PREDICTIVE MODELING OF POLLUTION IN RIVER BASINS USING MACHINE LEARNING TECHNIQUE

HASLINDA BINTI ABDUL SAHAK

MCS241004_MCST 1043

CHAPTER 5

UNIVERSITI TEKNOLOGI MALAYSIA

JANUARY 2025

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Summary

This research aimed to develop and evaluate advanced machine learning models for predicting water pollution levels in Malaysian river basins. The study used a holistic dataset of historical water quality data, meteorological factors, and other relevant parameters obtained from DOSM. A critical part of the research work was the collection and preprocessing of water quality data from DOSM with a great deal of caution, considering accuracy, consistency, and suitability for machine learning model training. This has used some heavy data cleaning and preprocessing methodology for handling missing values, outliers, and inconsistency in the dataset.

A range of machine learning algorithms was studied and assessed on their capabilities with respect to effective predictions of pollution levels. Among them were well-known powerful methods like Random Forest and Long Short-Term Memory-LSTM networks. It involved extensive tuning of hyperparameters to optimize each model's performance with the dual goal of maximizing its predictive accuracy and generalizability. Performance for these developed models is investigated rigorously with a suite of relevant evaluation metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared. This was to ascertain the most appropriate and reliable model that could deliver real-time or near-real-time pollution level prediction with higher accuracy.

The study identified many improvements within the water quality of Malaysian river basins, specifically for the following key indicators: Biochemical Oxygen Demand, BOD5, and Suspended Solids, SS. The upward trend of this series may reflect the effectiveness of continuous environmental interventions and policies. Some of the challenges faced in the development and deployment of such predictive models were underlined in the research. This would include good quality and consistent availability of data from multiple sources, interpretability of model output mechanisms, and smooth integration with current hydrological models to achieve complete and accurate predictions using machine learning models. Overall, the current study provides an overview of the role that machine learning could play in the further development of the potentials for water quality monitoring and management within Malaysia. These findings will be useful as guidelines for effective, data-driven strategy development aimed at mitigating pollution to protect water resources across the nation.

5.2 Future Works

The findings from this research give a very strong foundation for further studies in this field. Some of the possible promising directions that could be explored in order to further enhance the understanding and, therefore, prediction of pollution in river basins include the exploration of advanced machine-learning techniques. Deep learning architectures, such as convolutional neural networks and LSTM networks, can be employed in order to capture complex temporal and spatial patterns in pollution dynamics. Furthermore, hybrid models that marry strengths of physical and statistical approaches can result in better predictability with more insight into the process.

Integrating diverse data sources would largely help in improving the accuracy and completeness of the predictions regarding pollution. The use of IoT sensors, satellite imagery, and real-time data streams can give a better view of the patterns of pollution that could enable timely, more accurate interventions. Furthermore, explainable AI-XAI needs to be developed. The XAI models can make transparent and interpretable predictions that shall help stakeholders understand how the model made decisions and engender trust in the model outputs. This aspect of transparency acts as

a key driver for effective communication and successful implementation of model-driven interventions.

Finally, broadening the scope of research to cover global modeling of river basins can make useful inferences on the pattern and management of pollution in different geographical scopes. This would, in turn, demand global models that take into consideration the river basin characteristics for achieving an appropriate and effective approach to managing pollution globally. Besides, a study of socioeconomic impacts of pollution should be assessed in terms of human health, agriculture, tourism, among other sectors, which will ultimately help in formulating effective policies and mitigation strategies with consideration of multifaceted consequences of river basin pollution. These are the future research directions which will help us to come up with more accurate, reliable, and interpretable predictive models for river basin pollution. This will ensure better environmental management, sustainable water resource management, and a healthy planet for the future generation.