

ABSTRAK

Perbandingan Performa Model KNN, SVM, dan Random Forest pada Klasifikasi Kualitas Udara di Jakarta

Oleh

Ernawati Prahesta Kurnia Sari

21110011

Penurunan kualitas udara di Jakarta menimbulkan dampak serius terhadap kesehatan masyarakat, sehingga diperlukan metode analisis yang efektif untuk mengklasifikasikan tingkat pencemaran udara. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Random Forest (RF), dalam mengklasifikasikan kualitas udara di Jakarta. Pengujian dilakukan menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Data yang digunakan berasal dari pengukuran kualitas udara di Jakarta berdasarkan Indeks Standar Pencemar Udara (ISPU). Berdasarkan hasil perbandingan, model Random Forest menunjukkan performa terbaik dengan nilai akurasi sebesar 92%, precision 66%, recall 64%, dan F1-score sebesar 65%, sementara model KNN menghasilkan akurasi sebesar 89%, precision 64%, recall 58%, dan F1-score 60%, dan model SVM memperoleh akurasi sebesar 88% dengan nilai precision, recall, dan F1-score masing-masing sebesar 65%, 54%, dan 58%. Namun, data yang digunakan memiliki imbalance dataset yang menyebabkan penurunan kinerja pada kelas minoritas. Untuk mengatasi hal ini, diterapkan metode SMOTE pada model Random Forest. Setelah penerapan SMOTE, akurasi menjadi 83% dengan precision 87%, recall 83%, dan F1-score 82%, yang menunjukkan peningkatan pemerataan kinerja antar kelas meskipun akurasi keseluruhan menurun. Hasil penelitian ini diharapkan dapat memberikan rekomendasi model klasifikasi yang dapat diterapkan dalam sistem pemantauan kualitas udara berbasis machine learning.

Kata kunci: klasifikasi, k-nearest neighbors, kualitas udara, machine learning, random forest, support vector machine