

## UAS TI4141 Analitika Data

Dikerjakan Secara Individu, Deadline: Selasa, 12 Mei 2020 Pukul 23.55 WIB

Dikumpulkan di LMS ([kuliah.itb.ac.id](http://kuliah.itb.ac.id))

### INSTRUKSI:

UAS ini dikerjakan secara individu. Di awal laporan, harap tulis ulang pernyataan berikut:

*“Saya mengerjakan UAS ini secara individu, tanpa menerima bantuan dari mahasiswa TI4141 ataupun individu lainnya. Saya memahami bahwa melakukan tindak plagiarisme dan kerja sama dalam pengerjaan UAS ini adalah terlarang, dan jika saya terbukti melakukan upaya plagiarisme dan kerja sama, saya bersedia menerima sanksi sesuai dengan kebijakan ITB.”*

Pengumpulan UAS dilakukan sesuai dengan ketentuan berikut:

- Dokumen rekomendasi dikumpulkan dalam bentuk **powerpoint dan laporan berisi narasi penjelasan atas analisis yang sudah dilakukan** ke situs LMS.
- Script dari R maupun Python dikumpulkan via email ke [ajidarmap@itb.ac.id](mailto:ajidarmap@itb.ac.id), dengan subject email *“Script Tubes TI4141 NIM\_Nama”*

### SOAL.

Anda adalah seorang konsultan di Kementerian Kesehatan RI. Pemerintah Indonesia memberikan dua buah dataset terkait penyebaran virus COVID-19 di Italia untuk Anda analisis, agar negara kita dapat menyikapi pandemi ini dengan lebih baik ketimbang Italia. Tugas Anda adalah menganalisis data tersebut dengan berbagai tools analitika data dan menggunakan kerangka umum analisis data, misalnya CRISP-DM, sehingga didapat pemahaman atau insights yang dapat menjadi dasar rekomendasi bagi *policymaker* pemilik data tersebut. Langkah-langkah yang perlu dilakukan meliputi, namun tidak terbatas pada:

1. Lakukan identifikasi dan perumusan masalah yang dinyatakan dalam rumusan masalah umum dan beberapa turunan rumusan permasalahan (*sub-problem*) yang lebih spesifik,
2. Jelaskan strategi analisis yang diambil; bagaimana metode/task analitika data yang Anda lakukan mampu menyelesaikan masalah (*sub-problem*) tersebut.
3. Jelaskan proses transformasi data dan *data pre-processing* yang dilakukan.
4. Lakukan proses analisis data, baik dengan menggunakan model *supervised* dan atau *unsupervised learning*, mulai dari fase *training* model hingga evaluasi model.
5. Gunakan *tools* visualisasi data untuk memperjelas hasil analisis data yang dilakukan.
6. Simpulkan tugas besar ini dalam bentuk poin rekomendasi yang Saudara berikan pada *stakeholders* terkait di Indonesia, beserta implikasi manajerial dari rekomendasi tersebut.

Penjelasan variabel di dokumen *province\_data.csv*

No	Variable	Description
1	Date	The Date of the Medical Detection
2	Region_Code	The ISTAT Regional Code provided by the Government
3	Region_Name	The Name of the Italian Region
4	Province_Code	The ISTAT Provincial Code provided by the Government
5	Province_Name	The Name of the Italian Province
6	Province_GeoCode	The ISTAT GeoCode
7	Province_Latitude	The Province main Latitude
8	Prov_Longitude	The Province main Longitude
9	Total_Cases	Total Number of New Confirmed Cases

Penjelasan variabel di dokumen *regional\_data.csv*

No	Variable	Description
1	Date	The Date of the Medical Detection
2	Region_Code	The ISTAT Regional Code provided by the Government
3	Region_Name	The Name of the Italian Region
4	Region_Latitude	The Region main Latitude
5	Region_Longitude	The Region main Longitude
6	Hospitalized_with_Symptoms	The Total Number of People Hospitalized with Syntoms
7	Intensive_Care	The Total Number of People in Intensive Care for that day in that Region
8	Total_Hospitalized	The Total Number of People Hospitalized for that day in that Region
9	Tot_Home_Isolation	The Total Number at Home Isolation for that day in that Region
10	Total_Actually_Positive	The Total Number of Positive People for that day in that Region
11	New_Actually_Positive	The Number of New People Actually Positive for that Day in that Region
12	Healed	The Total Number of Healed People
13	Deceased	The Total Number of Deceased People
14	Total_Cases	The Total Number of Cases
15	Total_Medical_Swabs	The Total Number of Medical Swabs done to the Population