



**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS GADJAH MADA**

---

**QUIZ**

<b>Mata Kuliah</b>	<b>: Pembelajaran Mesin</b>
<b>Hari, tanggal</b>	<b>: Sabtu, 01 Juni 2019</b>
<b>Waktu</b>	<b>: Sabtu 9.00 – Minggu 24.00</b>
<b>Dosen</b>	<b>: Drs. Edi Winarko, MSc., PhD</b>

---

Quiz ini dikerjakan sendiri-sendiri. Gunakan bahasa pemrograman Python untuk menjawab pertanyaan di bawah ini. Tuliskan jawaban pada MS Word/Open Office, kemudian unggah jawaban dan program Python ke Elisa. Semua soal di bawah ini dijawab menggunakan File *Heart.csv*, yang dapat diunduh di Elisa.

---

1. Buat summary dari atribut-atribut bertipe kontinyu pada data Heart.csv dengan cara menampilkan nilai-nilai *count, mean, standard deviation (std), min, quartiles and max* dari atribut-atribut tersebut.
2. Tentukan jumlah missing value dari masing-masing atribut. Replace semua missing value.
3. Atribut klas dari data ini adalah atribut AHD. Tentukan jumlah data untuk masing-masing klas (klas YES dan NO).
4. Menggunakan *k-cross validation*, dimana  $k = 5$ , tentukan nilai accuracy, Precision, Recall dari model yang dibuat menggunakan algoritma Random Forest, AdaBoost, dan Gradient Boosting.
5. Tentukan nilai **K** terbaik jika anda melakukan klastering data Heart.csv menggunakan algoritma **K-means** dengan mengambil nilai  $k = 2, 3, 4, 5, 6$ . Buat plot untuk nilai SSE (sumbu-Y) terhadap nilai **K** (sumbu-X).