

KNN bekerja dengan cara menghitung jarak, baik itu Manhattan, Euclidean, atau Minkowski (generalisasi). Model akan memprediksi suatu data input baru dengan menghitung besaran jarak antara titik baru dengan semua titik yang ada pada training data. Kemudian, model melakukan voting antara kelas-kelas yang dimiliki oleh titik sebanyak $n_neighbors$ yang paling dekat dengan titik yang ingin diprediksi. Kelas dengan voting terbanyak akan menjadi prediksi yang diberikan oleh model.

Kinerja model yang dibangun dari scratch dengan scikit-learn tidak terlalu berbeda, dengan perbedaan lebih kecil daripada prediksi antara softmax regression milik scikit-learn dan dari scratch.

Karena tidak memiliki hyperparameter special, dengan $n_neighbors$ secara umum sebesar akar dari banyak data, maka holdout validation sudah baik. Namun tentu, yang dapat diperbaiki adalah efisiensi algoritma. Terdapat perbedaan signifikan waktu komputasi antara algoritma milik scikit-learn dan yang dibangun dari scratch. Scikit-learn menggunakan struktur data tree yang memberikan efisiensi lebih tinggi daripada list biasa, walau telah menggunakan numpy array.