

ABSTRAK

Menurut WHO, kebiasaan merokok adalah salah satu penyebab utama kematian global, dengan lebih dari 5 juta kematian per tahun akibat penyakit seperti kanker, penyakit jantung, penyakit hati, dan stroke. Rokok mengandung zat beracun dan karsinogenik yang terbagi dalam kategori gas dan partikel padat. Klasifikasi perokok meliputi perokok ringan, sedang, dan berat. Penelitian ini bertujuan mengklasifikasikan apakah seseorang adalah perokok atau bukan, guna membantu organisasi kesehatan dalam memahami perilaku perokok dan mengembangkan program pencegahan serta penghentian merokok. Metode K-Nearest Neighbor (kNN) digunakan untuk klasifikasi dengan memanfaatkan dataset pembelajaran. Berdasarkan pengujian dan analisa yang telah dilakukan, model kNN ini memiliki akurasi sekitar 73.56%, yang menunjukkan bahwa sekitar 73.56% dari semua prediksi model adalah benar. Precision dan recall untuk kedua kelas menunjukkan bahwa model ini lebih baik dalam memprediksi kelas 0 dibandingkan dengan kelas 1. Ada ruang untuk perbaikan, terutama dalam memprediksi kelas 1, yang menunjukkan bahwa model mungkin perlu disesuaikan lebih lanjut atau dioptimalkan untuk mencapai kinerja yang lebih baik. Untuk meningkatkan kinerja model kNN, langkah-langkah berikut dapat dilakukan: optimalkan nilai k, normalisasi fitur, gunakan jarak berbobot, terapkan PCA untuk reduksi dimensi, dan lakukan cross-validation untuk menghindari overfitting.

Kata Kunci: ***Rokok, K-Nearest Neighbor, Kesehatan.***