

Mata Kuliah : IBDA 3122 Knowledge Discovery
Hari, Tanggal : Jumat, 04 Maret 2022
Waktu : Batas Pengumpulan pk. 12.00 WIB
Sifat Ujian : Proyek

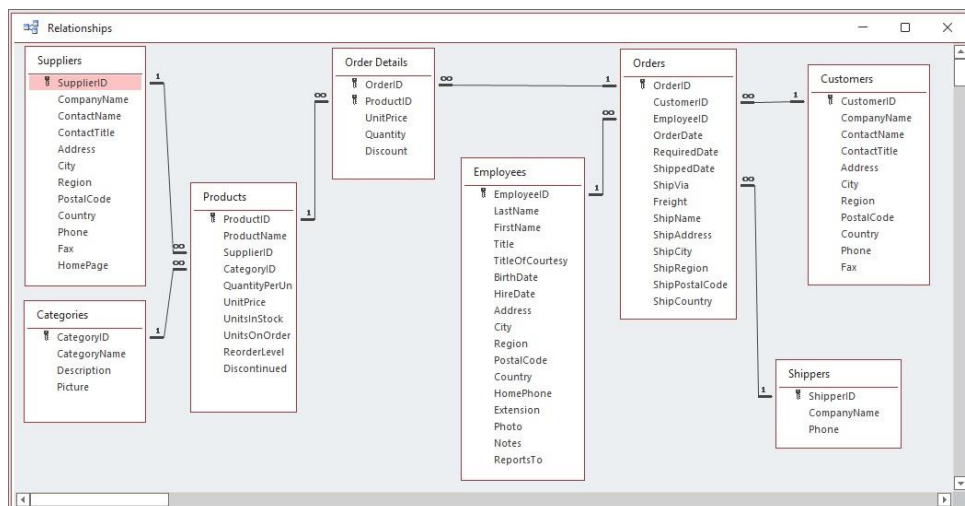
PETUNJUK

- Terdapat 2 Proyek, pilih satu untuk diselesaikan.
- 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

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

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 2017 (proyeksi data).

1. Kategori produk apa yang mencatatkan nilai penjualan tertinggi selama periode data? Berapakah angka proyeksi penjualannya? Produk apa dalam kategori itu yang mencatatkan penjualan tertinggi?
2. Tiga kategori yang mencatatkan rata-rata nilai penjualan tertinggi selama periode data?
3. Sepuluh produk yang mencatatkan rata-rata nilai penjualan tertinggi selama periode data? Sepuluh produk yang mencatatkan rata-rata nilai penjualan terendah selama periode data?
4. Pada proyeksi data, kategori produk mana saja yang memiliki kecenderungan penjualan rata-rata bulanan naik dan mana yang cenderung menurun?
5. Wilayah (*country*) mana yang paling banyak melakukan order (*count*) dan paling tinggi nilai order (*sum*) 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, klastering, regresi maupun proses statistik biasa (seperti heat-map). Bebas menggunakan alat bantu (excel, rapidminer, python, dll).

Laporan:

1. Susunlah temuan sedemikian sehingga sesuai poin yang ada dalam penugasan Anda.
2. Tampilkan tabel/visualisasi data pada setiap poinnya.
3. Kesimpulan yang berisi poin yang Anda anggap paling penting terkait temuan tersebut

Laporan buat dan dikumpulkan dalam bentuk zip yang berisi:

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

Proyek B

Bursa Efek Indonesia (BEI) merupakan istilah yang terkenal pada dunia saham di Indonesia. Jumlah investor pasar modal sudah mencapai angka 7,49 juta investor per akhir Desember 2021, naik diatas 90% di banding tahun sebelumnya (2019: 2,4 juta; 2020: 3,8 juta). Sektor perusahaan yang terdapat di BEI salah satunya adalah manufaktur. Kontribusi sektor manufaktur dalam Produk Domestik Bruto (PDB) tercatat yang paling besar dibandingkan sektor lainnya yakni sebesar 19,7%, meskipun Indonesia tidak luput dari pandemi Covid-19.

Pada perdagangan saham, harga mengalami fluktuasi naik maupun turun. Pergerakan harga saham pada dasarnya tidak dapat diprediksi secara pasti dan berfluktuasi seiring dengan bertambahnya waktu dan situasi yang berkembang.

Dataset

Dataset dapat Anda unduh melalui canvas

Deskripsi Tugas

Tugas:

Group Investor XYZ ingin berinvestasi saham jangka pendek (1 bulan/30 hari) pada Unilever (UNVR), Indofood (INDF), Astra (ASII), Semen (SMGR). Mereka meminta bantuan Anda untuk melakukan perkiraan nilai saham tersebut.

1. Siapkan data pergerakan harian selama 5 tahun (sampai 22 Feb 22).
2. Lakukan proses data mining pada saham-saham tersebut.
3. Bagi data menjadi 2 bagian, data uji dan data latih. Data 300 hari terakhir sebagai data uji peramalan. Sisanya sebagai data latih.
4. Set model untuk mengambil horison 30 hari. Lembar window ditentukan sendiri. Setting diatur sedemikian rupa sehingga menghasilkan *error* serendah mungkin atau akurasi setinggi mungkin.
5. Lewatkan data uji.
6. Tarik tabel hasil, kemudian buat tabel baru: *date, original, prediction*
7. Evaluasi akurasinya dengan MAPE atau parameter akurasi lainnya (pada rentang data prediksi yang memiliki data *original*). Sedangkan pada data harga prediksi 30 hari (tidak ada data original sebagai pembandingan), tarik garis trend (linier, kuadrat, dll).

Note:

- Akurasi poin 7 (*Fair* <45%, *Good* 45%-65%, *Top* >65%)
- Silahkan menggunakan model yang diinginkan dan alat bantu yang diperlukan (*excel, rapidminer, python*, dll).

Laporan:

1. Susunlah temuan sedemikian sehingga sesuai poin yang ada dalam penugasan Anda.
2. Tampilkan tabel/visualisasi/*screenshot* pendukung pada setiap poinnya.
3. Kesimpulan yang berisi poin penting terkait temuan tersebut

Laporan buat dan dikumpulkan dalam bentuk zip yang berisi:

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