Ujian Tengah Semester Ganjil 2023/2024

Kelompok :

1. Muhammad Audya Fadhlurrohman 202310041
2. Angelive Hilsunny 202310014

Pengerjaan

1. Soal dapat dikerjakan dalam kelompok beranggotakan 1-3 orang.
2. Soal nomor 2 s.d 4 dikerjakan menggunakan bahasa pemrograman Python.
3. File yang dikumpulkan :
4. Coding Python (dalam format .py atau .ipynb, boleh dikumpulkan dalam bentuk *link repository*/Github)
5. Laporan (dalam format .docx atau .pdf)
6. Kumpulkan file yang diminta melalui email ke alamat [isnan.mulia@ibik.ac.id](mailto:isnan.mulia@ibik.ac.id), dengan menggunakan *subject* “[UTS Machine Learning 2023/2024]”. Pengumpulan cukup diwakili oleh salah satu anggota kelompok, di-cc-kan kepada seluruh anggota kelompok.
7. Batas waktu pengumpulan adalah **1 pekan setelah jadwal UTS Machine Learning.**

Terdapat sebuah *file* CSV yang berisi dataset *smartphone*. Dataset ini mengandung 20 atribut predictor, dan 1 atribut label dengan nama “price\_range”. Dataset sudah dimodifikasi sedemikian rupa sehingga mengandung beberapa missing values. Dataset dapat diakses pada link <https://t2m.io/c3guBDpu>.

Jawaban

1. Dari dataset yang diberikan, berikut ini hasil identifikasi kami

Atribut Prediktor (Fitur):

1. battery\_power
2. blue
3. clock\_speed
4. dual\_sim
5. fc
6. four\_g
7. int\_memory
8. m\_dep
9. mobile\_wt
10. n\_cores
11. pc
12. px\_height
13. px\_width
14. ram
15. sc\_h
16. sc\_w
17. talk\_time
18. three\_g
19. touch\_screen
20. wifi

Atribut Label:

- price\_range

Untuk atribut-atribut kategorik seperti "blue," "dual\_sim," "four\_g," "three\_g," "touch\_screen," dan "wifi," biasanya memiliki nilai unik berupa 0 atau 1, yang mungkin mengindikasikan "Tidak" atau "Ya."

Jika Anda ingin mengekstrak nilai unik dari atribut tersebut, Anda dapat melakukannya dengan menghitung jumlah kemunculan nilai unik di setiap atribut. Misalnya, untuk atribut "blue":

Nilai Unik di "blue": [0, 1]

Nilai Unik di "four\_g": [0, 1]

Nilai Unik di "three\_g": [0, 1]

Nilai Unik di "touch\_screen": [0, 1]

Nilai Unik di "wifi": [0, 1]

1. Melakukan praproses data. Detail untuk pengerjaan praproses data adalah sebagi berikut.
2. Pisahkan bagian atribut prediktor dengan atribut label menjadi dua variable terpisah.
3. Gunakan SimpleImputer untuk menangani *missing values,* dengan nilai median dari atribut data sebagai strategi pengisian nilai.
4. Terapkan MinMaxScaler atau StandardScaler untuk mengubah nilai atribut predictor, agar nilai dari masing-masing atribut menjadi seragam.
5. Tampilkan statistic deskriptif dari atribut data sebelum dan setelah dilakukan pengisian *missing values* dan standarisasi.
6. Detail untuk pengerjaan klasifikasi adalah sebagai berikut.
7. Gunakan salah satu algoritma berikut untuk membangun model klasifikasi: KNN, *Decision Tree,* SVM (parameter yang digunakan boleh disesuaikan dengan bebas).
8. Gunakan metode *holdout* untuk membagi keseluruhan dataset menjadi data *training* dan data *testing*, dengan proporsi 80% data *training* dan 20% data *testing*.
9. Tampilkan *confusion matrix* dan akurasi dari model klasifikasi.
10. Membangun model clustering berdasarkan dataset yang telah dilakukan praproses data, jika memungkinkan. Detail untuk pengerjaan *clustering* adalah sebagai berikut.
11. Gunakan salah satu algoritma berikut untuk melakukan *clustering* : K-Means, DBSCAN, *Agglomerative clustering* (parameter yang digunakan boleh disesuaikan dengan bebas).
12. Tampilkan *silhouette score* dari model *clustering*.
13. Merangkum apa saja yang dilakukan pada keempat tugas di atas dalam bentuk laporan tertulis. Detail yang harus dituliskan adalah sebagai berikut.
14. Pembahasan mengenai nama dan jenis atribut dari setiap atribut prediktor dan atribut label yang terdapat pada dataset.
15. Pembahasan mengenai statistic deskriptif dari data, baik untuk data sebelum dilakukan praproses maupun data setelah dilakukan pengisian *missing values* dan standarisasi.
16. Pembahasan mengenai model klasifikasi.
17. Pembahasan mengenai model *clustering*.
18. Lengkapi laporan dengan nama dan NIM setiap anggota, serta kelas.