

SOAL UJIAN AKHIR SEMESTER GENAP 2021-2022

Telah diverifikasi:

27 Mei 2022

Desi Arisandi, M.T.I.
Kaprod TI

Prog. Studi : Teknik Informatika Kelas : A, B dan S
Matakuliah : Big Data Tgl : 02 Juni 2022
Dosen : Bagus Mulyawan, S.Kom., MM Waktu : 07:30-09:10 (100 Menit) A,B
Dr. Dedi Trisnawarman 17:00-18:40 (100 Menit) S
Teny Handhayani, Ph.D
Kode MK : TK13025 Sifat : Ujian melalui MS Teams
Catatan : -

Bila dalam pelaksanaan ujian peserta melanggar tata tertib ujian dan berlaku curang, maka nilai ujian langsung diberi nilai E tanpa memperhatikan komponen nilai lain.

Capaian Pembelajaran (CPL) dan Capaian Pembelajaran Mata Kuliah (CPMK)**

	CPL					
	S8	P1	P3	KU1	KK3	KU5
CPMK 01				√		
CPMK 02	√					
CPMK 03					√	
CPMK 04		√				
CPMK 05			√			
CPMK 06						√
CPMK 07					√	

Keterangan:

Capaian Pembelajaran (CPL): *

- S8 : Menunjukkan sikap bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri
P1 : Menguasai prinsip dan metode matematika untuk menyelesaikan masalah komputasi
P3 : Menguasai konsep dan prinsip-prinsip sistem cerdas untuk pengembangan aplikasi sistem cerdas di berbagai bidang
KU1 : Mampu menerapkan pemikiran logis, kritis, sistematis, dan inovatif dalam konteks pengembangan atau implementasi ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora yang sesuai dengan bidang keahliannya
KK3 : Mampu memproses dan menganalisis permasalahan yang berhubungan dengan big data dan mengimplementasikan hasil analisis tersebut dalam suatu aplikasi inovatif yang menunjang kewirausahaan di bidang teknologi informasi
KU5 : Mampu mengambil keputusan secara tepat dalam konteks penyelesaian masalah di bidang keahliannya berdasarkan hasil analisis informasi dan data

Capaian Pembelajaran Mata Kuliah:*

- CPMK 1 : Mahasiswa memahami bagaimana teknologi komputer berkembang, membantu manusia dalam melakukan pekerjaan sehari-hari, dan memungkinkan banyak bidang ilmu dan perkembangan teknologi baru. (KU1).
- CPMK 2 : Mahasiswa dapat mengerjakan tugasnya secara mandiri dengan penuh tanggungjawab (S8)
- CPMK 3 : Mahasiswa memahami dan mampu membangun suatu metode penyimpanan data, teknik optimalisasi, serta berbagai implementasi yang dapat dibangun dari suatu penyimpanan data tersebut. (KK3)
- CPMK 4 : Mahasiswa mampu memahami konsep dasar statistik dan probabilitas dalam komputasi. (P1)
- CPMK 5 : Mahasiswa memahami konsep, algoritma, dan aplikasi machine learning yang dapat dievaluasi secara terukur dan akurat (P3)
- CPMK 6 : Mahasiswa memahami konsep dan teknik analisis dasar Big Data (KU5)
- CPMK 7 : Mahasiswa dapat mengenal dan dapat menggunakan Teknologi Big Data (KK3)

No.	Bentuk Soal (PG/Essay /Kasus)	URAIAN MATERI	CPMK**							BOBOT SOAL (%)
			CPMK 01	CPMK 02	CPMK 03	CPMK 04	CPMK 05	CPMK 06	CPMK 07	
1	PG	Analisis kuantitatif						√		2
2	PG	Analisis kualitatif						√		2
3	PG	Data mining				√				2
4	PG	Analisis statistik				√				2
5	PG	Correlation				√				2
6	PG	A/B testing				√				2
7	PG	Pemahaman analisis statistik				√				2
8	PG	Analisis regresi					√			2
9	PG	Implementasi regresi					√			2
10	PG	Klasifikasi					√			2
11	PG	Clustering					√			2
12	PG	Visualisasi				√				2
13	PG	Filtering						√		2
14	PG	Analisis sentimen					√			2
15	PG	Natural Language Processing (NLP)					√			2
16	PG	Analisis visual						√		2
17	PG	Time series						√		2
18	PG	Analisis jaringan						√		2

No.	Bentuk Soal (PG/Essay/ /Kasus)	URAIAN MATERI	CPMK**							BOBOT SOAL (%)
			CPMK 01	CPMK 02	CPMK 03	CPMK 04	CPMK 05	CPMK 06	CPMK 07	
19	PG	Data spasial						√		2
20	PG	Manfaat analisis jaringan						√		2
21	PG	Penyimpanan disk			√					2
22	PG	Penyimpanan file terdistribusi			√					2
23	PG	Penyimpanan RDBMS			√					2
24	PG	NoSQL			√					2
25	PG	Karakteristik NoSQL			√					2
26	PG	Penyimpanan RDBMS			√					2
27	PG	Contoh database relational			√					2
28	PG	Penyimpanan key-value			√					2
29	PG	Pemahaman penyimpanan key-value			√					2
30	PG	Penyimpanan dokumen			√					2
31	PG	Perangkat penyimpanan dokumen			√					2
32	PG	Sharding			√					2
33	PG	Ciri penyimpanan dokumen			√					2
34	PG	Penyimpanan dokumen			√					2
35	PG	Ciri penyimpanan column family			√					2
36	PG	Teorema CAP			√					2
37	PG	Penyimpanan graph								2
38	PG	Contoh perangkat penyimpanan grafik			√					2
39	PG	Penyimpanan NewSQL			√					2
40	PG	Penyimpanan memory			√					2
41	Essay	NoSQL			√					5
42	Essay	Teknik analisis Big Data						√		5
43	Essay	Teknik Analisis Big Data						√		10

* : CPL dan CPMK disesuaikan dengan yang tertulis di RPS

** : beri tanda ✓

*** : materi pembelajaran yang terkait soal ujian

A. Pilihan Ganda

Jawablah soal – soal berikut dengan benar, **jawaban bisa lebih dari satu pilihan !**

1. Berikut pernyataan benar mengenai analisis kuantitatif yaitu

A	menggunakan sampel yang lebih kecil
B	berfokus pada kuantifikasi pola dan korelasi dalam data
C	melibatkan sejumlah besar data
D	hasil analisis tidak dapat diukur secara numerik
E	hasil analisis dapat digunakan untuk perbandingan numerik

2. Berikut pernyataan benar mengenai analisis kualitatif yaitu

A	hasil analisis tidak dapat digeneralisasikan ke seluruh kumpulan
B	melibatkan sejumlah besar data
C	hasil analisis tidak dapat diukur secara numerik
D	hasil analisis dapat digunakan untuk perbandingan numerik
E	menggunakan sampel yang lebih kecil

3. Berikut pernyataan benar mengenai data mining yaitu

A	memiliki tujuan mengidentifikasi pola yang sebelumnya tidak diketahui
B	digunakan untuk analitik prediktif
C	menggunakan kumpulan data besar
D	menggunakan sampel berukuran kecil
E	menggunakan teknik berbasis perangkat lunak otomatis

4. Analisis statistik selalu bersifat kuantitatif.

A	Benar
B	Salah

5. Correlation membandingkan dua versi elemen untuk menentukan versi mana yang lebih unggul berdasarkan metrik yang telah ditentukan sebelumnya.

A	Benar
B	Salah

6. Berikut pernyataan benar mengenai A/B testing yaitu

A	melibatkan elemen <i>control</i> dan <i>treatment</i>
---	---

B	melibatkan data sampel yang berukuran kecil
C	kedua elemen mengalami percobaan secara bersamaan
D	untuk membandingkan dua produk
E	percobaan tidak dilakukan bersamaan

7. Apa jenis teknik analisis yang tepat untuk mengetahui efektivitas beberapa jenis vaksin untuk yaitu

A	Correlation
B	Regression
C	A/B Testing
D	Clustering
E	Classification

8. Berikut pernyataan benar mengenai teknik analisis regresi yaitu

A	menganalisis keterkaitan variabel dependen dan variabel independen
B	hasil analisis tidak dapat diukur secara numerik
C	melibatkan elemen <i>control</i> dan <i>treatment</i>
D	beberapa variabel independen dapat diuji secara bersamaan
E	menggunakan teknik berbasis perangkat lunak otomatis

9. Teknik yang tepat untuk memprediksi harga rumah berdasarkan luas area yaitu

A	Correlation
B	Regression
C	A/B Testing
D	Clustering
E	Classification

10. Berikut pernyataan benar tentang klasifikasi yaitu

A	Supervised learning
B	Unsupervised learning
C	kategori secara implisit dihasilkan berdasarkan pengelompokan data
D	Klasifikasi dapat dilakukan untuk dua atau lebih kategori
E	menganalisis keterkaitan variabel dependen dan variabel independen

11. Berikut pernyataan benar tentang clustering yaitu

A	Supervised learning
B	Unsupervised learning
C	kategori secara implisit dihasilkan berdasarkan pengelompokan data

D	melibatkan elemen <i>control</i> dan <i>treatment</i>
E	Representasi visual biasanya menggunakan scatter plot

12. Grafik boxplot dapat digunakan untuk visualisasi data outlier.

A	Benar
B	Salah

13. Collaborative filtering adalah teknik pemfilteran item berdasarkan kolaborasi, atau penggabungan, dari perilaku masa lalu pengguna dengan perilaku orang lain.

A	Benar
B	Salah

14. Analisis sentiment merupakan bagian dari analisis semantik.

A	Benar
B	Salah

15. Berikut pernyataan benar mengenai Natural Language Processing (NLP) yaitu

A	digunakan untuk pengenalan teks dan ucapan
B	digunakan untuk mencari kelompok data
C	Supervised learning
D	Unsupervised learning
E	termasuk dalam jenis analisis semantik

16. Teknik analisis visual yang efektif untuk mengekspresikan pola, komposisi data melalui hubungan sebagian-keseluruhan dan distribusi geografis data yaitu

A	Heat map
B	Scatter plot
C	Box plot
D	Pie chart
E	Bar chart

17. Analisis yang melibatkan kumpulan nilai berurutan yang direkam dalam interval waktu yang teratur yaitu

A	Natural Language Processing
B	Text Analytics
C	Collaboratif filtering

D	Sentiment Analysis
E	Time series

18. Teknik yang berfokus menganalisis hubungan antara entitas dalam jaringan yaitu

A	Network graph
B	Text Analytics
C	Heat map
D	Sentiment Analysis
E	Time series

19. Analisis data spasial difokuskan pada analisis data berbasis lokasi untuk menemukan hubungan dan pola geografis yang berbeda antar entitas.

A	Benar
B	Salah

20. Manfaat analisis jaringan yaitu

A	Optimasi rute
B	Prediksi penyebaran penyakit
C	Analisis jaringan sosial
D	Prediksi cuaca
E	Prediksi harga rumah

21. Ciri penyimpanan pada disk yaitu

A	Biayanya tinggi
B	Bertujuan untuk penyimpanan jangka panjang
C	Menggunakan tape
D	Biayanya rendah

22. Berikut ciri penyimpanan file terdistribusi yaitu

A	Tidak menyediakan redundansi
B	Mendukung penyimpanan data tanpa skema
C	Menyediakan redundansi
D	Menggunakan penskalaan vertikal

23. Berikut ciri penyimpanan RDBMS

A	Menyediakan redundansi
B	Ideal untuk penyimpanan jangka panjang

C	menggunakan penskalaan horizontal
D	Tidak menyediakan redundansi

24. Berikut merupakan pernyataan benar tentang NoSQL yaitu ...

A	Hanya menerima data semi-terstruktur
B	Toleran terhadap kesalahan
C	Biaya operasional mahal
D	Dibangun di atas teknologi berbasis cluster

25. Karakteristik NoSQL yaitu

A	Kepatuhan terhadap BASE
B	Kepatuhan terhadap ACID
C	Mendukung kuiri berbasis API
D	Menyimpan data yang dinormalisasi

26. Apa yang membuat database relasional menjadi pilihan yang kurang ideal untuk menyimpan data berkecepatan tinggi?

A	Kepatuhan terhadap BASE
B	Kepatuhan terhadap ACID
C	Latensi
D	Sharding otomatis

27. Berikut contoh database relational yaitu

A	IBM DB2 pureScale, NoSQL
B	Sybase ASE Cluster Edition, Oracle Real Application Clusters (RAC)
C	NoSQL, Microsoft Parallel Data Warehouse (PDW)
D	Sybase ASE Cluster Edition, Microsoft Parallel Data Warehouse (PDW)

28. Berikut ciri penyimpanan key-value yaitu

A	Memungkinkan partial update.
B	Bertindak seperti tabel hash.
C	Tidak memungkinkan partial update.
D	Terdapat nilai indeks

29. Berikut pernyataan benar tentang penyimpanan key-value yaitu

A	Mendukung nilai kompresi
B	Memungkinkan latensi pada waktu baca

C	Memiliki struktur bersarang yang kompleks
D	Menggunakan skema pengkodean biner

30. Berikut pernyataan benar tentang penyimpanan dokumen yaitu

A	Mendukung nilai kompresi
B	Memungkinkan latensi pada waktu baca
C	Memiliki struktur bersarang yang kompleks
D	Menggunakan skema pengkodean biner

31. Contoh perangkat penyimpanan dokumen yaitu

A	Riak, Redis, Amazon Dynamodb
B	MongoDB, Terrastore, CouchDB
C	MongoDB , Amazon Dynamodb, CouchDB
D	MongoDB, Terrastore, Redis

32. Sharding pada basis data relasional dilakukan secara manual.

A	Benar
B	Salah

33. Berikut ciri penyimpanan dokumen yaitu

A	Memungkinkan partial update.
B	Bertindak seperti tabel hash.
C	Tidak memungkinkan partial update.
D	Mendukung nilai indeks

34. Berikut pernyataan benar mengenai penyimpanan dokumen yaitu

A	Memberikan dukungan skema yang fleksibel
B	menyimpan data berorientasi dokumen semi-terstruktur
C	Setiap dokumen dapat memiliki skema yang berbeda
D	Penyimpanan data tidak terstruktur

35. Berikut ciri penyimpanan column-family yaitu

A	Mendukung skema yang fleksibel
B	Menyimpan objek domain
C	Setiap baris diidentifikasi oleh kunci baris
D	Nilainya memiliki struktur yang relatif sederhana

36. Pada teorema CAP, perangkat penyimpanan NoSQL umumnya AP atau CP.

A	Benar
B	Salah

37. Berikut ciri perangkat penyimpanan grafik yaitu

A	Entitas disimpan sebagai node
B	Node dapat memiliki lebih dari satu jenis tautan
C	Node hanya dapat satu jenis tautan
D	Hubungan antar entitas disimpan sebagai edge

38. Contoh perangkat penyimpanan grafik yaitu

A	Riak, Redis, Amazon Dynamodb
B	Neo4J, OrientDB, Infinite Graph
C	Cassandra, HBase, Amazon SimpleDB
D	MongoDB, Terrastore, Redis

39. Contoh perangkat penyimpanan NewSQL yaitu

A	VoltDB, NuoDB, InnoDB
B	Riak, Redis, Amazon Dynamodb
C	Cassandra, HBase, Amazon SimpleDB
D	MongoDB, Terrastore, Redis

40. Berikut ciri penyimpanan memory yaitu

A	penskalakan secara vertikal
B	memungkinkan analitik dalam memori
C	menghilangkan latensi waktu transfer
D	memfasilitasi pembuatan informasi real-time

B. ESSAY

41. Sebutkan jenis perangkat penyimpanan NoSQL!
42. Jelaskan perbedaan antara supervise machine learning dengan unsupervised machine learning dan berikan contoh masing-masing satu metode dan contoh implementasi praktisnya
43. Berikut ini merupakan sampel data pasien Covid-19 yang bertahan hidup dan meninggal yang ditunjukkan oleh variabel survival (hidup = 0, meninggal = 1). Asumsikan bahwa Anda memiliki 10000 sampel data dengan jumlah pasien hidup 8000 orang dan meninggal 2000 orang. Anda diminta untuk melakukan analisis pada data berikut.

Detak jantung	SpO2	Pasang_O2	Temperatur	Respiratory Rate	Survival
120	97.711213	0	37	20	0
3	99	0	36.3	20	0
102	99	0	36.3	20	0
1	100	0	36	20	0
88	100	0	36	20	0
1	97.711213	0	36	20	0
89	97.711213	0	36	20	0
1	97	0	37	22	0
110	99	0	36.1	23	0
106	97.711213	0	37.5	20	0
120	97.711213	0	36.5	20	0
110	95.665794	0	36.4	22	1
90	100	0	36.5	20	1
90	100	0	36.5	20	1
122	95.665794	0	37	22	1
140	92	0	37.6	20	1

Sebutkan dan jelaskan lima teknik analisis Big Data yang Anda gunakan!