

Mata Kuliah : IBDA 3122 Knowledge Discovery
 Hari, Tanggal : Jumat, 24 Maret 2022
 Waktu : Batas Pengumpulan Pk. 17.00 WIB
 Sifat Ujian : Proyek

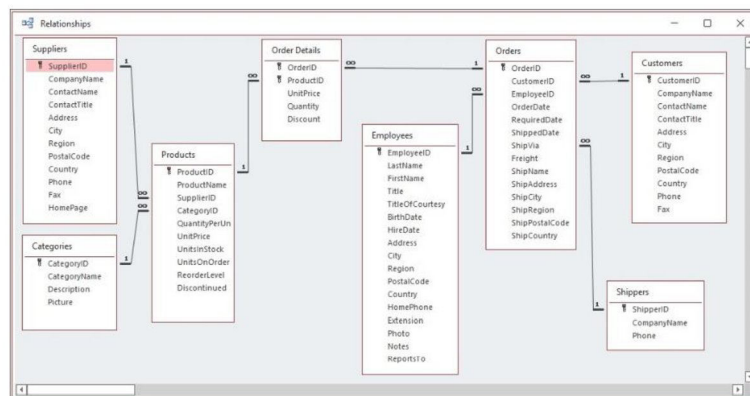
PETUNJUK

- Terdapat empat proyek, pilih proyek untuk diselesaikan dengan nilai maksimum 20 poin.
- Kumpulkan laporan dan *source code* dalam bentuk file zip.
- Dalam laporan pada bagian akhir, salinlah pernyataan komitmen integritas berikut ini dan bubuhkanlah tanda tangan di bawahnya:

"Di hadapan TUHAN yang hidup, saya menegaskan bahwa saya tidak memberikan maupun menerima bantuan apapun—baik lisan, tulisan, maupun elektronik—di dalam ujian ini selain daripada apa yang telah diizinkan oleh pengajar, dan tidak akan menyebarkan baik soal maupun jawaban ujian kepada pihak lain."

Proyek A (16 poin)

Database Northwind adalah database sampel, yang berisi data penjualan untuk perusahaan fiktif bernama "Northwind Traders", yang mengimpor dan mengekspor makanan khusus dari seluruh dunia. Database Northwind memiliki data sampel *customers*, *orders*, *product*, *suppliers*, *shipping*, dan *employees*. Hubungan tabel ditampilkan dalam diagram E-R berikut.



Database Northwind

Produk yang diperdagangkan terbagi atas 8 kategori:

Beverages, Condiments, Confections, Dairy Products, Grains/Cereals, Meat/Poultry, Produce dan Seafood.

Data order yang tercatat adalah periode Agustus 1994 – Juni 1996.

Dataset

Dataset dapat Anda unduh melalui canvas

Deskripsi Tugas

Manajemen Perusahaan Northwind ingin menjaga loyalitas pelanggan, memperluas pasar dan meningkatkan penjualan. Salah satu aspek yang ingin mereka ketahui adalah keadaan penjualan selama periode data (Agustus 1994 – Juni 1996) dan perkiraan penjualannya pada Juli-Desember 1996 dan sepanjang 1997 (proyeksi data).

1. Kategori produk apa yang mencatatkan nilai penjualan tertinggi selama periode data? Berapakah angka proyeksi penjualannya? (NIM Genap: Juni-Desember 1996; NIM Ganjil: sepanjang 1997)
Produk apa dalam kategori itu yang mencatatkan penjualan tertinggi?
2. Tiga kategori yang mencatatkan rata-rata nilai penjualan tertinggi selama periode data?
3. (NIM Genap) periode Januari-Juni 1996. (NIM Ganjil) periode selama tahun 1995.
Lima produk yang mencatatkan rata-rata nilai penjualan tertinggi selama periode tersebut?
Lima produk yang mencatatkan rata-rata nilai penjualan terendah selama periode tersebut?
4. Pada proyeksi data, kategori produk mana saja yang memiliki kecenderungan penjualan rata-rata bulanan naik dan mana yang cenderung menurun?
5. (NIM Genap) Wilayah (*country*) mana yang paling banyak melakukan order (*count*) dan paling tinggi nilai order (*sum*) selama periode data?
(NIM Ganjil) Wilayah (*country*) mana yang memiliki trend order (*count*) meningkat dan menurun selama periode data?
6. Apakah ada kaitan/relasi antara kategori produk (8 kategori) dengan asal *customer* (21 *country*)?
7. Apakah ada pola *order* kategori produk sepanjang periode data?
8. Kelompokkan wilayah (*country*) asal *customer* menjadi 5 kelompok dengan parameter 8 kategori produk berdasarkan nilai total *order* tiap kategori.

Mereka meminta bantuan Anda untuk menghasilkan informasi sebagai dasar pihak manajemen melakukan analisa bisnis. Untuk itu, lakukan proses *data mining* menggunakan pendekatan yang sesuai. Baik klasifikasi, klustering, regresi maupun proses statistik biasa (seperti heat-map). Bebas menggunakan alat bantu (excel, rapidminer, python, dll).

Proyek B (4 poin)

Sebuah toko aksesoris retail online ingin menganalisa perilaku konsumen dalam membeli barang yang mereka tawarkan, sehingga mereka dapat mengatur display barang dengan baik untuk kenyamanan konsumen dan mengatur promo dengan lebih tepat. Mereka meminta bantuan Anda menganalisis data untuk tujuan tersebut. Bebas menggunakan alat bantu (excel, rapidminer, python, dll).

Dataset

Dataset terkait dapat Anda unduh melalui canvas

Referensi

- <https://www.datacamp.com/tutorial/market-basket-analysis-r>
- <https://gifadn.medium.com/market-basket-analisis-mba-dengan-menggunakan-datasets-groceries-di-r-62f63f0278c4>
- <https://www.kaggle.com/code/aslanahmedov/market-basket-analysis-apriori-algorithm>

Proyek C (4 poin)

Suatu perusahaan peralatan khusus ingin melakukan segmentasi terhadap produk yang dijualnya sehingga manajemen akan lebih bisa tepat memilih strategi bisnis dan melakukan perencanaan marketing. Data yang telah direkap untuk Tahun 2021 ditampilkan pada tabel di bawah.

Terlihat pada tabel tersebut bahwa Produk B01 total 18 unit terjual yang tercatat dalam 5 buah invoice dengan total nilai produk 2,330 milyar rupiah dan mencatatkan keuntungan bersih 685 juta rupiah. Mereka meminta bantuan Anda menganalisis data sehingga setiap produk dapat ditentukan masuk ke golongan klaster tertentu dengan menggunakan Fuzzy C-Means. *Output* yang diharapkan berupa empat buah klaster, yaitu best-gain (terjual banyak dan untung), best-sales (dibeli banyak orang dan untung), normal (tidak banyak terjual namun untung), poor (kurang banyak terjual dan kurang memberi keuntungan). Buat kesimpulan berupa karakteristik setiap klaster. Anda hanya boleh menggunakan excel atau program lainnya tanpa penggunaan library khusus fcm.

Tabel Rekap Penjualan perusahaan XYZ
 Tahun 2021

Product ID	Number of Invoices	Number of products	Amount (IDR Mio)	Net Profit (IDR Mio)
B01	5	18	2330	685
B02	4	92	2079	286
B03	8	75	2306	506
B04	10	51	1843	840
B05	6	61	2320	993
B06	8	63	1334	855
B07	4	65	2891	281
B08	6	14	2716	624
B09	5	66	2305	611
B10	9	100	1601	747
B11	10	61	2844	215
B12	14	49	1084	283
B13	7	46	1931	251
B14	8	100	2927	934
B15	4	90	1071	870
B16	12	38	1118	631
B17	18	89	2669	818
B18	24	98	2597	927
B19	32	36	2485	697
B20	11	40	1359	572
B21	8	83	2499	208
B22	5	20	1520	216
B23	16	52	2618	521
B24	5	12	1998	527
B25	18	33	2139	282

Dataset

Dataset terkait dapat Anda unduh melalui canvas

Proyek D (4 poin)

Sebuah toko aksesoris retail online ingin melakukan segmentasi terhadap produk yang dijualnya sehingga manajemen akan lebih bisa tepat memilih strategi bisnis dan melakukan perencanaan marketing. Mereka meminta bantuan Anda menganalisis data untuk tujuan tersebut. Bebas menentukan jumlah kluster, bebas menggunakan alat bantu (weka, rapidminer, python, dll). *Output* yang diharapkan kluster produk dapat di *profiling* dengan baik.

Dataset

Dataset yang digunakan sama dengan proyek B.

Referensi:

- <https://www.kaggle.com/code/hossamrizk/wine-clustering-using-fcmeans>.

Laporan

1. Susunlah temuan sedemikian sehingga sesuai tujuan yang ada dalam proyek yang Anda pilih.
2. Tampilkan tabel/visualisasi data pada setiap poin temuan.
3. Kesimpulan yang berisi hal/komentar/rekomendasi yang Anda anggap paling penting terkait temuan tersebut.

Laporan dibuat dan dikumpulkan dalam bentuk zip yang berisi:

- Folder berisi code/file analisa
- Laporan, dalam bentuk PDF

Submit sebelum 24 Maret 2023, Pk. 17.00 WIB.

Penilaian

Aspek Penilaian: (Max Total 20 point)	Capaian yang diharapkan
1. Northwind (16 poin) <ul style="list-style-type: none"> Ketepatan penyiapan data (2) Ketepatan penggunaan metode klasifikasi (1.5) Ketepatan penggunaan metode klustering (1.5) Ketepatan penggunaan metode regresi (1.5) Ketepatan penggunaan proses statistik biasa (1.5) Kelengkapan item temuan (2) Penjelasan temuan (2) Penyajian tabel/grafik sesuai dan relevan dengan penjelasan temuan (2) Kejelasan kesimpulan (2) 	<ul style="list-style-type: none"> <i>Sub-CPMK-3: Mampu menjelaskan bagaimana memahami dan menyiapkan set data.</i> <i>Sub-CPMK-4: Mampu menjelaskan dan menerapkan metode klasifikasi dan prediksi</i> <i>Sub-CPMK-5: Mampu menjelaskan dan menerapkan metode relasi antar catatan</i>
2. Online Retail-Market Basket (4 poin) <ul style="list-style-type: none"> Ketepatan penggunaan metode asosiasi (1.5) Penjelasan temuan dan dilengkapi tabel/grafik yang sesuai (1.5) Kejelasan kesimpulan berupa karakteristik kluster (1) 	<ul style="list-style-type: none"> <i>Sub-CPMK-5: Mampu menjelaskan dan menerapkan metode relasi antar catatan</i>
3. Segmentasi Produk-Kluster (4 poin) <ul style="list-style-type: none"> Ketepatan penggunaan metode kluster (1.5) Penjelasan temuan dan dilengkapi tabel/grafik yang sesuai (1.5) Kejelasan kesimpulan berupa karakteristik kluster (1) 	<ul style="list-style-type: none"> <i>Sub-CPMK-5: Mampu menjelaskan dan menerapkan metode relasi antar catatan</i>
4. Online Retail-Klustering (4 poin) <ul style="list-style-type: none"> Ketepatan penggunaan metode kluster (1.5) Penjelasan temuan dan dilengkapi tabel/grafik yang sesuai (1.5) Kejelasan kesimpulan berupa karakteristik kluster (1) 	<ul style="list-style-type: none"> <i>Sub-CPMK-5: Mampu menjelaskan dan menerapkan metode relasi antar catatan</i>

~ Selamat merayakan ujian tengah semester ~