

Muhammad Hafidh Raditya

1301184079

IF-42-03

A. Teknik Prapemrosesan Data

Pada program kali ini bahasa pemrograman yang saya gunakan adalah Python. Saya mengimport data yang berformat .csv menggunakan library pandas pada python. Dataset saya masukkan menjadi bentuk tuple yang memiliki 9 elemen per tuple sesuai dengan banyaknya parameter yang ada pada file .csv. setelah dataset berhasil diimport, dataset saya bagi menjadi 5 dataset baru dengan metode 5-fold-cross-validation, yang berarti dari 768 record data, nantinya akan dibagi menjadi 1/5 data untuk datatest dan sisanya untuk data train. Dengan ini berarti jumlah datatest menjadi kurang lebih 154 record data.

B. Perhitungan Ukuran Jarak

Pada program saya kali ini, saya memakai metode Euclidean dan metode Manhattan. Nantinya, kedua metode ini akan dianalisis, metode mana yang memiliki akurasi yang lebih baik.

C. Strategi kNN

Karena ukuran datatrain berjumlah 154 record data, maka akan dilakukan looping sebanyak record datatrain tersebut yaitu 154. Nantinya 154 data tersebut akan dibandingkan satu persatu kepada seluruh datatest yang ada untuk diketahui berapa jaraknya. Selanjutnya, akan diseleksi hasil mana yang lebih banyak kemiripan, yaitu dengan menyeleksi distance terkecil sebanyak k input, lalu dari k distance terkecil yang sudah diseleksi, akan dilihat datatrain tersebut lebih mirip ke positif diabetes atau negatif diabetes. Setelah itu akan dihitung akurasinya untuk setiap k yang diinputkan.

D. Pemilihan nilai k terbaik

Karena ukuran datatest ada 154 record data, maka saya menggunakan input sebanyak 154 sebagai ukuran nilai k paling besar. Namun karena 154 ini merupakan nilai yang sangat besar, maka proses eksekusi program juga akan menjadi lumayan lama. Nantinya dari nilai k 1 sampai 154, akan dilihat nilai k berapa yang menghasilkan akurasi paling besar.

E. Analisis dan Output Program

Skenario running program saya bagi menjadi dua, yang pertama menggunakan algoritma euclidean distance, dan yang kedua menggunakan algoritma manhattan distance.

```
Pengantar Kecerdasan Buatan/TUGAS 3/IF  
4203_1301184079/IF4203_1301184079.py"  
NILAI K MAKSIMUM = 14  
RATA RATA AKURASINYA = 75.87%  
PS D:\SEMESTER 5\Pengantar Kecerdasan  
Buatan\TUGAS 3\IF4203_1301184079>
```

Gambar diatas merupakan hasil running program menggunakan algoritma euclidean distance. Nilai k maksimum yang didapat adalah 14 dengan rata-rata akurasi mencapai 75.87%.

```
Pengantar Kecerdasan Buatan/TUGAS 3/IF  
4203_1301184079/IF4203_1301184079.py"  
NILAI K MAKSIMUM = 33  
RATA RATA AKURASINYA = 76.52%  
PS D:\SEMESTER 5\Pengantar Kecerdasan  
Buatan\TUGAS 3\IF4203_1301184079>
```

Gambar diatas merupakan hasil running program menggunakan algoritma manhattan distance. Nilai k maksimum yang didapat adalah 33 dengan rata-rata akurasi mencapai 76.52%.

Setelah mendapatkan hasil running dari kedua program, didapatkan nilai k yang berbeda. Pada pengujian menggunakan euclidean distance, didapatkan nilai k maksimumnya 14 namun akurasi rata-ratanya hanya 75.87%. Sedangkan pada pengujian menggunakan manhattan distance, didapatkan nilai k maksimumnya 33 dengan akurasi rata-rata yang lebih besar yaitu 76.52%. Maka dapat disimpulkan, untuk dataset yang diberikan pada tugas pemrograman 3 kali ini, lebih efektif menggunakan algoritma manhattan distance, karena menghasilkan akurasi rata-rata yang lebih besar jika dibandingkan dengan algoritma euclidean distance.