

# PENGGALIAN DATA

## TUGAS KELOMPOK III

### ASSOCIATION RULES MINING

#### A. Permasalahan

Tugas ketiga berkaitan dengan analisis data melalui penggalian aturan asosiasi (*association rules mining*) menggunakan R. Data yang digunakan tepat sama dengan data yang digunakan untuk tugas klasifikasi (pemasaran perbankan). Data tersebut dapat digunakan untuk menganalisis asosiasi antar beberapa atribut non-kelas dengan atribut kelas disertai dengan statistik beberapa atribut bertipe numerik. Seperti halnya tugas kelompok sebelumnya, sangat diharapkan bahwa tugas kelompok ini dikerjakan sendiri-sendiri dalam kelompoknya dan TIDAK berusaha untuk menyontek hasil pekerjaan dari kelompok lain, sehingga tugas ini dapat memberikan manfaat yang semaksimal mungkin. Adanya pertanyaan teknis penggunaan R sehubungan dengan implementasi tugas ini dapat dikonsultasikan dengan Asisten tutorial.

#### B. Deskripsi data

**1. Title:** Bank Marketing

**2. Sources**

Created by: Paulo Cortez (Univ. Minho) and Sérgio Moro (ISCTE-IUL) @ 2012

**3. Relevant Information:**

The data is related with direct marketing campaigns of a Portuguese banking institution (from May 2008 to November 2010). The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be (or not) subscribed.

**4. Number of Instances:** 45.211

**5. Number of Attributes:** 16 + class label

**6. Attribute information:** lihat tabel berikut

No	Nama Atribut	Nilai Atribut
1	age	numeric
2	job	type of job (categorical) : admin., unknown, unemployed, management, housemaid, entrepreneur, studen, blue-collar, self-employed, retired, technician, services
3	marital	marital status (categorical): married, divorced, single; note: divorced means divorced or widowed
4	education	categorical: unknown, secondary, primary, tertiary
5	default	has credit in default? (binary: yes, no)
6	balance	average yearly balance, in euros (numeric)
7	housing	has housing loan? (binary: yes, no)
8	loan	has personal loan? (binary: yes, no)
9	contact	contact communication type (categorical: unknown, telephone, cellular)
10	day	last contact day of the month (numeric)
11	month	last contact month of year (categorical: jan, feb, mar, ..., nov, dec)
12	duration	last contact duration, in seconds (numeric)
13	campaign	number of contacts performed during this campaign and for this client (numeric, includes last contact)
14	pdays	number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)
15	previous	number of contacts performed before this campaign and for this client (numeric)
16	poutcome	outcome of the previous marketing campaign (categorical): unknown, other, failure, success)
17	Class label	has the client subscribed a term deposit? (binary: yes, no)

### C. Tugas

1. Lakukan eksplorasi data dari berbagai perspektif untuk memahami karakteristik data. Gambarkan hasil eksplorasi dalam berbagai bentuk grafik/chart yang menurut anda paling sesuai untuk menggambarkan karakteristik data.
2. Lakukan praproses data (dapat menggunakan praproses yang disediakan dalam library R atau menggunakan praproses manual/menggunakan *spreadsheet*). Praproses difokuskan untuk mentransformasikan data agar dapat diperlakukan sebagai item, sehingga dapat dilakukan analisis asosiasi.
3. Lakukan proses pembangkitan *frequent itemsets* dengan menggunakan algoritma *apriori* dan *FP-growth*. Lakukan perbandingan yang diperoleh menggunakan kedua algoritma tersebut. Gunakan library R yang tersedia untuk keduanya. Lakukan uji coba untuk berbagai nilai ambang batas *support* dan tentukan nilai ambang batas *support* yang pas menurut hasil uji coba anda.
4. Bangkitkan sejumlah aturan asosiasi (*association rules*) yang menarik dari satu set *frequent itemsets* yang diperoleh sebelumnya, di mana atribut "*class-label*" sebagai target bersama-sama dengan atribut lain dengan tipe numerik, seperti rata-rata usia (*age*) dan/atau rata-rata saldo tahunan (*balance*) sebagai bagian dari target. Lakukan analisis kemenarikan (*interestingness*) dari aturan yang dihasilkan menggunakan berbagai ukuran kemenarikan (selain hanya menggunakan *confidence*). Lakukan uji coba untuk berbagai nilai ambang batas ukuran kemenarikan dan buat kesimpulan dari hasil uji coba tersebut.

### 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 hari Minggu, tanggal 17 Mei 2020 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 (50%). Tugas ini akan memberikan kontribusi 25% dari keseluruhan nilai tugas mata kuliah. **Isi laporan yang mengindikasikan adanya plagiarisme tidak akan dinilai.**

-----oooOooo-----