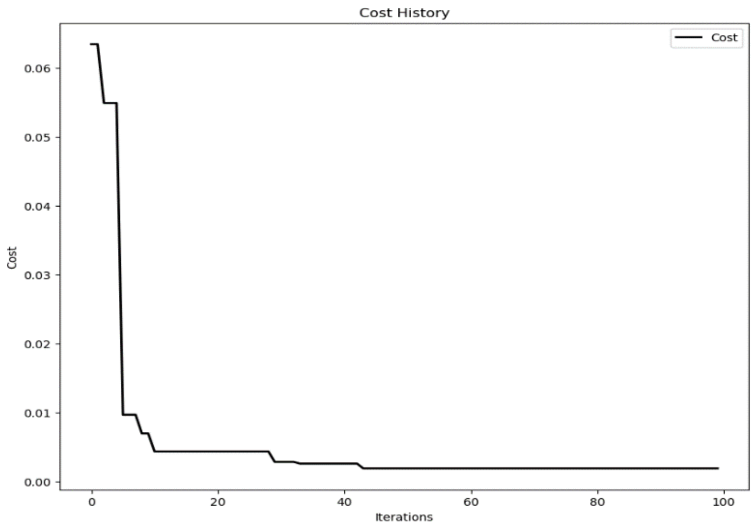

Figure 9. Iterations, when used to train the LSSVM model.

The inner iteration goal of PSO with LSSVM is to find the set of parameters that produces the model LSSVM by providing the most accurate ROP predictions. During the iteration, the particle moves through the search space, trying different combinations of parameters, and gradually approaches the optimal solution. The optimal number of iterations may differ based on the complexity of the problem. This process is carried out until it reaches a predetermined stopping criterion, such as reaching the maximum number of iterations or the desired level of

Journal of Earth Energy Engineering
Vol. 12 No. 3, December 2023, pp 121-128

convergence. Several factors can influence iteration and convergence performance, including the size of the data set used, particle population size, and learning factors.

In context, Particle Swarm Optimization (PSO), which is used in conjunction with LSSVM (Least Squares Support Vector Machine) in ROP prediction with "iterations," refers to the number of cycles or steps executed by the algorithm *PSO* to find the optimal solution. Each iteration involves updating the position and velocity of the particles in the search space with the optimum value (best pos) and optimum value (best cost), which in this research obtained the optimum value (best pos): -0.0429913, 0.00350291, and the optimum value (best cost): 0.00186. The optimum value (best cost) is found by the PSO algorithm during the iteration process, which is usually measured from the error between the ROP predictions produced by the LSSVM model.

The goal is to find the LSSVM model parameters that produce the lowest "best cost." Thus, the PSO algorithm finds the best solution is "best post," which refers to producing the optimal or best post values for the LSSVM model parameters to predict ROP with high accuracy.

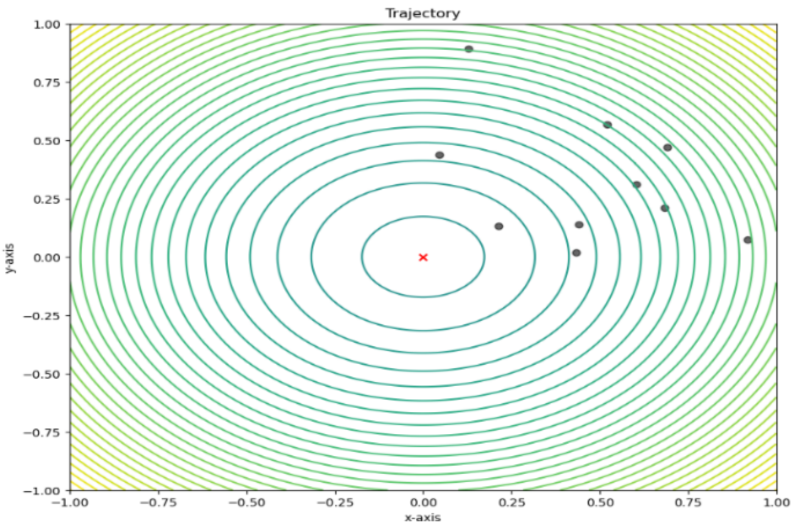

Figure 10. PSO iterations, when used to train the LSSVM model (2D).

CONCLUSIONS

In this paper, we develop a methodology to predict ROP using the technique of *machine learning*. The model used is PSO-LSSVM. It takes drilling data sequentially, predicts ROP continuously, and achieves better ROP prediction results. In this case, Hole Depth (ft), weight on bit (klbf), Rotation per minute (RPM), Torque (lbft), Hook Load (lbf), and Standpipe Pressure (psig) show significant influence in maintaining ROP at a high level. Then, use the PSO-LSSVM prediction model to predict. As we can see, R2=0.83 and R=0.91 in the ROP train, R2=0.70, and R=0.84 in the ROP test, indicating that the prediction accuracy is high because the smaller RMSE, MAE, and MSE indicate a better prediction model and also in this research PSO which is used in conjunction with LSSVM in the ROP prediction, the optimum value (best post): -0.0429913, 0.00350291, and the optimum value (best cost): 0.00186. The optimum value (best cost) is found by the PSO algorithm during the iteration process, which is usually measured from the error between the ROP predictions produced by the LSSVM model. The results show that the LSSVM optimized with PSO has stronger search and convergence capabilities and higher prediction accuracy for ROP prediction in well X.

ACKNOWLEDGEMENTS

First, I thank God Almighty so that this writing can be completed to Pattimura University, Ambon, Department of Geology, Petroleum Engineering Study Program, and National Development University "Veteran", Yogyakarta. And special for the family.

REFERENCES

Balaji, K., Rabiei, M., Suicmez, V., Canbaz, C. H., Agharzeyva, Z., Tek, S., Bulut, U., & Temizel, C. (2018, June 11). Status of Data-Driven Methods and their Applications in Oil and Gas Industry. The SPE Europec Featured at 80th EAGE Conference and Exhibition. https://doi.org/10.2118/190812-MS

Bizhani, M., & Kuru, E. (2022). Towards drilling rate of penetration prediction: Bayesian neural networks for uncertainty quantification. Journal of Petroleum Science and Engineering, 219(September), 111068. https://doi.org/10.1016/j.petrol.2022.111068

Bodaghi, A., Ansari, H. R., & Gholami, M. (2015). Optimized support vector regression for drilling rate of penetration estimation. Open Geosciences, 7(1), 870–879. https://doi.org/10.1515/geo-2015-0054

Elkatatny, S. (2018). New Approach to Optimize the Rate of Penetration Using Artificial Neural Network. Arabian Journal for Science and Engineering, 43(11), 6297–6304. https://doi.org/10.1007/s13369-017-3022-0

Elkatatny, S., Al-Abduljabbar, A., & Abdelgawad, K. (2020). A new model for predicting rate of penetration using an artificial neural network. Sensors (Switzerland), 20(7). https://doi.org/10.3390/s20072058

Hossain, M. E., & Al-Majed, A. A. (2015). Fundamentals of Sustainable Drilling Engineering. Wiley. https://doi.org/10.1002/9781119100300

HOSSAIN, M.E., ISLAM. M. R. (2018). Drilling Engineering Problems and Solution. 627.

Ji, H., Lou, Y., Cheng, S., Xie, Z., & Zhu, L. (2023). An Advanced Long Short-Term Memory (LSTM) Neural Network Method for Predicting Rate of Penetration (ROP). https://doi.org/10.1021/acsomega.2c06308

Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., & Saputelli, L. (2020). Toward Oil and Gas 4.0. In Machine Learning in the Oil and Gas Industry (pp. 1–40). Apress. https://doi.org/10.1007/978-1-4842-6094-4_1

Riazi, M., Mehrjoo, H., Nakhaei, R., Jalalifar, H., Shateri, M., Riazi, M., Ostadhassan, M., & Hemmati-Sarapardeh, A. (2022). Modelling rate of penetration in drilling operations using RBF, MLP, LSSVM, and DT models. Scientific Reports, 12(1), 1–24. https://doi.org/10.1038/s41598-022-14710-z

S, A. A., Elkatatny, S., Ali, A. Z., & Mahmoud, M. (2019). IPTC-19548-MS Rate of Penetration Prediction in Shale Formation Using Fuzzy Logic Fuzzy Logic (FL).

Shah, M., Kshirsagar, A., & Panchal, J. (n.d.). Applications of Artificial Intelligence (AI) and Machine Learning (ML) in the Petroleum Industry.

Sobhi, I., Dobbi, A., & Hachana, O. (2022a). Prediction and analysis of penetration rate in drilling operation using deterministic and metaheuristic optimization methods. Journal of Petroleum Exploration and Production Technology, 12(5), 1341–1352. https://doi.org/10.1007/s13202-021-01394-w

Sobhi, I., Dobbi, A., & Hachana, O. (2022b). Prediction and analysis of penetration rate in drilling operation using deterministic and metaheuristic optimization methods. Journal of Petroleum Exploration and Production Technology, 12(5), 1341–1352. https://doi.org/10.1007/s13202-021-01394-w