

Model Machine Learning untuk Klasifikasi Cepat Kualitas Air Menggunakan Feature Selection dan Parameter yang Dipilih Secara Fundamental = Machine Learning Model for Quick Water Quality Classification Using Feature Selection and Fundamentally Selected Parameters

Adib Muhammad Prawirahutama, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920517500&lokasi=lokal>

Abstrak

Air merupakan sumber daya yang paling penting bagi kehidupan, oleh karena itu perlu diperhatikan dan dijaga kualitasnya. Dalam studi air, ML menawarkan banyak peluang untuk mengklasifikasikan kualitas air. Hasil akurasi klasifikasi kualitas air bergantung pada model yang digunakan, ukuran kumpulan data, dan parameter air yang digunakan untuk melatih model pembelajaran. Dalam makalah ini, model SVM, NB, DT, RF, dan CATBoost digunakan untuk memodelkan proses klasifikasi kualitas air. Metode feature selection: filter, wrapped, dan embeded akan dibandingkan, bersama dengan model dengan pemilihan parameter manual yang dipilih berdasarkan kemudahan pengukurannya. Menggunakan embedded feature selection dan DT classifier dengan SMOTE sebagai metode penyeimbangan kelas, model ini dapat mencapai akurasi 99,33%, presisi 99,43%, daya ingat 99,33%, dan skor F1 99,34%. Model untuk indikasi kualitas air secara realtime juga diperoleh dengan classifier CatBoost, dengan akurasi 92,31%, presisi 91,72%, recall 92,31%, dan skor F1 91,75%.

.....Water is the most important resource for life, hence it's quality needs to be checked and maintained. In water studies, ML offers numerous opportunities for classifying Water Quality (WQ) indicators. Results of WQ classification accuracy depend on the model used, the size of the data set, and the water parameters used to train the learning models. In this paper, SVM, NB, DT, RF, and CATBoost models are used to model a WQ classification. Filter, wrapped, and embeded feature selection methods will be compared, along with a model with a manual selection of parameters that are selected based on their ease of measurement. Using embedded feature selection and DT classifier with SMOTE as class balancing method, the model can achieve 99.33% accuracy, 99.43% precision, 99.33% recall, and 99.34% F1-score. Model for realtime water quality indication is also obtained with CatBoost classifier, it achieve 92.31% accuracy, 91.72% precision, 92.31% recall, and 91.75% F1-score.