

ABSTRACT

This research is motivated by the awareness of the growing use of Deep Learning (DL) across various fields. The focus of this study is the prediction of dissolved oxygen (DO) levels and nitrate concentrations in river water. Commonly used methods for predicting water indicators include Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), etc. Among these prediction methods, LSTM is known for its superiority in maintaining long-term information. This method was chosen for the study due to its compatibility with the characteristics of the dataset. Although LSTM has proven effective, there are issues related to low data quality that can reduce prediction accuracy.

To enhance data quality, improvements were made during the preprocessing stage. *Outliers* were removed using the Interquartile Range (IQR), noise was reduced with Moving Average (MA), and missing data were filled using Linear Interpolation (LI). Additionally, the use of the K-means clustering method was employed to further improve data quality. Combining locations with similarity enhanced the learning performance of the model, thereby increasing the accuracy of predicting DO levels and nitrate concentrations.

The results of the study indicate that the approach of grouping data based on data collection locations using K-means was successfully applied under various dataset conditions. When compared to methods without grouping, the proposed method yielded satisfactory accuracy levels. For predicting dissolved oxygen, there was an increase in the R^2 value by 0.0044%, a decrease in MAE by 54.4267%, and a decrease in MSE by 78.8535%. In nitrate prediction, there was an increase in the R^2 value by 61.9218%, a decrease in MAE by 74.5407%, and a decrease in MSE by 91.3194%. These results confirm the significant potential of using K-means to enhance the modeling of DO and nitrate level predictions. Although challenges related to accuracy persist in some locations, the successful application of this strategy demonstrates its effectiveness in predicting DO and nitrate levels over one month, considering the location factor in data collection using K-means.

Keywords: Prediction of DO and Nitrate levels, Preprocessing, Location clustering, K-means, Correlation analysis, Spearman, LSTM

INTISARI

Penelitian ini didasari kesadaran akan penggunaan *Deep Learning* (DL) yang semakin berkembang di berbagai bidang. Fokus penelitian ini adalah untuk prediksi tingkat oksigen terlarut (DO) dan kadar nitrat dalam air sungai. Metode yang biasa digunakan untuk memprediksi indikator air adalah *Artificial Neural Networks* (ANN), *Recurrent Neural Networks* (RNN), *Long Short-Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), dll. Dari banyaknya metode prediksi, LSTM dikenal sebagai metode yang unggul karena dapat mempertahankan informasi jangka panjang. Metode ini dipilih untuk penelitian ini, karena sesuai dengan karakteristik *dataset* yang dimiliki. Walaupun LSTM telah terbukti efektif, terdapat permasalahan yang melibatkan kualitas data rendah sehingga dapat menurunkan akurasi prediksi.

Peningkatan kualitas data dilakukan dengan menambah penggunaan metode pada tahap *preprocessing*. Menghapus *outliers* dengan IQR, *noise* dengan MA serta mengisi data kosong dengan LI dilakukan. Selain itu, penggunaan metode pengelompokan *K-means* juga dapat meningkatkan kualitas data. Dengan menggabungkan lokasi yang memiliki kecocokan, dapat meningkatkan kinerja pembelajaran pada model sehingga meningkatkan akurasi prediksi tingkat DO dan kadar nitrat

Hasil dari penelitian menunjukkan bahwa, pendekatan pengelompokan data berdasarkan lokasi pengambilan data menggunakan *K-means* berhasil diterapkan pada berbagai kondisi *dataset*. Bila dibandingkan dengan metode tanpa pengelompokan, metode usulan menghasilkan tingkat akurasi yang memuaskan. Pada prediksi oksigen terlarut, peningkatan nilai R^2 sebesar 0,0044%, MAE menurun sebesar 54,4267% dan MSE menurun sebesar 78,8535%. Pada prediksi nitrat, peningkatan nilai R^2 sebesar 61,9218%, MAE menurun sebesar 74,5407% dan MSE menurun sebesar 91,3194%. Hasil ini menegaskan potensi besar penggunaan *K-means* dalam meningkatkan pemodelan prediksi tingkat DO dan nitrat. Meskipun tantangan terkait akurasi masih ada pada beberapa lokasi, penggunaan strategi ini membuktikan keberhasilannya dalam memprediksi tingkat DO dan nitrat selama satu bulan dengan mempertimbangkan faktor lokasi pengambilan data menggunakan *K-means*.

Kata Kunci: Prediksi tingkat DO dan nitrat, *Preprocessing*, Pengelompokan lokasi, *K-means*, Analisa korelasi, *Spearman*, LSTM