

ABSTRACT

Semarang-Demak is a coastal area that utilizes clean water sourced from groundwater to support various households, agriculture and industry. Problems that often occur in coastal areas are in the form of sea water intrusion which can be caused by natural factors and human activities. Seawater intrusion can be detected using water quality data. Nowadays, prediction of water quality parameters such as total dissolved solids (TDS), electrical conductivity (EC) and turbidity in water is among the challenging issues in water resources management. One of the parameters used to determine the characteristics and quality of groundwater is the EC parameter. Incompleteness of EC data and limitations of direct measurements can interfere with analysis.

Machine learning models are known to provide the most accurate predictions. This work will find the best machine learning to find groundwater quality based on EC parameters. Three machine learning algorithms are used in the performance investigation: artificial neural networks (ANN), gaussian processes (GP), and multiple linear regression (MLR) algorithms. Prediction using hydrochemical groundwater datasets, seven hydrochemical parameters (K, Ca, Mg, Na, SO₄, Cl and HCO₃) and three physical parameters of groundwater (TDS, pH and EC). R-squared (R^2) and Root Mean Squared Error (RMSE) are used to measure the performance of the three techniques. In addition to the Semarang data set, two data sets were used when testing the accuracy and timing of modeling.

Testing the three sets of data shows that the MLR has $R^2 = 0.985$ and RMSE = 0.030, which is slightly better than the other models. In addition to accuracy, MLR offers better modeling time than GP and ANN. So it can be concluded that the MLR model can be a solution to difficult problems of EC prediction and incomplete data in water resources management.

Keywords : prediction, electrical conductivity, water quality, groundwater

INTISARI

Semarang-Demak merupakan wilayah pesisir yang memanfaatkan air bersih yang bersumber dari air tanah untuk menunjang berbagai kegiatan rumah tangga, pertanian, dan industri. Masalah yang sering terjadi di wilayah pesisir berupa intrusi air laut yang dapat disebabkan oleh faktor alam maupun aktivitas manusia. Intrusi air laut dapat dideteksi dengan menggunakan data kualitas air. Saat ini, prediksi parameter kualitas air seperti padatan terlarut atau *Total Dissolved Solid* (TDS), daya hantar listrik atau *Electrical Conductivity* (EC), dan kekeruhan dalam air merupakan salah satu isu yang menantang dalam pengelolaan sumber daya air. Salah satu parameter yang digunakan untuk mengetahui karakteristik dan kualitas air tanah adalah parameter EC. Ketidaklengkapan data EC dan keterbatasan pengukuran langsung dapat mengganggu analisis.

Teknik pembelajaran mesin dikenal dapat memberikan prediksi yang akurat. Studi ini akan menemukan model prediksi dari teknik pembelajaran mesin terbaik untuk mengetahui kualitas air tanah berdasarkan parameter EC. Tiga algoritma pembelajaran mesin digunakan dalam investigasi kinerja: jaringan saraf tiruan (JST), proses gaussian (GP), dan regresi linier ganda (MLR). Prediksi menggunakan set data hidrokimia air tanah, tujuh parameter hidrokimia (K, Ca, Mg, Na, SO₄, Cl dan HCO₃) dan tiga parameter fisik air tanah (TDS, pH dan EC). *R-squared* (R^2) dan *Root Mean Squared Error* (RMSE) digunakan untuk mengukur kinerja ketiga teknik tersebut. Selain kumpulan data Semarang, dua kumpulan data yang lain digunakan saat menguji akurasi dan waktu pemodelan.

Pengujian tiga set data menunjukkan bahwa MLR memiliki akurasi $R^2 = 0,985$ dan RMSE = 0,030 yang berarti sedikit lebih baik dari pada model lainnya. Selain akurasi, MLR juga menunjukkan waktu pemodelan yang lebih baik daripada GP dan ANN. Oleh sebab itu, dapat disimpulkan bahwa model MLR dapat menjadi solusi dari permasalahan sulitnya prediksi EC dan ketidaklengkapan data dalam pengelolaan sumber daya air.

Kata kunci -- prediksi, daya hantar listrik, kualitas air, air tanah