

Mata Kuliah : IBDA 3111, Prapemrosesan dan Rekayasa Data
Hari, Tanggal : -
Waktu : -
Sifat Ujian : Proyek

PETUNJUK

- Dalam naskah ini terdapat 1 soal esai.
- Kerjakanlah soal-soal di bawah ini dan dikumpulkan sesuai dengan keterangan yang ada di dalam soal
- Setelah selesai mengerjakan ujian, salinlah pernyataan komitmen integritas berikut ini di bagian laporan dan bubuhkanlah tanda tangan di bawahnya:

“Di hadapan TUHAN yang hidup, saya menegaskan bahwa saya tidak memberikan maupun menerima bantuan apapun—baik lisan, tulisan, maupun elektronik—di dalam ujian ini selain daripada apa yang telah diizinkan oleh pengajar, dan tidak akan menyebarkan baik soal maupun jawaban ujian kepada pihak lain.”

1. Soal Nomor 1

Beasiswa LPDP merupakan salah satu beasiswa paling bergengsi di Indonesia. Setiap tahunnya, kurang lebih 10,000 mahasiswa Indonesia penerima beasiswa ini mendapatkan kesempatan untuk melanjutkan studi di universitas ternama baik di dalam dan luar negeri. Ada tiga tahap seleksi untuk mendapatkan beasiswa ini, yakni: seleksi administrasi, seleksi akademik, dan sesi wawancara. Diantara tiga jenis seleksi diatas, biasanya seleksi akademik menjadi hal yang paling sulit bagi calon *awardee*. Kurang lebih 50% peserta yang sudah lolos seleksi administrasi terseleksi kembali di dalam seleksi akademik. Ada tiga bagian di dalam seleksi ini, yakni: tes penalaran, tes kuantitatif, dan tes pemecahan masalah. Diduga hal yang paling berpengaruh terhadap kelulusan dari calon *awardee* di dalam tahap ini adalah skor dari ketiga bagian tersebut.

Berdasarkan hasil survey yang dilakukan oleh rektorat CIT, terdapat kurang lebih 50% mahasiswa IBDA yang berencana untuk studi lanjut dan beasiswa LPDP ini menjadi pilihan yang tepat bagi mereka. Oleh karena itu, mahasiswa IBDA perlu mempersiapkan diri sedini mungkin dengan mempelajari bagaimana pola penerimaan beasiswa LPDP terkhusus di seleksi akademik. Buatlah model pembelajaran mesin dengan menggunakan data “Hasil Tes SBK LPDP Gelombang I – 2021.csv” dan aplikasikan semua teknik prapemrosesan data yang telah dipelajari untuk mendapatkan akurasi setinggi-tingginya,

Keterangan data berdasarkan kolom:

- Output: “Apakah Anda dinyatakan Lulus Seleksi Substansi Akademik Pada Gelombang I Tahun 2021” -> 2 jawaban: Ya dan Tidak
- Input: Semua kolom selain kolom Output*

*Anda tidak diperkenankan untuk membuang salah satu baris atau kolom input jika tidak dibuktikan bahwa kolom tersebut layak untuk dibuang lewat metode prapemrosesan data beserta penjelasan yang logis. Pengabaian hal ini akan mengurangi penilaian.

Kolom Output

Jenis Beasiswa	Tingkat Pendidikan yang Dituju	Negara Tujuan Kuliah	Bidang Studi yang dipilih saat mendaftar	Mendaftar dengan menggunakan LoA sesuai ketentuan LPDP	Apakah anda dinyatakan Lulus Seleksi Substansi Akademik Pada gelombang I Tahun 2021?	Jumlah Jawaban Benar pada Tes Penalaran Verbal	Jumlah Jawaban Benar pada Tes Kuantitatif	Jumlah Jawaban Benar pada Tes Pemecahan Masalah	[optional] Kampus/universitas tujuan saat ini (pilihan 1)
Targeted	S2	Dalam negeri	TRP	Tidak	Ya	10	12		6 Universitas Indonesia
Reguler	S2	Luar Negeri	ENE	Ya, dengan LoA	Ya	17	22		8 UCL
Targeted	S2	Luar Negeri	ELE	Ya, dengan LoA	Ya	16	18		7 KTH Sweden
Reguler	S2	Dalam negeri	LIN	Tidak	Tidak	15	12		4 UGM
Afirmasi	S2	Luar Negeri	LIN	Tidak	Ya	16	9		7 University of Birmingham
Targeted	S2	Luar Negeri	Dev	Ya, dengan LoA	Ya	17	11		4 KU leuven

Kolom Input

Kolom Input

Gambar 1. Ilustrasi isi data "Hasil Tes SBK LPDP Gelombang I – 2021.csv"

Berikut ini adalah hal yang perlu dilakukan:

- Lakukan teknik prapemrosesan data yang sudah pernah diajarkan di dalam kelas untuk mengolah data ini.
- Anda perlu menerapkan minimal 10 buah teknik prapemrosesan data dan minimal 5 jenis model lalu catat akurasi modelnya. Simpan hasil tabel ke dalam file dalam format "Laporan.pdf" lalu sertakan hasil analisa dan penjelasan Anda akan faktor apa yang paling menentukan kelulusan di seleksi akademik bagi calon *awardee*.
- Simpan setiap model yang sudah dibuat di dalam format "<Nama>_<NIM>_Model_<urutan>.ipynb" lalu dikumpulkan di dalam .zip dengan format "IBDA3111_<Nama>_<NIM>_Soal1.zip"

Ilustrasi isi file .zip:

- Lucas_13312024_Model_1.ipynb
- Lucas_13312024_Model_2.ipynb
- .
- .
- Lucas_13312024_Model_5.ipynb
- Laporan.pdf

Semua model di atas disimpan di dalam file "IBDA3111_Lucas_13312024.zip"

Tabel 1. Contoh pengisian tabel

Model ke-	Metode pengolahan data yang dilakukan	Algoritma yang digunakan	Akurasi model
1	1. <i>Feature importance</i> 2. Menerapkan skala pada data numerik 3. Mengubah distribusi menjadi Gaussian	Logistic Regression	90%

	4. ..		
	5. ..		
2			
.			
.			
5			

Q&A: (Mohon dibaca terlebih dahulu sebelum bertanya kepada dosen terkait soal)

Q: Apakah di dalam satu model harus terdapat 10 buah teknik prapemrosesan data?

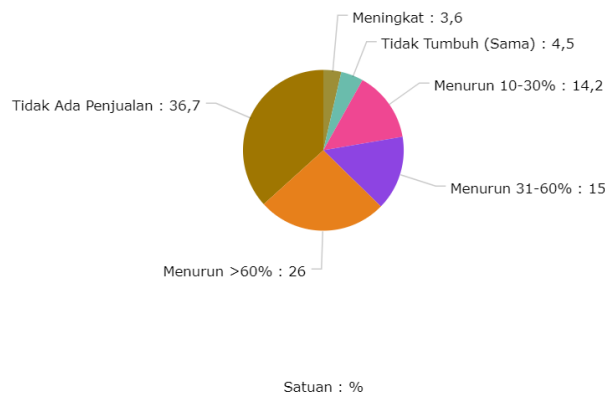
A: Tidak. Ke-10 teknik ini harus terangkum dari total 5 buah model yang dibuat. Jika ada sebuah model yang sudah memiliki 10 teknik prapemrosesan dipersilahkan.

Kriteria	Tidak memuaskan	Kurang memuaskan	Memuaskan	Sangat Memuaskan
	0 poin	3 poin	7 poin	10 poin
Tugas pemrograman: Menerapkan teknik prapemrosesan data	Tidak menerapkan sama sekali	Menerapkan 10 teknik prapemrosesan data dan tidak ada penjelasan sama sekali mengenai penggunaannya	Menerapkan 10 teknik prapemrosesan data dan penjelasan yang kurang detail + akurat mengenai penggunaannya	Menerapkan 10 teknik prapemrosesan data dan penjelasan yang detail + akurat mengenai penggunaannya
	0 poin	3 poin	7 poin	10 poin
Tugas pemrograman: Membuat 5 buah model pembelajaran mesin	Program yang dikumpulkan tidak dieksekusi	Berhasil menunjukkan 3 buah model dengan akurasi masing-masing	Berhasil menunjukkan 4 buah model dengan akurasi masing-masing	Berhasil menunjukkan 5 buah model dengan akurasi masing-masing
	1 poin	3 poin	4 poin	5 poin
Tugas pemrograman: Akurasi dari model yang dibuat	Menunjukkan akurasi dari semua model yang dibuat	Nilai untuk mahasiswa yang mendapatkan akurasi tertinggi ke-3 dibandingkan rekan sekelas	Nilai untuk mahasiswa yang mendapatkan akurasi tertinggi ke-2 dibandingkan rekan sekelas	Nilai untuk mahasiswa yang mendapatkan akurasi paling tinggi dibandingkan rekan sekelas
	2 poin	5 poin	10 poin	15 poin
Tugas laporan: Kerapihan, kejelasan penulisan, dan ketajaman analisa pada file "Laporan.pdf"	Laporan tidak rapi atau sulit dipahami dan analisisnya tidak tajam	Laporan rapi dan mudah dipahami, tetapi analisisnya tidak tajam	Laporan rapi dan mudah dipahami, tetapi analisisnya kurang tajam	Laporan sangat rapi dan mudah dibaca, dan terdapat analisa yang tajam

2. Soal Nomor 2

Di dalam masa pandemi Covid-19 selama hampir dua tahun kebelakang, banyak usaha mikro kecil menengah (UMKM) yang terdampak. Sebuah artikel dari katadata.id menyebutkan hanya 3.6% saja yang masih mengalami pertumbuhan dan 96.4% sisanya stagnan atau mengalami penurunan dengan tingkat yang berbeda-beda.

Penurunan Penjualan Sektor UMKM Akibat Covid-19



Sumber : Asosiasi Business Development Services Indonesia

Gambar 2. Penurunan penjualan UMKM akibat Covid-19

Bersyukur, semenjak dua bulan terakhir, jumlah kasus Covid-19 yang sempat meledak sampai menyentuh angka 52,000/hari sekarang menyusut menjadi 300/hari. Oleh karena itu pemerintah sudah menurunkan tingkat pemberlakuan pembatasan kegiatan masyarakat (PPKM) dari level-4 menjadi level-1. Sebagai mahasiswa CIT, Anda diminta untuk membantu pemerintah di dalam mempercepat proses pemulihan dari bisnis UMKM yang terdampak. Hal yang akan Anda lakukan di dalam tugas kali ini adalah membuat sebuah model pembelajaran mesin yang mampu memberikan rekomendasi lokasi yang tepat untuk membuka restoran padang “Ganto Minang” di daerah Jakarta Selatan. Model pembelajarn mesin ini nantinya akan mempelajari pola *point of intereset* (POI) yang ada di sekitar restoran lalu mencari pola terbaik untuk membuka cabang.



Gambar 3. Restoran padang “Ganto Minang”

Untuk mengerjakan tugas ini, Anda memiliki informasi sebagai berikut:

- Kumpulan grid berbentuk hexagon yang memiliki informasi POI dan restoran Ganto Minang yang ada di dalam grid tersebut. Grid ini berjumlah 306, yang memiliki pola seperti gambar di bawah ini (perhatikan nilai i dari gambar 3 dan 4) selama grid tersebut tidak ada di lokasi ujung. Data terkait grid bisa dilihat di dalam tessellation.csv. Di dalam data csv ini terdapat informasi POI sebagai kolom input dan restoran Ganto Minang sebagai output serta di grid nomor berapa terdapat POI dan restoran tersebut.

Tabel 2. Isi tessellation.csv

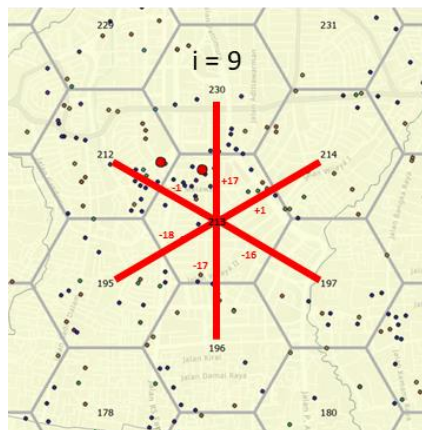
No.grid di dalam hexagon

Kolom Input

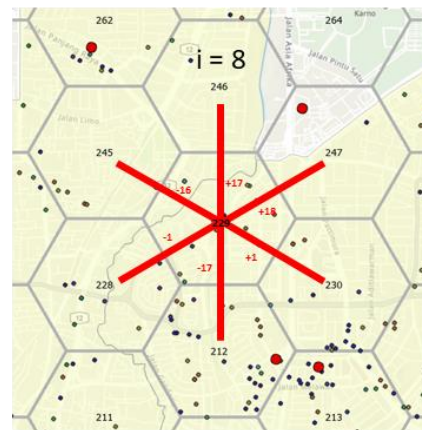
Kolom Output

grid	traffic_station	sport_fix	shopping_mall	school	medical	life_service	landmark	hotel	financial_facilities	factories	cultural	communal_facilities	catering	auto_service	attraction	ganto_minang
177	0	0	0	0	0	0	0	0	1	4	0	0	0	0	0	1
266	0	0	1	0	3	0	0	6	3	0	0	0	3	0	0	1
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

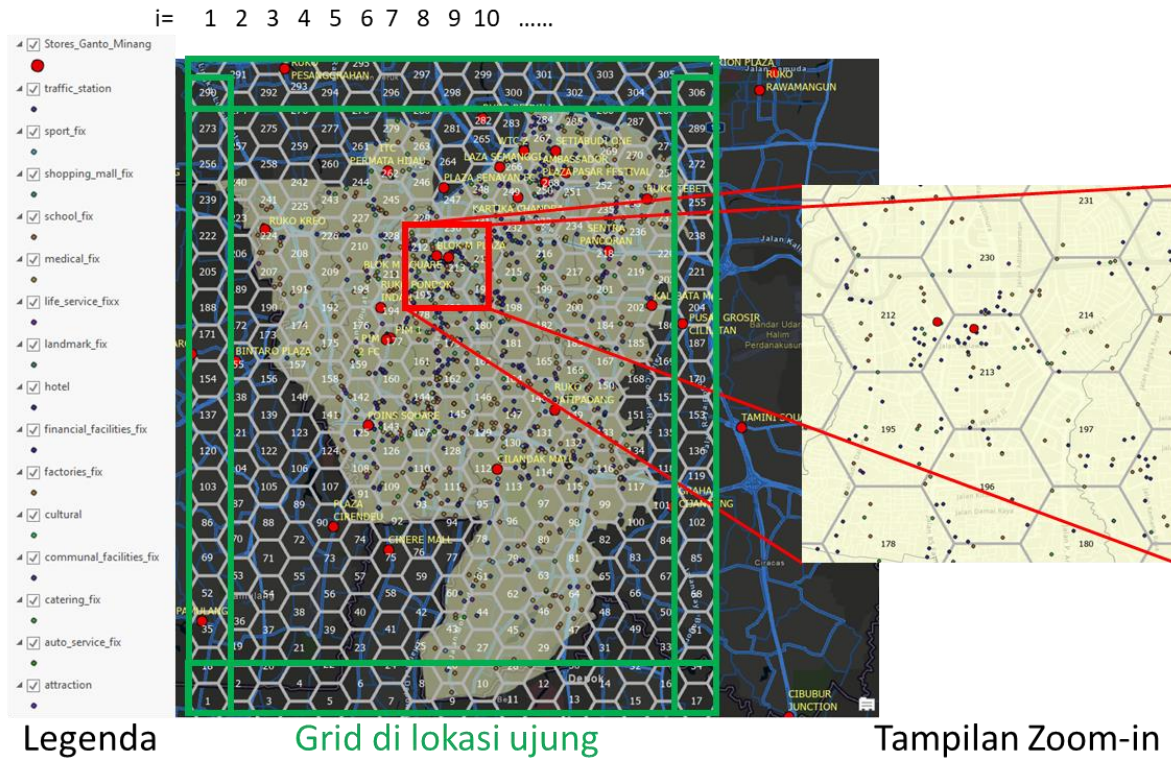
Jika i bilangan ganjil



Jika i bilangan genap



Gambar 4. Ilustrasi penomoran grid jika i bilangan ganjil dan genap



Gambar 5. Ilustrasi hexagon di di Jakarta Selatan

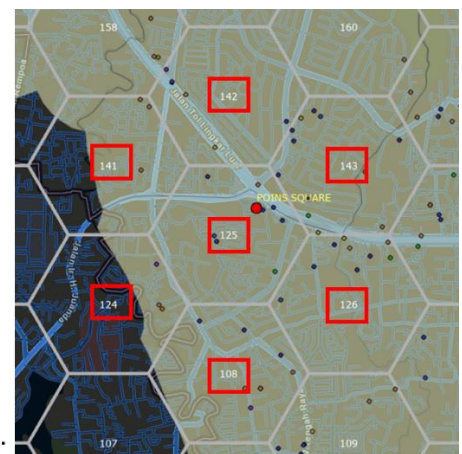
Berikut ini hal yang perlu dilakukan di dalam penilaian:

1. Buatlah tabel seperti di bawah ini untuk semua grid yang berada di Jakarta Selatan dengan menggunakan code di dalam Jupyter Notebook dengan menggunakan informasi dari tessellation.csv.

No. Grid	Ada Ganto Minang?	Nama Ganto Minang	Kategori POI				Attraction
			Traffic Station	Sport Fix		
125	Ya	POINS SQUARE	10	2			5
126	Tidak	-	2	1			8
127	Tidak	-					

Total POI dari restoran Ganto Minang di grid 108, 124, 125, 126, 141, 142, dan 143

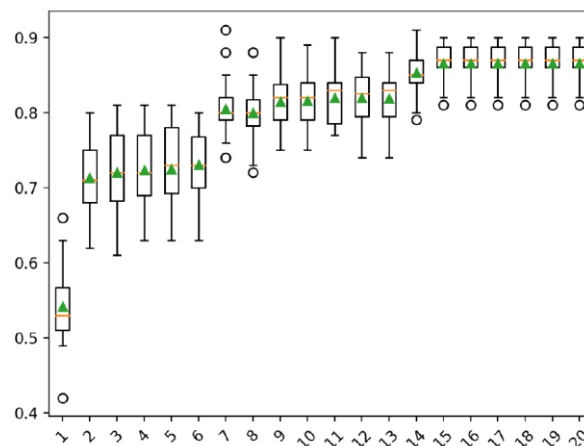
Jika restoran Ganto Minang ada di grid X, maka perlu dihitung total POI yang ada di dalam grid X beserta 6 grid di sekelilingnya.



Gambar 6. Ilustrasi pengisian tabel dari grid

2. Jika tabel pada langkah 1 sudah diperoleh, gunakan kolom "Ada Ganto Minang?" sebagai kolom output dan kolom "Kategori POI" sebagai kolom input.

- Lakukan teknik prapemrosesan data dengan menggunakan PCA dan LDA untuk mengurangi jumlah kolom kategori POI, lalu tentukan jumlah komponen yang menghasilkan akurasi tertinggi baik untuk metode PCA dan LDA. Pastikan Anda menampilkan plot seperti gambar di bawah ini di dalam menentukan jumlah komponen.



Gambar 7. Box Plot dengan sb. x = jumlah komponen dan sb. y = akurasi

- Simpan hasil pekerjaan dari langkah 1-3 di dalam file berikut dengan format <Nama>_<NIM>_Soal2.ipynb.
 Sebagai contoh: "Lucas_13312024_Soal2.ipynb"
- Lakukan analisa sebagai berikut di dalam file "Laporan.pdf":
 - Jika sudah diketahui jumlah komponen yang memberikan akurasi paling maksimal, carilah nilai score masing-masing kategori POI lalu lakukan interpretasi terhadap hasilnya.
 - Bandingkan hasil akurasi model dengan menggunakan PCA dan LDA, jelaskan penyebab perbedaan akurasi diantara keduanya.
- Simpan file ke dalam zip dengan format "IBDA3111_<Nama>_<NIM>_Soal2.zip"

Ilustrasi isi file .zip:

- Lucas_13312024_Soal2.ipynb
- Laporan.pdf

Semua model di atas disimpan di dalam file "IBDA3111_Lucas_13312024_Soal2.zip"

- Gabungkan file soal 1 dan soal 2 ke dalam file dengan format "IBDA3111_<Nama>_<NIM>.zip" lalu kumpulkan melalui Canvas.

Kriteria	Tidak memuaskan	Kurang memuaskan	Memuaskan	Sangat Memuaskan
	0 poin	3 poin	7 poin	10 poin
Tugas pemrograman: Membuat tabel <i>summary</i> dari <i>tesselation.csv</i>	Tidak menerapkan sama sekali	Logika <i>code</i> hampir benar dan hasilnya salah	Logika <i>code</i> benar dan hasilnya salah	Logika <i>code</i> benar dan hasilnya benar

Tugas pemrograman: Penerapan PCA dan LDA	0 poin	3 poin	7 poin	10 poin
	Tidak menerapkan sama sekali	Berhasil menerapkan salah satu diantara PCA dan LDA dan	Berhasil menerapkan PCA dan LDA dan dilakukan dengan konteks yang tidak tepat	Berhasil menerapkan PCA dan LDA dan dilakukan dengan konteks yang tepat
Tugas laporan: Kerapihan, kejelasan penulisan, dan ketajaman analisa pada file "Laporan.pdf"	2 poin	5 poin	10 poin	15 poin
	Laporan tidak rapi <u>atau</u> sulit dipahami <u>dan</u> analisisnya tidak tajam	Laporan rapi <u>dan</u> mudah dipahami, tetapi analisisnya tidak tajam	Laporan rapi <u>dan</u> mudah dipahami, tetapi analisisnya kurang tajam	Laporan sangat rapi <u>dan</u> mudah dibaca, <u>dan</u> terdapat analisa yang tajam
Pengumpulan tugas	0 poin		5 poin	
	Tugas dikumpulkan tidak tepat waktu		Tugas dikumpulkan tepat waktu	