

Perbandingan Decision Tree dan Kneighbors dalam memodelkan faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia

Daniel Morantha Predict 2022

Pendahuluan

Latar belakang masalah

Dalam mencapai pembangunan kualitas hidup manusia Indonesia tidak terlepas dari program pembangunan yang dilaksanakan pemerintah di tiap wilayah Indonesia baik itu tingkat pusat maupun daerah. Selain itu Penentuan program pembangunan yang dilaksanakan harus tepat sasaran dan sesuai dengan prioritas daerah berdasarkan kategori IPM yang dimiliki daerah tersebut. Indonesia perlu adanya suatu sistem keputusan yang dapat menentukan klasifikasi kategori IPM secara cepat dan akurat.

Rumusan masalah

- Bagaimana membangun sistem keputusan yang dapat menentukan klasifikasi kategori IPM secara cepat dan akurat?
- Bagaimana hasil perbandingan Decision Tree dan KNeighbors dalam memodelkan faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia?

Data Acquisition

Dataset

- Dataset yang digunakan ialah IPM.csv

Variable data

- Variable Independen:
 - Harapan_Lama_Sekolah dengan tipe data float (bilangan Desimal)
 - Pengeluaran_Perkapita dengan tipe data integer
 - Rerata_Lama_Sekolah dengan tipe data float
 - Usia_Harapan_Hidup dengan tipe data float
- Variable Dependen: IPM

Preprocessing data

Data yang dilakukan ialah menyeleksi fitur dengan jenis seleksi matriks korelasi dengan heatmap. Pada heatmap di sebelah kanan terlihat baris terakhir korelasi antara IPM dengan fitur lain dimana ada relasi kuat dengan variabel Usia_Harapan_Hidup, Pengeluaran_Perkapita, dan diikuti oleh var Rerata_Lama_Sekolah dan Harapan_Lama_Sekolah.



Model dan parameter model

Model

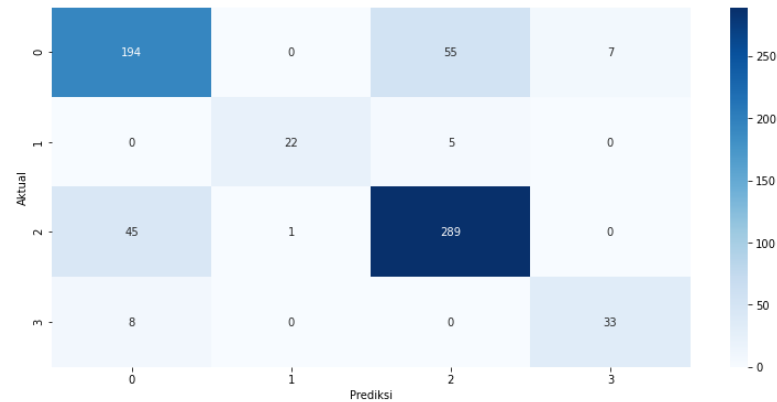
- Model menggunakan 2 algoritma klasifikasi sebagai perbandingan yakni K-nearest neighbor dan DecisionTree. Model klasifikasi dengan algoritma K-nearest neighbor memiliki akurasi model sebesar 81.18%. Sedangkan algoritma DecisionTree memiliki akurasi model sebesar 95.14%. Sehingga algoritma Decision Tree cocok terhadap dataset untuk dijadikan sebagai model.

Parameter model

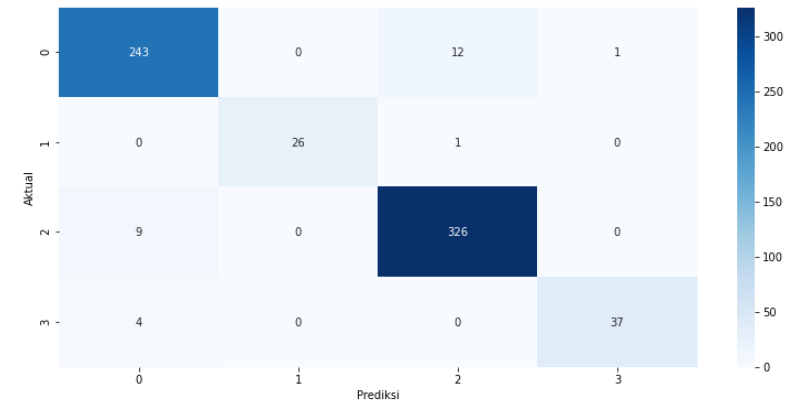
- Pada model K-NN menggunakan parameter `n_neighbors = 7`, `metric='minkowski'`, `p=1` yang artinya `p=1` ialah manhattan distance.
- Sedangkan model DecisionTree menggunakan parameter `criterion='entropy'`, `random_state=0`

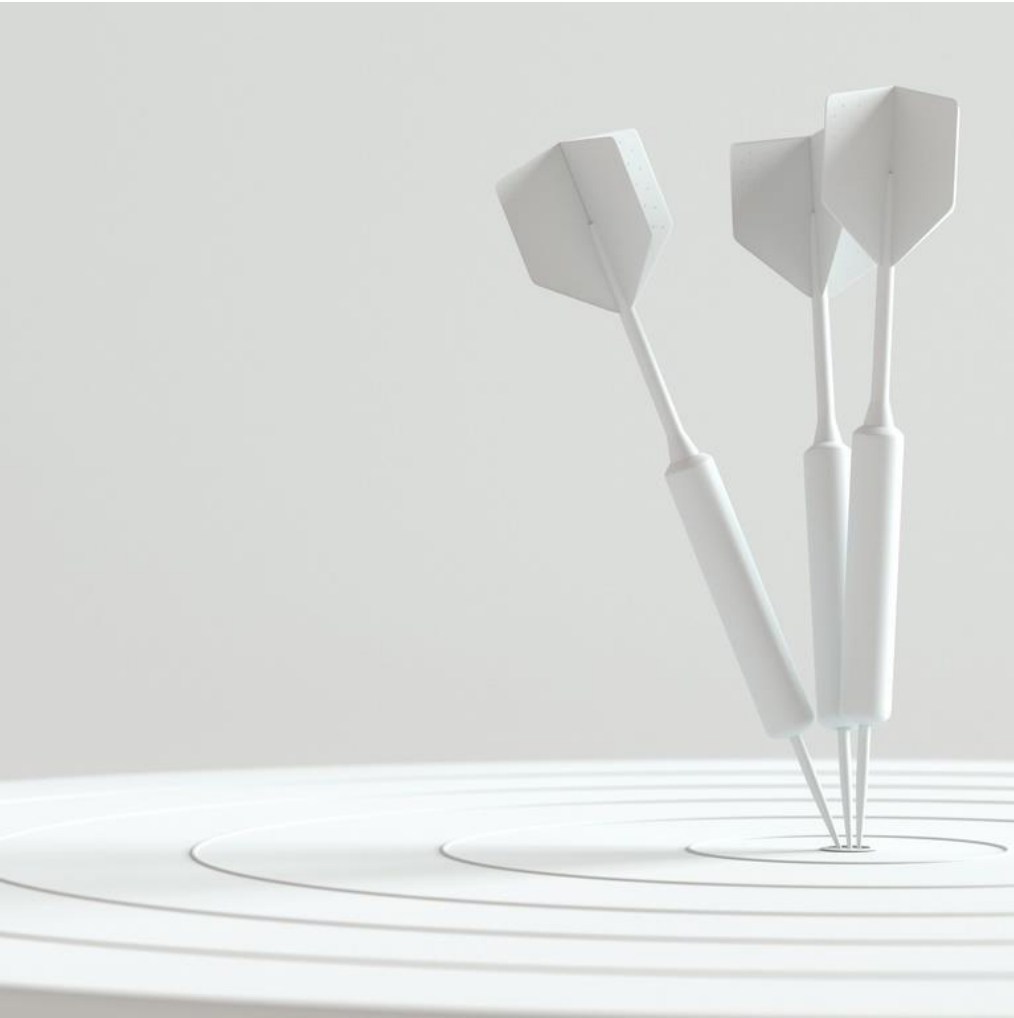
Evaluasi model

K-Nearest Neighbor



Decision Tree





Kesimpulan

- Model klasifikasi dengan algoritma K-nearest neighbor memiliki akurasi model sebesar 81.18%. Sedangkan algoritma DecisionTree memiliki akurasi model sebesar 95.14%. Sehingga algoritma Decision Tree cocok terhadap dataset untuk dijadikan sebagai model deployment.