



Prediction of Biochemical Oxygen Demand and Chemical Oxygen Demand Using Artificial Neural Network and Long Short-Term Memory

Sekarlangit^{1*} Catur Edi Widodo² Tarno³

¹*Master Program of Information System, School of Postgraduate Studies, Universitas Diponegoro, Indonesia*

²*Department of Physics, Faculty of Science and Mathematics, Universitas Diponegoro, Indonesia*

³*Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, Indonesia*

Article details:

Published: 30th 4 2024



Abstract

River water quality is a significant factor in environmental sustainability and public health. The identification and monitoring of major pollutant indices such as chemical oxygen demand (COD) and biochemical oxygen demand (BOD) are critical in monitoring river water quality. Several studies have employed artificial neural networks (ANN) and long-short-term memory (LSTM) to predict river water quality based on pollutant data such as COD and BOD; however, their performance has to be evaluated further. The purpose of this study is to assess how well ANN and LSTM predict these parameters in river water. In this study, two machine learning models, ANN

and LSTM, are implemented in Python. The BOD and COD parameters are trained using both models, and to see if the ANN and LSTM models perform well on this data, the first step is to examine the loss model on the ANN and LSTM. The discussion results reveal that both models capture river water quality trends and are consistent when compared to real and forecasted data. Model performance evaluation using mean squared error (MSE) and root mean squared error (RMSE) shows that LSTM outperforms ANN in predicting BOD and COD parameters as river water quality indicators.

Keywords: Artificial Neural Network, Biochemical Oxygen Demand, Chemical Oxygen Demand, Long Short-Term Memory, River Water Quality

To cite this article :

Sekarlangit, Catur Edi Widodo, T. (2024). Prediction of Biochemical Oxygen Demand and Chemical Oxygen Demand Using Artificial Neural Network and Long Short-Term Memory. *International Journal of Health, Education & Social (IJHES)*, 7(3), 81–88.



Introduction

River water quality is the main indicator of environmental sustainability and public health. One important aspect in monitoring river water quality is the identification and monitoring of key pollutant parameters, including Chemical Oxygen Demand (COD) and Biochemical Oxygen Demand (BOD). The COD and BOD are important indicators in determining the level of organic pollutants in water, which can provide an overview of the level of pollution and its potential negative impacts on ecosystems and human health.(Q. Zhou et al., 2023). Typically, monitoring of these parameters is carried out using conventional methods involving periodic water sampling at specific locations and time-consuming laboratory analysis. However, these methods are often inefficient in providing real-time information about river water quality conditions, which can limit the ability to take rapid preventive or corrective action (Thamrin et al., 2023).

In an effort to improve the prediction and monitoring capabilities of river water quality, Machine Learning approaches have become increasingly popular and relevant. The two types of Machine Learning models used in this research are Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) (Chen et al., 2020). The ANN and LSTM have been proven capable of modelling complex patterns in data, including sequence data such as time series, which is an important characteristic of river water quality data (Salaeh et al., 2022).

Although there have been several studies using ANN and LSTM to predict river water quality based on pollutant parameters such as COD and BOD, a more in-depth evaluation of their performance is still needed (Rustam et al., 2022). A comprehensive evaluation will help in determining the relative superiority of ANN and LSTM in terms of river water quality prediction (Xu et al., 2023). Therefore, this research aims to evaluate the performance of ANN and LSTM in predicting pollutant parameters such as COD and BOD in river water. By comparing the evaluation performance of these two models.

Methods

This research uses machine learning models, which are Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM). The results of the ANN model are compared with the results of the LSTM model to determine the performance of each model (Siami-



Namini et al., 2018). The data used is river water quality monitoring data with BOD and COD parameters from January 2013 to December 2022 obtained from the Environmental Laboratory Unit in Indonesia with a division of 80% as training data and 20% as test data. The dataset can be seen in Table 1 with total data 1559. The ANN model is used to learn underlying patterns in the data, while the LSTM model is used to capture temporal dependencies. The stages from data preprocessing to performance measurement and model performance evaluation can be seen in Figure 1.

Table I. Sample Dataset

Date	BOD	COD
06/01/2013	2,8	14,0
15/01/2013	7,1	11,0
25/01/2013	6,3	14,0
...	...	...
10/12/2022	6,3	10,0
20/12/2022	7,8	7,0
29/12/2022	9,2	7,0

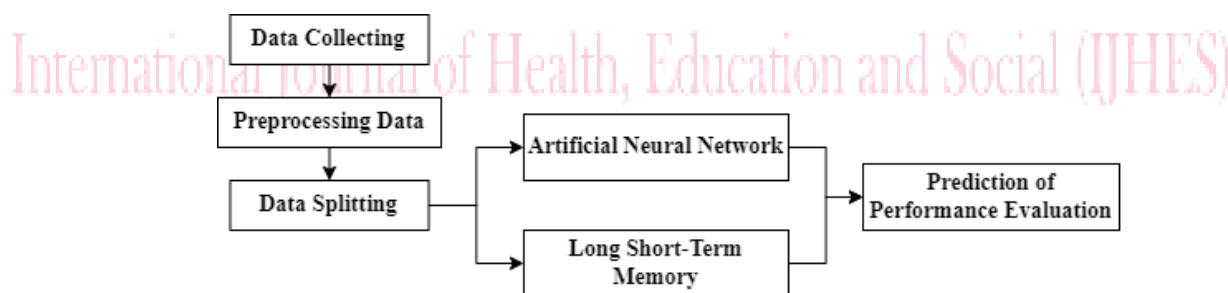


Figure 1. Research Process

The first stage is preprocessing data which consists of data cleaning, data normalization and dividing the data into training data and test data. In the data normalization process, the method used is min-max normalization. The next steps, the data is trained on the model and the final result obtained is an evaluation of the model performance using R-Squared, MSE and RMSE.

R-Squared (R^2) is an evaluation metric used to measure the extent of variation in the target variable by a regression model. R^2 ranges from 0 to 1 with higher values indicating better model



performance. R-Squared formula is defined y_i = actual value, $\hat{y}$ = predicted value, N = number of samples.

$$R - Squared = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

Mean square error (MSE) is a method of calculating error by squaring the error value in each period. If the MSE value is smaller, it indicates the model is more stable (Niknam et al., 2023). MSE formula is defined y_i = actual value, $\hat{y}$ = predicted value, N = number of samples.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (2)$$

Root mean square error (RMSE) is the mean value of the square roots used to evaluate the size of the error. RMSE evaluation is calculated based on the average value of the sum of the squared errors between the forecast results and the actual value. The value is used to measure the magnitude of the error produced by the forecast model. A forecast model approaches the actual value variation if the RMSE calculation results are low (X. Zhou et al., 2024). RMSE formula is defined y_i = actual value, $\hat{y}$ = predicted value, N = number of samples

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (3)$$

International Journal of Health, Education and Social (IJHES)

Results and Discussion

This research uses Python and applies two different machine learning models, which are ANN and LSTM, and each model has different results. The ANN model and LSTM model are trained and get a model loss process is carried out including training loss and validation loss for each BOD and COD parameter. This is shown that the model is running well on the data because the test data and training data continue to decrease. In this process, model loss is also done to avoid overfitting the model. It can be seen in Figure 2-5.

As can be seen in Figure 6-9, from observing the results, it is evident that the model accurately captures trends in river water quality parameters. In the figure, the predicted values of the time series are compared with the actual values in the same period. Evaluation of both the ANN model and the LSTM model shows consistent agreement with the actual data in most of the prediction examples. This shows that the model effectively predicts river water quality with BOD and COD parameters.

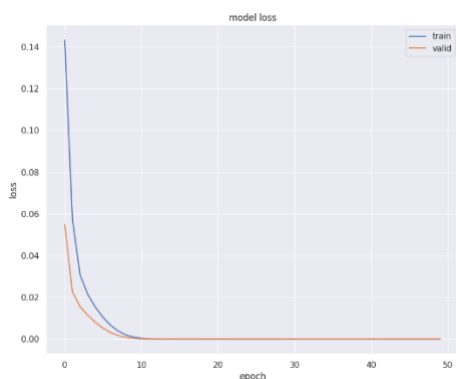


Figure 2. Model Loss ANN BOD

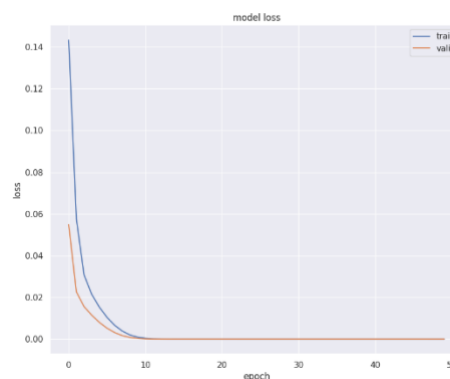


Figure 3. Model Loss LSTM BOD

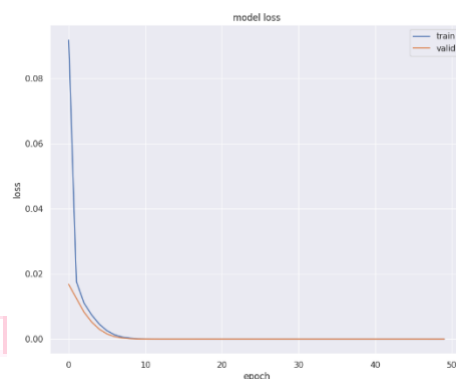


Figure 4. Model Loss ANN COD

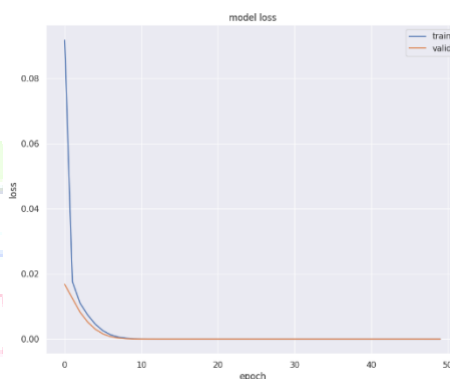


Figure 5. Model Loss LSTM COD

The results of the predicted values and actual values of the BOD and COD parameters in the ANN model and LSTM model are shown in Figure 6-9. The ANN model and LSTM model, and using historical data to train the model.

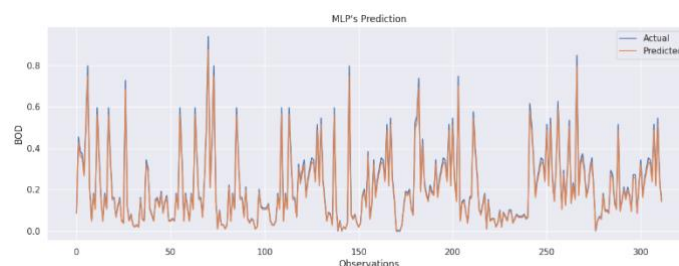


Figure 6. Actual and predicted value BOD of the ANN model

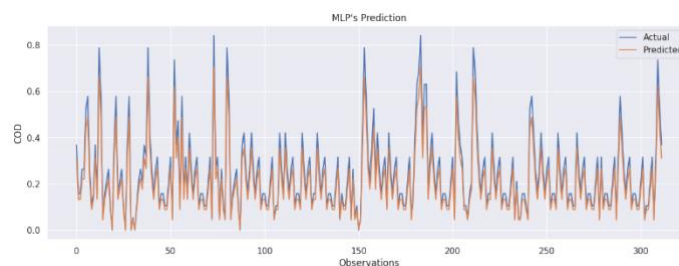


Figure 7 Actual and predicted value COD of the ANN model

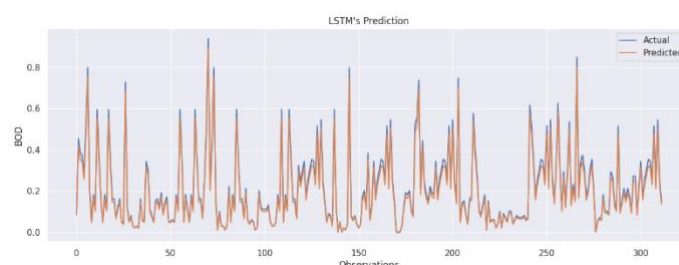


Figure 8 Actual and predicted value BOD of the LSTM model

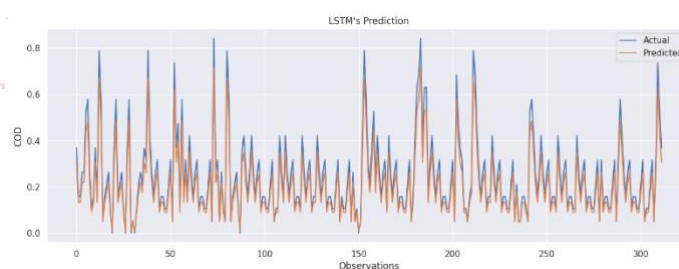


Figure 9 Actual and predicted value COD of the LSTM model

The results of the model evaluation of the BOD and COD parameters in the ANN model and LSTM model are shown in Table 2 and Figure 10-11.

Tabel 2. The Result of a Comparison of R^2 , MSE and RMSE value for ANN and LSTM

Parameter	Model	R^2	MSE	RMSE
BOD	ANN	0,992	0,00027	0,01647
	LSTM	0,986	0,00046	0,02163
COD	ANN	0,923	0,00220	0,04698



	LSTM	0,918	0,00235	0,04854
--	------	-------	---------	---------

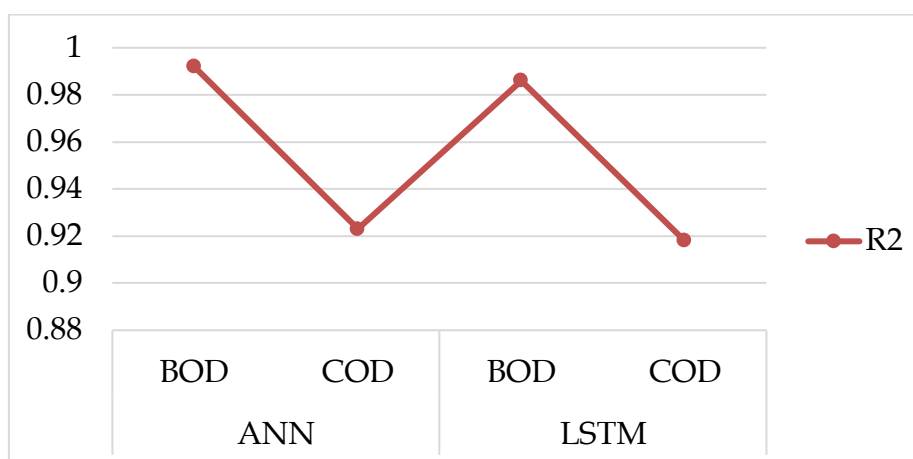


Figure 10. Comparison of R² for ANN and LSTM Model

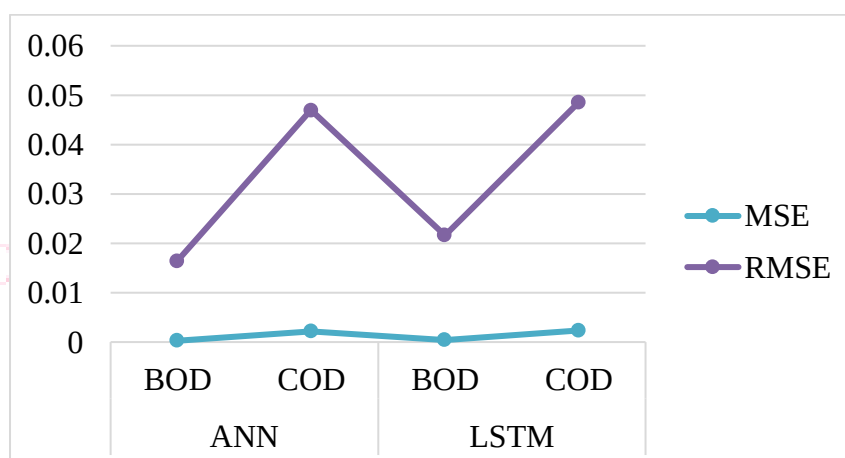


Figure 11. Comparison of MSE and RMSE for ANN and LSTM Model

In this research, the evaluation method uses R², MSE and RMSE model evaluation to present the performance of different prediction models. Based on Table 2, it can be seen that the ANN model has quite good performance in predicting BOD with an MSE value of 0.00027 and an RMSE value of 0.01647 and COD with an MSE value of 0.00220 and an RMSE value of 0.004698. Meanwhile, the LSTM model also has quite good performance in predicting BOD with an MSE value of 0.00046 and an RMSE value of 0.04854.

Both the ANN and LSTM models have high R² values because both models are close to 1 value for the BOD and COD parameters, which indicates that the model has a good fit to



the data. The performance of the ANN model is slightly better than LSTM in predicting BOD, with higher R^2 values, which are ANN 0.992 and LSTM 0.986 and lower MSE and RMSE values. However, the ANN model has a slightly higher R^2 value, which are LSTM 0.918 and ANN 0.923 and lower MSE and RMSE values in predicting COD compared to LSTM model.

Although the ANN model had a higher R^2 value for predicting BOD and COD, than the LSTM model. The result is indicating that the ANN model has lower MSE and RMSE values for predicting BOD and COD than the LSTM model, which indicates that the ANN model has a better overall fit and is more accurate in predicting BOD and COD.

Conclusion and Recommendation

In conclusion, analysis performed the ANN model outperforms the LSTM model in predicting the BOD and COD parameters, as evidenced by the high R^2 values are close to 1 value and the lower MSE and RMSE values. The LSTM model has a higher MSE and RMSE for the COD parameter compared to the BOD parameter, indicating that the COD parameter is more challenging to predict. The ANN model has a lower MSE and RMSE for the BOD parameter, suggesting that the ANN model is better suited for predicting the BOD parameter. Based on performance evaluation the ANN model is more effective in predicting the BOD and COD parameters than the LSTM model. Future research could include more variables related to river water quality to develop a comprehensive prediction framework. This approach will allow the model to provide accurate predictions in diverse environmental conditions, thereby increasing its effectiveness.

References

- Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. In *Applied Sciences (Switzerland)* (Vol. 10, Issue 17). MDPI AG. <https://doi.org/10.3390/app10175776>
- Niknam, A., Zare, H. K., Hosseininiasab, H., & Mostafaeipour, A. (2023). Developing an LSTM model to forecast the monthly water consumption according to the effects of the climatic factors in Yazd, Iran. *Journal of Engineering Research*, 11(1), 100028. <https://doi.org/10.1016/j.jer.2023.100028>
- Rustam, F., Ishaq, A., Kokab, S. T., de la Torre Diez, I., Mazón, J. L. V., Rodríguez, C. L., & Ashraf, I. (2022). An Artificial Neural Network Model for Water Quality and Water



- Consumption Prediction. *Water* (Switzerland), 14(21).
<https://doi.org/10.3390/w14213359>
- Salaeh, N., Dittthakit, P., Pinthong, S., Hasan, M. A., Islam, S., Mohammadi, B., & Linh, N. T. T. (2022). Long-Short Term Memory Technique for Monthly Rainfall Prediction in Thale Sap Songkhla River Basin, Thailand. *Symmetry*, 14(8).
<https://doi.org/10.3390/sym14081599>
- Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1394–1401.
<https://doi.org/10.1109/ICMLA.2018.00227>
- Thamrin, N. M., Jaffar Politeknik Ungku Omar, A., Jaffar, A., Syahirul Amin Megat Ali, M., Farid Misnan, M., Ihsan Mohd Yassin, A., Mohamed Zan, N., & Nor Liyana Nik Ibrahim, N. (2023). *Application of Long-Short Term Memory for Accurate Biochemical Oxygen Demand Prediction in Rivers through Water Quality Parameters* (Vol. 7).
<http://arqiipubl.com/ams>
- Xu, Q., Huang, X., Zhang, B., Zhang, Z., Wang, J., & Wang, S. (2023). TBM performance prediction using LSTM-based hybrid neural network model: Case study of Baimang River tunnel project in Shenzhen, China. *Underground Space (China)*, 11, 130–152.
<https://doi.org/10.1016/j.undsp.2022.11.002>
- Zhou, Q., Hu, X., Wang, D., Zhang, W., Chen, Z., Wang, J., Wang, P., & Ren, X. (2023). Prediction of Dissolved Oxygen in Rivers Based on LSTM Model with Improved Attention Mechanism. *Research of Environmental Sciences*, 36(6), 1135–1146.
<https://doi.org/10.13198/j.issn.1001-6929.2023.02.18>
- Zhou, X., Meng, X., & Li, Z. (2024). ANN-LSTM-A Water Consumption Prediction Based on Attention Mechanism Enhancement. *Energies*, 17(5).
<https://doi.org/10.3390/en17051102>