

PENGALIAN DATA

TUGAS KELOMPOK II

CLUSTERING

A. Permasalahan

Tugas kedua berkaitan dengan analisis data melalui proses clustering menggunakan R. Data diperoleh dari bechmark internasional dan terkait dengan sebagian data perusahaan mutinasional untuk bisnis ritel. Data tersebut merupakan set data transaksional yang terjadi antara tanggal 01/12/2009 s.d. 09/12/2011 untuk bisnis ritel online non-toko di UK. Perusahaan utamanya menjual barang-barang cinder mata. Kebanyakan kastemer dari perusahaan adalah pedagang grosir.

B. Deskripsi data

Dalam tabel berikut diberikan deskripsi dari data transaksi bisnis ritel tersebut.

Variable	Description
InvoiceNo	Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation
StockCode	Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product
Description	Product (item) name. Nominal
Quantity	The quantities of each product (item) per transaction. Numeric
InvoiceDate	Invoice Date and time. Numeric, the day and time when each transaction was generated
UnitPrice	Numeric, Product price per unit in pound sterling
CustomerID	Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer
Country	Country name. Nominal, the name of the country where each customer resides

C. Tugas

1. Pelajari beberapa referensi mengenai bisnis ritel, sehingga dapat dipahami apa yang menjadi fokus dalam bisnis tersebut, terutama terkait dengan manajemen pelayanan pelanggan (*customer relationship management*).
2. Lakukan eksplorasi data dari berbagai perspektif untuk memahami karakteristik data bisnis ritel tersebut. Gambakan hasil eksplortasi dalam berbagai bentuk grafik/chart yang menurut anda paling sesuai untuk menggambarkan karateristik data.
3. Lakukan praproses data (dapat menggunakan praproses yang disediakan dalam library R atau menggunakan praproses manual/menggunakan *spreadsheet*). Anda juga dapat memodifikasi bentuk data, sehingga sesuai dengan tujuan analisis yang akan dilakukan berdasarkan pemahaman terhadap referensi yang digunakan pada poin nomor 1 (gunakan referensi yang kredibel dan dapat diakses secara online).
4. Lakukan proses clastering menggunakan tiga pendekatan yang berbeda: partisional (K-means), metode hirarki, dan metode berbasis densitas. Carilah library R yang sesuai untuk masing-masing metode. Untuk metode K-means lakukan pebgulangan dengan berbagai titik pusat awal yang berbeda

5. Lakukan perbandingan hasil clustering dari ketiga metode yang digunakan. Berikan penjelasan secukupnya untuk mendukung hasil perbandingan.
6. Untuk hasil analisis, tentukan jumlah klaster yang paling tepat (misalnya menggunakan metode silhouette) dan juga analisis mendalam terhadap masing-masing klaster yang dihasilkan dengan merujuk pada referensi tentang analisis CRM yang digunakan.

D. Laporan dan Batas Waktu

Laporan ditulis pada kertas berukuran A4 dengan spasi tunggal. Laporan dalam format PDF diserahkan per kelompok dan diunggah dalam menu "Assignment" pada aplikasi TEAMS **paling lambat pada tanggal 27 November 2019 pukul 16.00 WIB** (hanya satu orang dari setiap kelompok yang mengunggah). Masukkan *screenshot* dari script R yang digunakan disertai penjelasan seperlunya. Penilaian akan didasarkan pada aspek: sistematika penulisan dan kelengkapan laporan (25%), eksplorasi dan praproses data (25%), dan hasil clustering, ketajaman dan kedalaman analisis, termasuk rujukan terhadap referensi analisis CRM yang digunakan (50%). Tugas ini akan memberikan kontribusi 35% dari keseluruhan nilai tugas mata kuliah. Isi laporan yang mengindikasikan adanya plagiarisme tidak akan dinilai.

-----oooOooo-----