**Laporan Observasi**

**Tugas Pemrograman 03 KNN**

**CII-2M3 Pengantar Kecerdasan Buatan**

**Genap 2020/2021**



Disusun Oleh :

Kelompok 6

1. Irwan Reza Firmansyah (1301194018)
2. Ramadito Ferdian Assa (1301194005)
3. Artisa Bunga Syahputri (1301194007)

**FAKULTAS INFORMATIKA**

**PROGRAM STUDI S1 INFORMATIKA**

**UNIVERSITAS TELKOM**

**2020/2021**

**DAFTAR ISI**

[PENDAHULUAN 2](#_Toc75677218)

[HAL YANG DI OBSERVASI 3](#_Toc75677219)

[1. Penggunaan ukuran jarak yang digunakan 3](#_Toc75677220)

[2. Teknik prapemrosesan data 3](#_Toc75677221)

[3. Strategi penggunaan algoritma kNN 3](#_Toc75677222)

[PROSES YANG DIBANGUN 4](#_Toc75677223)

[1. Membaca file inputan (mobil.xls) 4](#_Toc75677224)

[2. Prapemrosesan data 4](#_Toc75677225)

[3. Perhitungan Ukuran Jarak 6](#_Toc75677226)

[ Euclidean distance 6](#_Toc75677227)

[ Manhattan distance 7](#_Toc75677228)

[ Minkowski distance 7](#_Toc75677229)

[ Supremum distance 8](#_Toc75677230)

[4. Klasifikasi kNN 8](#_Toc75677231)

[5. Mengeluarkan file rekomendasi mobil (rekomendasi.xls) 10](#_Toc75677232)

[HASIL DAN KESIMPULAN 12](#_Toc75677233)

# PENDAHULUAN

Pada Tugas Pemrograman kali ini kami akan membuat sebuah sistem rekomendasi mobil berbasis kNN untuk memilih tiga mobil terbaik sesuai dengan inputan user.

Sistem akan membaca masukan dari file mobil.xls dan akan mengeluarkan output berupa file rekomendasi.xls yang akan berisi satu vektor kolom berisi tiga baris string yang menyatakan nama mobil yang direkomendasikan dari file mobil.xls

Pada file mobil.xls berisi himpunan data mobil dengan lima atribut:

1. Ukuran akan bernilai dari 0-10 dimana semakin besar angkanya semakin besar ukuran mobil.
2. Kenyamanan akan bernilai 0-10 dimana semakin besar angkanya semakin nyaman mobilnya.
3. Irit akan bernilai 0-10 dimana semakin besar angkanya akan semakin irit.
4. Kecepatan akan bernilai 0-10 dimana semakin besar akan semakin cepat.
5. Harga akan bernilai ratus juta, jika bernilai 1 maka akan senilai 100 juta, 3,75 artinya mobil akan berharga 375 juta.

Akan bangunlah sebuah sistem rekomendasi mobil berbasis kNN untuk memilih 3 (tiga) mobil terbaik sesuai dengan inputan user. Sistem membaca masukan file mobil.xls dan mengeluarkan output berupa sebuah file rekomendasi.xls yang berisi satu vektor kolom berisi 3 (tiga) baris string yang menyatakan nama mobil yang direkomendasikan pada file mobil.xls.

# HAL YANG DI OBSERVASI

Hal yang di observasi pada tugas ini adalah:

## Penggunaan ukuran jarak yang digunakan

Pada tugas ini ukuran yang digunakan dalam penyelesaian masalah ada 4 yaitu Euclidean, Manhattan, Minkowski, dan Supremum. Dari hasil perhitungan jarak dengan menggunakan 4 metode tersebut, kemudian akan dilakukan perbandingan antara hasil yang satu dengan yang lain nya dan melihat perbedaan jarak yang dihasilkan dengan menggunakan setiap metode yang berbeda

## Teknik prapemrosesan data

Pada prapemprosesan dilakukan normalisasi terhadap data dengan menggunakan min-max scaling. Metode ini memungkinkan kita untuk melakukan transformasi linear terhadap data asli. Setiap nilai pada kolom ‘Harga’ akan dikurangi dengan nilai minimum, lalu dibagi dengan nilai maksimum yang dikurangi dengan nilai minimum dari data yang sama.

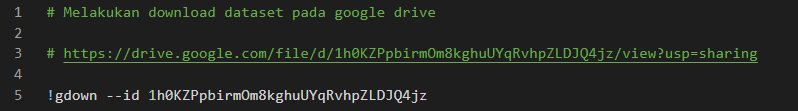
## Strategi penggunaan algoritma kNN

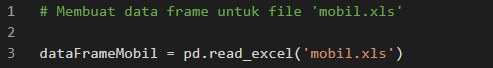
Hal pertama yang dilakukan dari program yang kamu buat adalah meminta inputan dari user berupa ukuran, kenyamanan, irit, kecepatan dan harga. Lalu program akan mencari rekomendasi 3 mobil terbaik berdasarkan inputan yang diberikan user menggunakan perhitungan jarak berdasarkan setiap metode yang digunakan .

# PROSES YANG DIBANGUN

## Membaca file inputan (mobil.xls)

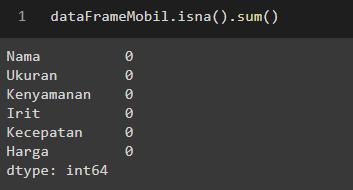
Proses yang dibangun pertama kali adalah melakukan pembacaan file inputan mobil.xls yang terdiri dari kolom nama mobil, ukuran, kenyamanan, keiritan, kecepatan, dan harga. Berikut jika diimplementasikan pada programnya :



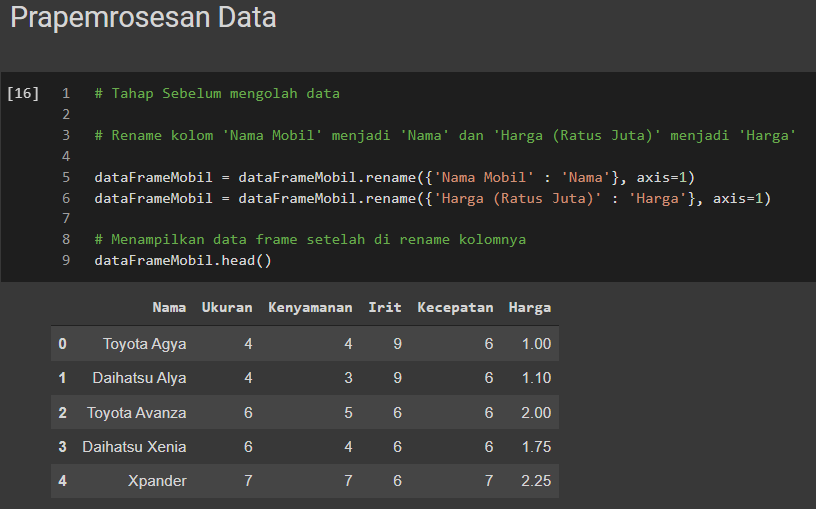


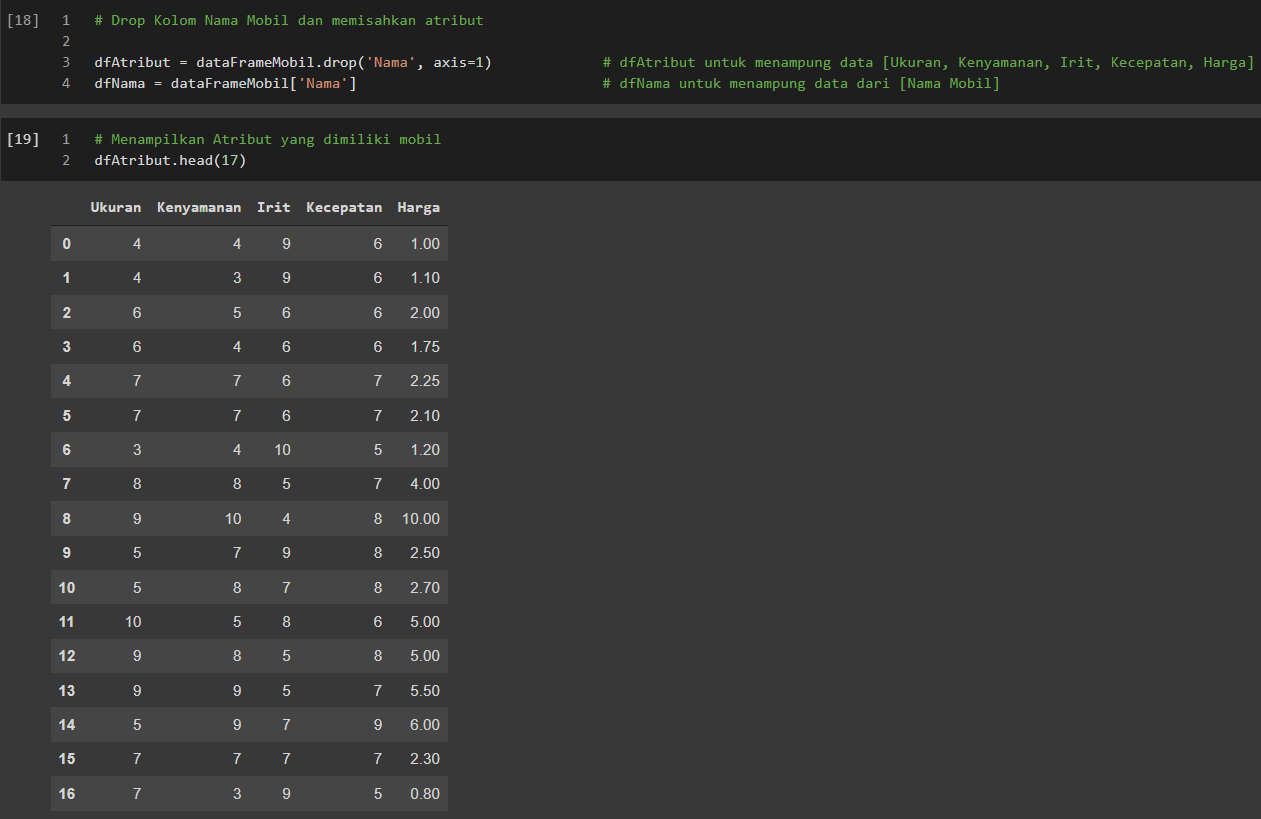
## Prapemrosesan data

Tahap selanjutnya yang kami lakukan adalah prapemprosesan data, dimana pada tahapan ini akan dilakukan pengecekan apakah ada data yang bernilai null pada data frame atau tidak.

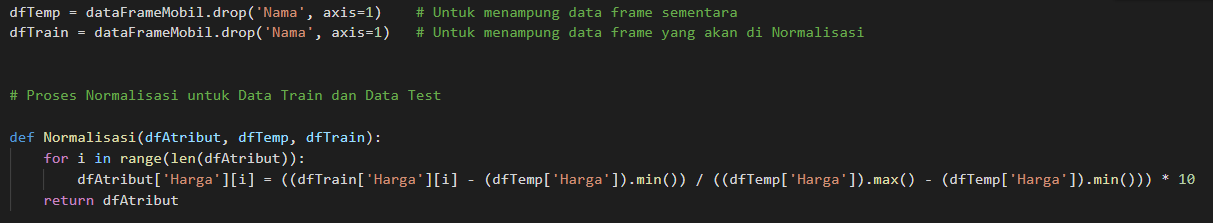


Lalu kami melakukan rename nama kolom yang memiliki lebih dari satu kata. Yang kami rename adalah kolom ‘Nama Mobil’ menjadi ‘Nama’ dan kolom ‘Harga dalam juta rupiah)’ menjadi ‘Harga’. Selanjutnya,kami melakukan drop data terhadap kolom yang tidak digunakan, kolom tersebut kemudian akan dimasukan ke dalam variabel baru. Kolom yang kami drop pada program ini adalah kolom ‘Nama’.





Selanjutnya kami melakukan proses normalisasi dengan menggunakan metode **min-max scaling** untuk tabel ‘Harga’. Setelah melakukan proses normalisasi dilanjutkan dengan proses perhitungan jarak



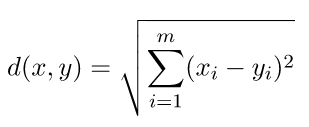
## 3. Perhitungan Ukuran Jarak

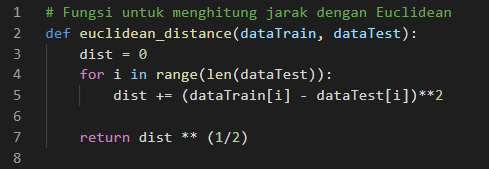
Pada program ini kami menggunakan 4 metode perhitungan jarak terhadap data yang kami miliki, metode perhitungan jarak yang kami gunakan antara lain:

### Euclidean distance

Euclidean distance adalah proses perhitungan jarak antara dua titik yang berada pada Euclidean space. Euclidean berhubungan dengan teorema Phytagoras yang pada umumnya diterapkan pada satu, dua, dan tiga dimensi .

Pada tugas kali ini kami menggunakan perhitungan jarak dengan menggunakan rumus euclidean pada 1 dimensi, dengan rumus:

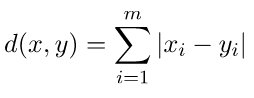


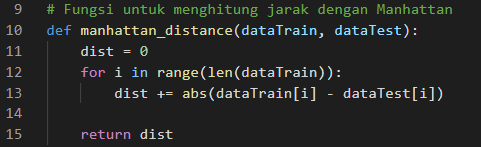


### Manhattan distance

Manhattan distance adalah jumlah panjang ruas garis dari dua titik terhadap tiap sumbu pada koordinat kartesius.

Rumus Manhattan distance yang kami gunakan pada tugas ini adalah :

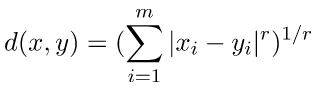


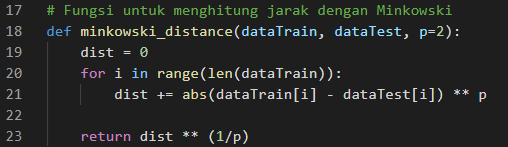


### Minkowski distance

Minkowski distance adalah perhitungan jarak matriks dalam ruang vektor bernorma, jarak ini biasa disebut sebagai bentuk generalisasi dari jarak euclidean dengan jarak manhattan.

Rumus dari minkowski distance yang kami gunakan dalam tugas ini adalah:



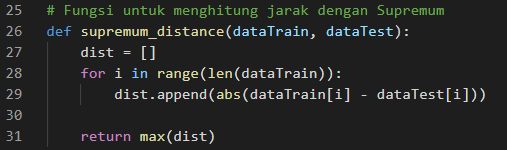


### Supremum distance

Supremum distance adalah teknik perhitungan jarak yang biasanya digunakan dalam analisis regresi pada matematika.

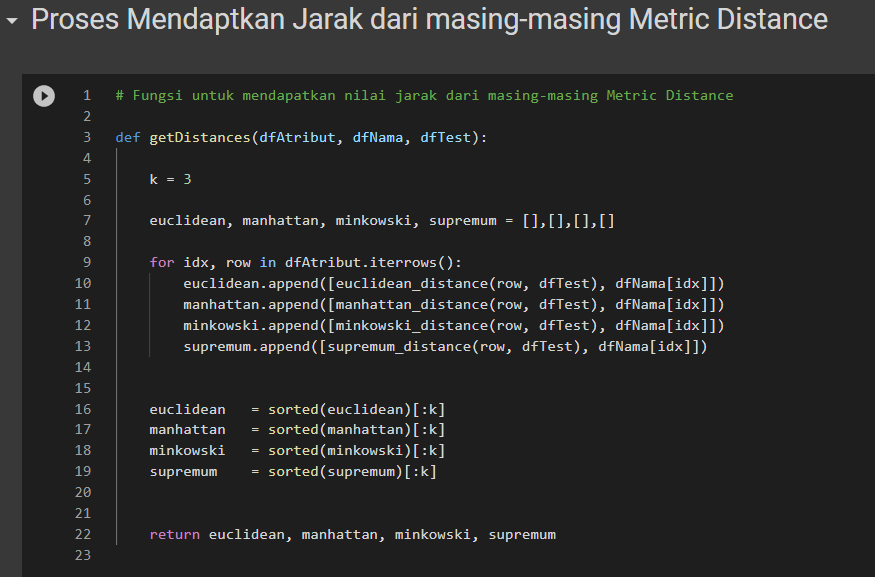
Rumus dari supremum distance yang kami gunakan dalam tugas ini adalah:



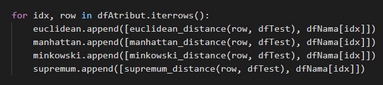


## 4. Klasifikasi kNN

Berikut adalah algoritma dari klasifikasi kNN yang kami bangun. Kami membuat terlebih dahulu function getDistance yang memiliki parameter dfAtribut yaitu data frame yang tidak memiliki kolom nama didalamnya, lalu ada df nama , yaitu data frame yang hanya memuat kolom nama mobil saja, dan dfTest, yaitu data test yang diinputkan oleh user.



Fungsi ini akan bekerja dengan menyediakan 4 list kosong yang akan diisi dengan hasil perhitungan jarak dari setiap metode yang dimiliki.



Perhitungan jarak dari data train mobil.xls dengan menggunakan 4 metode yang berbeda dan di tampung pada setiap variabel yang berbeda tadi kemudian akan di sorting secara ascending untuk menentukan 3 data terbaik dari setiap metode yang digunakan, dengan menset nilai k = 3.



## 5. Mengeluarkan file rekomendasi mobil (rekomendasi.xls)

Pada proses mengeluarkan file rekomendasi mobil adalah meminta inputan dari user terhadap nilai ukuran, kenyamanan, irit, kecepatan, dan harga dengan rentang nilai dari 0-10 yang akan dijadikan sebagai data test.

Pada tugas ini kami menetapkan data test nya adalah:

Ukuran = 7

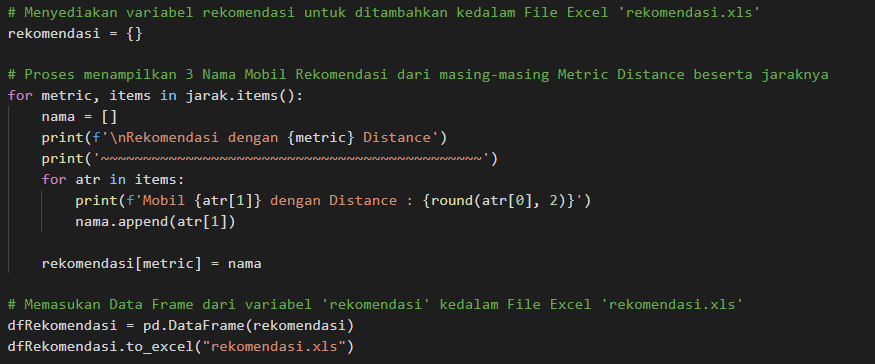
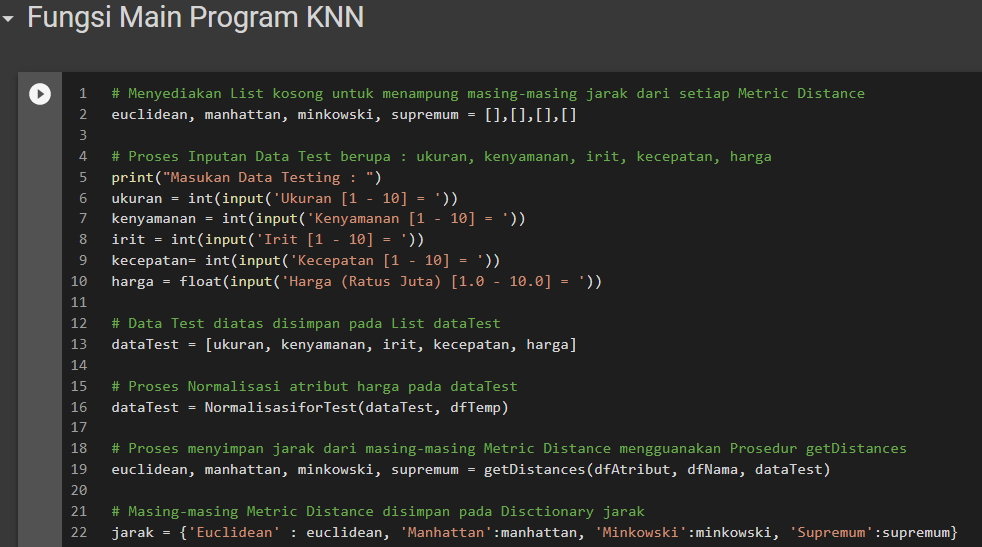
Kenyamanan = 8

Irit = 4

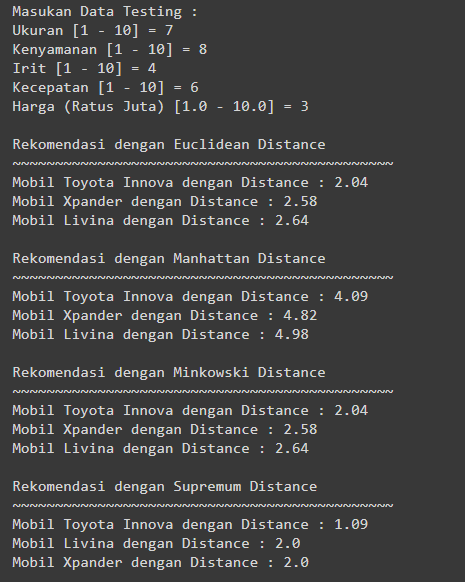
Kecepatan = 6

Harga = 3

Selanjutnya dilakukan perhitungan jarak dengan memanggil fungsi getDistance yang kemudian hasil dari perhitungan jarak tersebut akan memberikan rekomendasi 3 mobil terbaik berdasarkan metode yang digunakan.

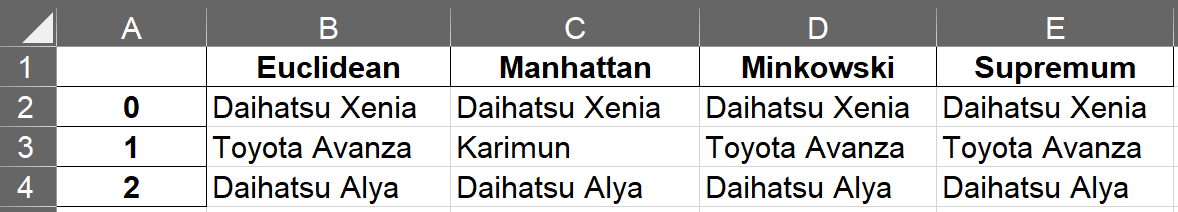


Hasil dari perhitungan jarak akan ditampilkan dalam file rekomendasi.xls



Dan didapatkan hasilnya:

“rekomendasi.xls”



# HASIL DAN KESIMPULAN

Algoritma k-nearest neighbor (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Algoritma ini dapat menghitung jarak terdekat (*near*) antara titik satu dengan lainnya. Kami diminta mengambil 3 neighbour yang paling dekat dengan data yang diminta untuk menjadi hasil keluaran program kami. Kami melakukan perhitungan menggunakan 4 metode perhitungan jarak yaitu Euclidean, Manhattan, Minkowski, dan Supremum. Tentunya dari masing-masing metode diatas memiliki nilai (value) jarak yang berbeda-beda namun untuk persoalan 3 rekomendasi mobil terbaik ini metode yang paling cocok dipakai yaitu metode Manhattan karena metode ini lebih cocok dengan data-data yang ada pada mobil.xls dimana datanya memiliki interval antara 0 - 10 ini yang menyebabkan algoritma Manhattan cukup baik dalam perhitungannya. Walaupun nilai jarak yang dihasilkan dari masing-masing Metric Distance tersebut berbeda, namun hasil rekomendasi yang dikeluarkan memiliki kesamaan.

Link Presentasi :