

ABSTRAK

Kualitas air di Provinsi DKI Jakarta mengalami penurunan yang signifikan akibat pencemaran yang semakin parah. Kondisi ini menimbulkan urgensi untuk melakukan analisis dan klasifikasi kualitas air, terutama pada sumber air sumur yang masih banyak digunakan oleh masyarakat. Penelitian dilakukan untuk mengklasifikasikan kualitas air sumur di Provinsi DKI Jakarta dengan menggunakan metode *Random Forest* dan *K-Nearest Neighbor(KNN)*. Penelitian ini menerapkan *Synthetic Minority Over-sampling Technique (SMOTE)* sebagai solusi untuk mengatasi ketidakseimbangan data dalam dataset klasifikasi. Variabel yang digunakan meliputi nitrat (kandungan nitrogen dalam NO_3), nitrit (NO_2), krom heksavalen ($Cr(VI)$), besi (Fe), mangan (Mn), total *coliform* dan *E.Coli*. Kualitas air diklasifikasikan ke dalam empat kategori, yaitu kondisi baik, tercemar ringan, tercemar sedang, dan tercemar berat. Hasil penelitian menunjukkan bahwa model *Random Forest* menghasilkan akurasi tertinggi sebesar 96,29%. Model ini dilatih menggunakan 80% data dari total 266 sampel air sumur dan diuji pada 20% sisanya. Penerapan SMOTE terbukti meningkatkan representasi kelas minoritas pada dataset tidak seimbang, sehingga berkontribusi pada peningkatan kinerja model secara keseluruhan. Namun, penerapan SMOTE pada data pengujian tidak memberikan keuntungan signifikan dibandingkan model tanpa SMOTE pada rasio pembagian data yang lebih seimbang. Algoritma *Random Forest* terbukti lebih unggul dibandingkan KNN dalam hal akurasi klasifikasi. KNN hanya mencapai akurasi sebesar 52,68%. Penelitian ini menegaskan pentingnya pemilihan algoritma yang tepat dalam analisis kualitas air untuk menghasilkan prediksi yang lebih akurat dan konsisten.

Kata Kunci: Kualitas air sumur, Klasifikasi, *Random Forest*, KNN, SMOTE