

ABSTRAK

Permasalahan yang terjadi dalam penerapan *K-Nearest Neighbors* sebagai algoritma klasifikasi adalah sering terjadinya *overfitting* dalam pengolahan data. Hal ini dapat diatasi dengan menggunakan teknik *cross-validation* dalam mengevaluasi model algoritma dan meminimalkan *overfitting*. Kemudian belum diketahuinya kinerja akurasi prediksi diabetes menggunakan algoritma *K-Nearest Neighbors* dengan teknik *cross-validation*. Data yang digunakan berasal dari Institusi Nasional Diabetes Penyakit Pencernaan dan Ginjal pada tahun 2021. Studi kasus dalam penelitian ini adalah mengetahui *screening* awal penyakit diabetes didukung dari hasil akurasi algoritma dan penerapan *real time* terhadap pengguna berbasis *streamlit*. Tujuan dari penelitian ini adalah untuk mengoptimalkan hasil akurasi dengan teknik *cross validation* yang didukung oleh algoritma *K-Nearest Neighbors* dalam kajian data diabetes. Metode yang digunakan adalah algoritma *K-Nearest Neighbors* yang didukung dengan teknik *cross validation* untuk hasil akurasi yang optimal. Kemudian penerapan aplikasi web interaktif berbasis *streamlit* untuk pengujian hasil akurasi yang digunakan oleh pengguna untuk melihat probabilitas bahwa pengguna menderita diabetes. Hasil penelitian menunjukkan bahwa optimasi teknik *Cross Validation* yang didukung dengan model algoritma *K-Nearest Neighbors* bekerja dengan baik. Hasil dari *confusion matrix* dengan menggunakan teknik *cross validation* lebih akurat ditinjau dari keuntungan menggunakan teknik *cross validation* itu sendiri. Sehingga laporan klasifikasi yang memiliki nilai 95% lebih akurat dibandingkan dengan akurasi yang bernilai 92% karena penggunaan teknik *cross validation* yang dapat meminimalkan *overfitting* selain pertimbangan dari value akurasinya dan penerapan aplikasi web interaktif berbasis *streamlit* untuk pengujian pengguna berjalan dengan baik.

Kata kunci: *Optimasi, Machine Learning , Algoritma, K-Nearest Neighbors, Diabetes, Cross Validation, Streamlit*