

UTS MR4103 Sains Data dan Pembelajaran Mesin

Dikerjakan Secara Mandiri, Deadline: Senin, 28 Oktober 2019 Pukul 23.59 WIB

Dikumpulkan ke ajidarmap@itb.ac.id

Instruksi: Ujian ini wajib dikerjakan secara individu. Anda tidak boleh berdiskusi dengan mahasiswa lain di kelas MR4103 tentang aspek apapun dari kelas selama periode waktu ujian ini. Melakukan hal tersebut akan berakibat pemberian nilai nol untuk UTS bagi pihak yang terlibat. Anda pun diharapkan tidak meminta saran ataupun bantuan dari orang lain saat mengerjakan ujian ini. Pengumpulan dokumen UTS dikirimkan via email ke ajidarmap@itb.ac.id, berisikan 4 file:

- 13410094-Bagian1.pdf
- 13410094-Bagian2.R
- 13410094-Bagian2.pdf
- 13410094-Bagian2.RData

dengan subjek email "NIM_Anda-UTSMR4103", Contoh: 13410094-UTSMR4103.

Bagian 1. Anggap Anda sekarang menjadi CEO sebuah *online travel agent* (OTA) di Indonesia (bisa dibayangkan perusahaannya *Traveloka* atau *Tiket.com*). Bisnis OTA sangat dinamis dan memiliki tingkat kompetisi yang sangat tinggi akhir-akhir ini. Kebanyakan strategi pemasaran yang diambil masih menggunakan strategi promosi untuk mencapai harga terendah. Padahal strategi ini tidak sustainable dalam jangka panjang. Anda tidak ingin perusahaan Anda terjebak dalam perang promosi dan diskon. Untuk itu Anda menyatakan bahwa perusahaan Anda tidak akan masuk ke mass market tetapi lebih ke traveler, baik bisnis maupun liburan, yang ingin mendapatkan value maksimum dari waktu dan uang yang mereka keluarkan (tidak harus harga termurah).

Buatkan narasi singkat bagaimana Anda dapat menggunakan data science (DS) dan machine learning (ML) untuk membantu Anda menghasilkan strategi dan taktik pemasaran yang tepat. Anda bisa mengikuti langkah-langkah umum sebagai berikut:

- Mendefinisikan business problem
- Menentukan task bagi DS/ML
- Menentukan jenis data yang akan digunakan (bentuk, sumber, cara memperoleh)
- Memberikan gambaran umum pemrosesan data sehingga diperoleh output bagi perumusan strategi

Maksimum narasi, termasuk grafik, tabel, adalah 2 halaman. Dikumpulkan dalam satu file pdf dengan nama "NIM_Anda-Bagian2.RData", contoh: "13410094-Bagian1.pdf"

Bagian 2. Anda diminta mengolah data ‘Midterm_FittingData.csv’. Tugas Anda adalah menganalisis data tersebut dengan berbagai tools sains data dan pembelajaran mesin, hingga didapat sebuah rekomendasi bagi *policy maker* pemilik data tersebut. Ikuti petunjuk di bawah ini dengan cermat:

1. Penggunaan tools harus sesuai dengan yang diajarkan dalam kelas atau textbook.
2. Anda harus mengirimkan file. RData berisikan seluruh objek hasil pengolahan data yang sudah Anda lakukan dalam UTS ini. File RData Anda harus disimpan dalam format seperti "NIM_Anda-Bagian2.RData", contoh: "13410094-Bagian2.RData"
3. Anda harus mengirimkan skrip R Anda, menamai file Anda seperti yang dijelaskan di atas, contoh: "13410094-Bagian2.R"
4. Anda harus mengirimkan laporan tertulis, menamai file Anda seperti yang dijelaskan di atas (format PDF, font 12 poin, spasi tunggal, margin 1 inci), contoh: "13410094-Bagian2.pdf". yang berisikan analisis sebagai berikut:
 1. Jelaskan proses Anda mengolah data dengan model dan tools yang diajarkan. Jika Anda menggunakan transformasi data, Anda perlu mendiskusikannya dengan jelas. Bab I ini tidak boleh melebihi 2 halaman.
 2. Tuliskan model prediktif akhir yang Saudara gunakan, termasuk persamaan atau gambar yang merangkum model akhir atau model akhir dengan semua parameter yang ditentukan secara numerik. Bab II ini tidak boleh melebihi ½ halaman.
 3. Jelaskan mengapa Anda memilih model akhir yang Anda lakukan. Bab III ini tidak boleh melebihi 1 halaman.
 4. Lampirkan grafik dan diagram pendukung dalam Bab Lampiran, jumlah halaman tidak dibatasi.

Terlampir file data ‘Midterm_FittingData.csv’. Deskripsi terkait dataset terlampir pada halaman ketiga di soal ujian ini.

Description of Variables in the Data Set:

Variables	Description
FIPS	FIPS code
region	State name
County	The name of county
ill PREV	Percentage of adults with a given health issue in a given county
LTPIA_PREV	Physical activity rate
SNAP Authorized stores	SNAP Authorized stores per 1000 individuals
WIC Authorized Stores	WIC Authorized stores per 1000 individuals
REDEMP_SNAPS12	SNAP redemptions in 2012
PCT_SNAP14	SNAP participants (%pop) in 2014
PCT_NSLP14	National school lunch program participants (%pop) in 2014
PCT_SBP14	School breakfast program participants (%pop) in 2014
PCT_SFSP14	Summer food service program participants (%pop) in 2014
PC_WIC_REDEMP12	Percentage WIC redemptions in 2012
REDEMP_WICS12	WIC redemptions in 2012
PCT_WIC14	WIC participants (%pop) in 2014
PCT_CACFP14	Child and Adult Care (%pop) in 2014
FDPIR12	Food Distribution Program on Indian Reservations 2012
health.RECFAC12	Recreation and Fitness Facilities in 2013
health.RECFACPTH12	Recreation and Fitness Facilities/1000 pop in 2012
restaurants.FFR12	Fast-food restaurants in 2012
restaurants.FFRPTH12	Fast-food restaurants per thousand in 2012
restaurants.FSR12	Full service restaurants in 2012
restaurants.FSRPTH12	Full service restaurants per thousand in 2012
GROC12	Grocery stores in 2012
GROCPH12	Grocery stores per thousand 2012
SUPERC12	Super centers in 2012
SUPERCPTH12	Super centers per thousand in 2012
CONVS12	Convenient stores in 2012
CONVSPTH12	Convenient stores per thousand in 2012
SPECS12	Specialized food stores in 2012
SPECSPTH12	Specialized food stores per thousand in 2012
SNAPS12	SNAP in 2012
SNAPSPTH12	SNAP per thousand in 2012
WICS12	WIC in 2012
WICSPTH12	WIC per thousand in 2012
foodstamps	% of households receiving food-stamps
percent.white	Ratio of number of white to total population
Pop.density	Population Density Per square mile
age.Estimate	Median age of the population
education.percent	% Population with at least X level education
unemployment	% of labor force unemployed
MF Ratio	Male to female ratio
maritalStatus	% of married individuals
drive.alone	% Population driving alone
povertyRate	County level poverty rate
medianIncome	Median household income
Primary.RUCA.Code.Mode	Mode of 2010 Rural-Urban Commuting Area (RUCA) Codes for a given county