

Case-Based 2 (Point 15)
Pembelajaran Mesin CII3C3
Fakultas Informatika – Universitas Telkom

Durasi waktu pengerjaan +/- 2 minggu mulai tanggal 14 Nov sampai dengan 04 Des 22 via LMS masing-masing

Tugas ini merupakan pengukuran pemahaman individu mahasiswa untuk mengimplementasikan dan menganalisis algoritma unsupervised learning (kmeans/dbscan/hierarchical) dari kumpulan data yang tersedia.

Scenario

Mengikuti keberhasilan tugas sebelumnya, Anda diberi kesempatan lebih lanjut untuk mengesankan Atasan atau Dosen Anda mengenai kemampuan Anda untuk mengimplementasikan algoritma unsupervised dan menganalisis data. Anda diminta untuk melakukan beberapa analisis dan menghasilkan luaran yang berguna menggunakan dataset yang tersedia online sebagai berikut:

[untuk NIM akhir GENAP gunakan data ini]

<https://www.kaggle.com/datasets/rohan0301/unsupervised-learning-on-country-data>

[untuk NIM akhir GANJIL gunakan data ini]

<https://archive.ics.uci.edu/ml/datasets/Water+Treatment+Plant>

Dataset masih terdapat missing value atau outlier. Harap lakukan perbaikan terhadap hal ini, selanjutnya anda harus menganalisa data tersebut. Jika perlu konversi variable kategori menjadi integer. Jika perlu lakukan normalisasi data melalui fitur rescaling. Jika perlu lakukan analisa elbow. Jika perlu lakukan analisa dengan plot data secara visual. Jika perlu lakukan transformasi data secara logaritmik. Dan masih banyak kemungkinan Analisa data yang dapat anda lakukan.

Hint: Anda bebas memilih satu dari tiga alat analisis data yaitu Weka, R, atau Python untuk membantu Anda menganalisis data dan menunjukkan pra-pemrosesan data yang diperlukan.

Tugas Anda

Tujuan dari tugas ini yaitu Anda diharapkan mampu menjelaskan, mengimplementasikan, menganalisis, dan mendesain teknik pembelajaran mesin unsupervised learning yaitu kmeans/dbscan/hierarchical.

Pertama, selidiki masalah kualitas data yang telah diberikan di atas. Jelaskan keputusan Anda mengenai pendekatan pra-pemrosesan data. Jelajahi kumpulan data dengan meringkas data menggunakan statistik dan mengidentifikasi masalah kualitas data apa pun. Tidak ada batasan jumlah ringkasan yang akan dilaporkan tetapi Anda diharapkan hanya melaporkan yang paling relevan.

Kedua, pilih salah satu dari metode unsupervised learning yang telah dipelajari yaitu kmeans/dbscan/hierarchical. Anda hanya perlu memilih satu metode untuk diterapkan. Gunakan algoritma tersebut untuk memberikan beberapa output/outcome dengan menggunakan variasi parameter, kemudian membuat laporannya dan menganalisis hasilnya.

Struktur Laporan

Struktur laporan berikut direkomendasikan, tetapi Anda dipersilakan untuk membuatnya lebih lengkap dan informatif.

Bagian-0: Buat cover yang berisi Nama, NIM, Kelas, dan kode Dosen serta tuliskan pernyataan bahwa ***Anda mengerjakan tugas ini dengan cara yang tidak melanggar aturan perkuliahan dan kode etik akademisi.***

Bagian-1: Ikhtisar kumpulan data yang dipilih (15 poin).

Tulis atau buat dalam bentuk Tabel yang terstruktur mengenai kumpulan data yang Anda pilih dan jelaskan mengenai data tersebut dengan cara seperti meringkas properti data, dll. Jelaskan penyelidikan kualitas data yang telah Anda lakukan dan poin-poin kunci yang menginformasikan pilihan pendekatan pra-pemrosesan data Anda. Sediakan beberapa plot/diagram yang diperlukan yang mewakili kumpulan data dan tulis beberapa kalimat yang menjelaskan plot/diagram tersebut.

Bagian-2: Ringkasan pra-pemrosesan data yang diusulkan (15 poin)

Jelaskan pendekatan pra-pemrosesan data yang Anda usulkan, dan ini harus mencakup teknik/metode pra-pemrosesan data serta alasannya. Nyatakan hasil yang Anda harapkan dari pendekatan prapemrosesan data yang diusulkan.

Berikan ringkasan singkat tentang alat analisis data mana yang telah Anda gunakan. Berikan beberapa screenshot dengan beberapa kode contoh. Tuliskan juga beberapa kalimat yang menjelaskan setiap screenshot tersebut.

[Catatan: jika ada yang gagal, Anda masih dapat membuktikan kesalahan dan mendapatkan pertimbangan penilaian untuk bagian ini.]

Bagian-3: Menerapkan algoritma yang di pilih (30 poin)

Jelaskan dengan detail algoritma unsupervised learning yang Anda pilih (kmeans/dbscan/hierarchical). Nyatakan hasil atau output yang Anda harapkan dari penggunaan algoritma yang dipilih.

Berikan beberapa analisis yang menjelaskan pemilihan parameter yang berbeda-beda, dengan sebelumnya menuliskan penggunaan parameter dalam bentuk tabel. Jelaskan bagaimana metode Anda memberikan hasil, jika diperlukan dapat membuat flowchart. Tulis beberapa kalimat yang menjelaskan setiap screenshot hasil code Anda dan apa makna yang ditampilkannya.

Kemudian tuliskan link google colab atau github untuk code anda. Berikan komentar penjelas di setiap block-code anda. **Penggunaan library tidak diperbolehkan, diharapkan mahasiswa dapat mendvelop codenya sendiri.**

[Catatan: jika ada yang gagal, Anda masih dapat membuktikan kesalahan dan mendapatkan pertimbangan penilaian untuk bagian ini.]

Jangan lupa, tuliskan referensi dari mana ada dapat source-code & url pendukung.

Bentuk plagiasi code dari orang lain tanpa mencantumkan sumber, otomatis mendapat nilai NOL pada tugas ini.

Bagian-4: Evaluasi hasil (20 poin)

Tuliskan evaluasi terkait hasil yang Anda (atau belum) temukan, buat dalam bentuk tabel. Kemudian, jelaskan pengukuran performansi yang Anda lakukan dan berikan analisis mendalam mengenai hasil yang diperoleh.

Bagian-5: presentasi video (20 poin)

Tuliskan link laporan, slide dan video presentasi pada bagian ini. Buat video presentasi berdurasi +/- 10 menit dimana pada awal video diawali dengan pengenalan yang memperlihatkan muka Anda, selanjutnya Anda dapat menjelaskan menggunakan slide, kemudian demo program dan terakhir tunjukkan laporan yang di buat.