# SOAL RESPONSI : BIG DATA ANALYTIC

# JURUSAN/KELAS : TI / 3

# Hari, Tanggal ` : Rabu, 16 Juni 2021, Pukul 15.00 WIB

# Waktu Pengumpulan : Rabu, 16 Juni 2021,Pukul 20.00 WIB(via ELA)

# Dosen : Maria Mediatrix S., S.Kom.,M.Eng.

|  |
| --- |
|  |

1. Gunakan file ‘insurance1.csv”, lakukan clustering menggunakan metode K-means untuk mengetahui berapa banyak jumlah perokok berdasarkan usia (variabel smoker dan age). Kemudian tampilkan hasil clustering (gunakan library matplotlib). Note : karena kolom smoker isinya adalah string, maka lakukan konversi string terlebih dahulu menjadi integer (angka). Lakukan normalisasi data terlebih dulu agar rentang nilai untuk setiap kolom tidak terlalu jauh (gunakan fungsi minmaxscaler saja)

Note:

* Lakukan proses apapun yang anda anggap bisa ditambahkan (selain yang disebutkan diatas) agar hasilnya bisa maksimal ya..
* File jawaban yang dikumpulkan dalam bentuk .ipynb atau .py
* Source code harus yang sudah di-running dan ada hasilnya