Ideation Phase

Literature Survey on The Selected Project & Information Gathering

Date	19 September 2022
Team ID	PNT2022TMID17967
Project Name	Project – Efficient Water Quality Analysis and Prediction using Machine Learning
Maximum Marks	4 Marks

Literature Surveys:

S.no	Title	Author	Description	Advantages	Disadvantages
1.	Emulating process-based water quality modelling in water source reservoirs using machine learning	Hadi Mohammed, Hoese Michel Tornyeviadzi, Razak Seidu	Demonstrated the potential of machine learning model (Long Short-Term Memory(LSTM)) . A Hydro dynamic and water quality model was first calibrated to predict time series, profiles, and contours of water variables namely Eschericha coli(E.coli), faecal coliforms, zinc, and lead concentrations. The results obtained were combined with the input data to train a suite of LSTM models to emulate the results achieved with the process-based modeling.	A validated model as developed in this study can offer an opportunity for real-time simulation of water quality.	While setting up and training the LSTM model to emulate the process-based model simulations was very time-consuming.
2.	Detecting Water Quality Using KNN, Bayesian and Decision Tree	Xudong Jia	Proposed a model using sklearn K Nearest Neighbor(KNN), Bayesian and decision tree.These models are used to classify water quality data. Comparison results show that decision tree algorithm performs best among the three supervised classification algorithms.	Method can reach high accuracy if we add the complexity of the model. If we select other features or insert more blank data, the performance can be much better. In these data set, the water features are not highly relevant and it also miss some median values. These two features perfectly correspond to the	KNN algorithm high precision which also needs more effort when calculating. Bayesian algorithm is very allergic to the preprocessing of the data.Accuracy of this method is very low. Decision tree algorithm can process irrelevant characteristic data and it is insensitive to missing median values.

				characteristics o f this algorithm	
3.	Efficient Water Quality Prediction Using Supervised Machine Learning	Umair Ahmed, Rafia Mumtaz, Hirra Anwar, Asad A.Shah, Rabia Irfan and Jose Garcia-Nieto	Proposed a methodology which employs four input parameters namely temperature, turbidity, pH and total dissolved solids.	Gradient boosting and polynomial regression predict the WQI most efficiently. It achieves reasonable accuracy using minimal number of parameters.	No application framework was built .
4.	Water Quality Index Classificati on Based on Machine Learning:A Case from the Langat River Basin Model	Illa Iza Suhana Shamsudd in, Zalinda Othman, and Nor Samsiah Sani	Proposed three machine learning models Artificial Neural Networks (ANN), Decision Trees (DT), and Support Vector Machines (SVM) to classify river water quality. Comparative performance analysis between the three models indicates that the SVM is the best model for predicting river water quality.	Use of the kernel function, the grid search method, and the multiclass classification technique impac ts the effectiveness of the SVM model. SVM can be used to forecast WQI with a high degree of accuracy.	More extensive research of water quality measures can improve the model. Supervised machine learning algorithms can be used to solve time series prediction issues on a raw water quality dataset.
5.	Machine learning based marine water quality prediction for coastal hydro-environmen t manageme nt	Tianan Deng, Kwok-Wing Chau, Huan-Feng Duan	Proposed two different ML methods – Artificial Neural Networks (ANN) and Support Vector Machine (SVM) – are implemented and improved by introducing different hybrid learning algorithms for the simulations and comparative analysis.	ANN is preferable to achieve satisfactory results with quick response. SVM accurately identify the optimal model.	SVM has lower computational efficiency due to the inclusion of the nonlinear relationships among variables and outputs.
6.	Robust machine learning algorithms for predicting coastal water quality index	Md Galal Uddin, Stephen Nash, Mir Talas Mahamm ad Diganta, Azizur Rahman,	Proposed eight commonly used algorithms, namely Random Forest (RF), Decision Tree (DT), K Nearest Neighbors (KNN), Extreme Gradient Boosting (XGB), Extra Tree (ExT), Support Vector Machine (SVM), Linear Regression (LR), and Gaussian Naïve Bayes (GNB). DT, ExT, and GXB models could be effective, robust and	The findings of this study are also useful for reducing model uncertainty and optimizing the WQM-WQI model architecture for	Inadequacy to assess the water quality in terms of temporal resolution.

		Agnieszka I. Olbert	significantly reduce model uncertainty in predicting WQIs.	predicting WQI values. The findings of this research would also have been much more useful in predicting WQIs in order to reduce the uncertainty in the WQI model.	
7.	A novel machine learning application: Water quality resilience prediction Model	Maryam Imani, Md Mahmudul Hasan, Luiz Fernando Bittencourt, Kent McClymont, Zoran Kapelan	Proposed utilises Artificial Neural Network (ANN) to develop a novel application to predict water quality resilience to simplify resilience evaluation of new attacks.	Key advantage of the predictive model is that it can be integrated into complex process-based models	The integration of different machine learnig techniques with resilence-based methods could be further explored in order to create new techniques.