

Soal Kuis Data Mining 2020
Kelas Selasa Siang (P2) Lab 1
Kode soal: **Ganjil**

1. Buka lah **diabet.csv** pada R! Lakukan eksplorasi data tersebut! (Lampirkan screenshot kode R serta jelaskan hasilnya)
2. Lakukanlah tahapan praproses terhadap data! (Lampirkan screenshot kode R serta jelaskan hasilnya)
 - a. Pengisian missing value.
 - b. Lakukanlah normalisasi pada atribut **plas** dan **insu** , menggunakan metode **min max**.
 - c. Reduksi Dimensi (Pilihlah atribut yang menurut Anda penting atau berpengaruh dan hapus atribut yang tidak digunakan).
3. Lakukanlah tahapan clustering terhadap data hasil praproses! (Tahapan yang dilakukan pada clustering yaitu clustering menggunakan **K-Means** (terhadap atribut plas dan insu) dan **K-Medoids** dengan **k=5** serta tampilkan hasil plotnya!)

Deskripsi data:

Preg	: Jumlah kali hamil
Plas	: Konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral
Pres	: Tekanan darah diastolik (mmHg)
Skin	: Ketebalan lipatan kulit triceps (mm)
Insu	: Insulin serum 2 jam (mu U / ml)
Mass (bmi)	: Indeks massa tubuh (berat dalam kg / (tinggi dalam m) ^ 2)
Pedi	: Fungsi silsilah diabetes
Age	: Umur (tahun)
Class	: Variabel kelas

Ketentuan Pengumpulan:

1. Jawaban dituliskan dalam file .doc atau .docx dilengkapi dengan **nama, nim, tipe soal, dan nomor meja**
2. **Laporan dan kode R** diunggah dalam LMS sesuai jadwal yang ditentukan.
3. Ketentuan nama fail laporan: Kuis1_Paralel<kelas paralel>_NIM_<Tipe Soal>.docx, contoh: Kuis1_Paralel1_G64160039_Ganjil.docx
4. Ketentuan nama Zip: Kuis1_Paralel<kelas paralel>_NIM_<Tipe Soal>.zip, contoh: Kuis1_Paralel1_G64160039_Ganjil.zip