

NIM : A11.2021.13919

NAMA : Gilang Djati P.

Klasifikasi Platform Media Sosial Berdasarkan Jumlah Followers dan Jumlah Post Menggunakan Algoritma K-NN

Dataset yang digunakan berisi kolom N, PROFILE, FOLLOWERS, POSTS, dan Platform, yang menunjukkan platform media sosial dari 100 pengguna media sosial teratas.

Pada tahap pertama yaitu tahap Preprocessing Data yang didalamnya berisi import library yang dibutuhkan, kemudian import data pada “session storage” yaitu SocialMediaTop100.csv, kemudian langkah selanjutnya yaitu membaca data csvnya, lalu datanya dibagi dalam dua variabel yaitu x, dan y, sebagaimana x untuk menampung kolom FOLLOWERS, dan POSTS sebagai atribut reguler, kemudian y untuk menampung kolom Platform sebagai label atau kelasnya, setelah itu langkah selanjutnya yaitu menghilangkan nan(missing value) menggunakan cara dengan mengganti rata-rata dari kolomnya, kemudian langkah selanjutnya adalah mengencoding labelnya dengan mengubahnya kedalam bentuk numerik.

Untuk langkah selanjutnya yaitu membagi dataset kedalam data training dan data testing dengan pembagian data testingnya 25% dan data trainingnya 75%, kemudian langkah terakhir untuk Preprocessing Data yaitu mengubah skala jarak yang terlalu besar dari kolom FOLLOWERS, dan POSTS dengan menggunakan fitur “StandardScaler”.

Pada tahap kedua yaitu tahap Pelatihan Model K-NN yang didalamnya berisikan modeling menggunakan K-NN untuk pembelajaran menggunakan “x_train”, dan “y_train” dengan menentukan K untuk tetangga terdekatnya = 5, dan menggunakan perhitungan jaraknya dengan “minkowski” untuk standartnya, kemudian untuk langkah selanjutnya yaitu membuat variabel baru untuk memprediksi label, setelah langkah tersebut perlu dilakukan untuk menghitung evaluasi atau validasinya menggunakan “confusion_matrix” dan dapat diperoleh nilai akurasi dari hasil evaluasi tersebut sekitar 89%, maka model yang dihasilkan dapat dikatakan baik.

Kemudian untuk langkah terakhir pada tahap Pelatihan Model K-NN berikut yaitu menampilkan visual gambarnya atau visual persebaran data training dan data testing menggunakan Klasifikasi K-NN. Pada langkah ini hal pertama yang perlu dilakukan yaitu membuat settingannya pada data training terlebih dahulu dengan “x_set”, dan “y_set” dengan mengambil dari “x_train”, dan “y_train”, kemudian membagi datanya menjadi dua variabel yaitu x1 untuk FOLLOWERS, dan x2 untuk POSTS, kemudian untuk selanjutnya yaitu dari “x1”, “x2”, “classifier.predict” menggunakan ListedColormap red, green, blue, yellow, dan purple. Apabila tampilan dari visual persebaran training telah selesai, maka perlu dilakukan juga untuk melihat visual persebaran dari data testing, dimana tahapannya sama seperti tahap untuk menampilkan visual gambar dari data training dengan mengubah settingannya yang mengambil dari “x_test”, dan “y_test”.

Dari seluruh tahapan diatas, maka dapat disimpulkan bahwa tujuan eksperimen ini dapat menghasilkan temuan-temuan yang membantu dalam hubungan, kesamaan, perbedaan, serta memberikan rekomendasi atau pemetaan terkait platform-platform sosial media tersebut, serta dapat juga memberikan gambaran mengenai kemampuan algoritma K-NN dalam mengklasifikasikan platform media sosial berdasarkan atribut tersebut.

Link Repository Github : <https://github.com/GilangDjati/Blok-Perkuliahan-Data-Mining>