

**UAS Machine Learning**



**NAMA : GALIH DWI PUTRA G**

**KELAS : INFORMATIKA A2**

**NPM : 41155050190043**

1. Apa itu *Linear dan Logistic Regresion* dan apa gunanya?

Jawaban : Dalam *machine learning*, program komputer yang disebut algoritme menganalisis set data besar dan bekerja mundur dari data tersebut untuk menghitung persamaan regresi linier. Ilmuwan data melatih algoritme pada set data yang diketahui atau diberi label terlebih dahulu, kemudian menggunakan algoritme tersebut untuk memprediksi nilai yang tidak diketahui. Data dalam kehidupan nyata lebih rumit daripada contoh sebelumnya. Itu adalah alasan mengapa analisis regresi linier harus memodifikasi atau mengubah nilai data secara matematis untuk memenuhi empat asumsi berikut.

Hubungan linier harus ada antara variabel independen dan dependen. Untuk menentukan hubungan ini, ilmuwan data membuat plot sebar—kumpulan acak nilai  $x$  dan  $y$ —untuk melihat apakah nilai tersebut berada di sepanjang garis lurus.

2. Apa itu [\*Support Vector Machine\*](#) dan apa gunanya?

Jawaban : Support Vector Machine atau SVM adalah algoritme pembelajaran mesin yang diawasi yang dapat digunakan untuk klasifikasi dan regresi. Cara kerja SVM didasarkan pada SRM atau Structural Risk Minimization yang dirancang untuk mengolah data menjadi Hyperplane yang mengklasifikasikan ruang input menjadi dua kelas.

SVM merupakan suatu teknik untuk menemukan hyperplane yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1999). SVM memiliki kelebihan diantaranya adalah dalam **menentukan jarak menggunakan support vector sehingga proses komputasi menjadi cepat**

3. Apa itu *K-Nearest Neighbor* dan apa gunanya?

Jawaban : K-Nearest Neighbor merupakan salah satu metode untuk mengambil keputusan menggunakan pembelajaran terawasi dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN merupakan algoritma supervised learning dimana hasil dari query instance yang baru diklasifikan berdasarkan mayoritas dari kategori pada algoritma KNN.

4. Apa itu [\*Naive Bayes\*](#) dan apa gunanya?

Jawaban : Naive Bayes adalah **algoritma machine learning untuk masalah klasifikasi**. Ini didasarkan pada teorema probabilitas Bayes. Hal ini digunakan untuk klasifikasi teks yang melibatkan set data pelatihan dimensi tinggi.

Tidak hanya dikenal karena kesederhanaannya, tetapi juga karena keefektifannya. Sangat cepat untuk membangun model dan membuat prediksi dengan algoritma Naive Bayes. Naive Bayes adalah algoritma pertama yang harus dipertimbangkan untuk memecahkan masalah klasifikasi teks. Oleh karena itu, Anda harus mempelajari algoritma ini secara menyeluruh.

5. Apa itu [\*Decision Tree\*](#) dan apa gunanya?

Jawaban : **Decision tree** adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko.

alasan untuk menggunakan algoritma decision tree;

- 1) Decision tree biasanya meniru kemampuan berpikir manusia saat membuat keputusan, sehingga mudah dimengerti;
- 2) Logika dibalik decision tree dapat dengan mudah dipahami karena menunjukkan struktur seperti pohon.

6. Apa itu [\*Random Forest\*](#) dan apa gunanya?

Jawaban : Random Forest adalah algoritma pembelajaran yang supervised. "Forest" yang dibangunnya adalah kumpulan pohon keputusan, biasanya dilatih dengan metode "bagging". Ide umum dari metode bagging adalah kombinasi model pembelajaran meningkatkan hasil keseluruhan. Random Forest membangun beberapa decision tree dan menggabungkannya untuk mendapatkan prediksi yang lebih akurat dan stabil. Satu keuntungan besar dari Random Forest adalah dapat digunakan untuk masalah klasifikasi dan regresi, yang merupakan mayoritas sistem pembelajaran mesin saat ini.

7. Apa itu *K-Means* dan apa gunanya?

Jawaban : K-means merupakan salah satu algoritma yang bersifat unsupervised learning. K-Means memiliki fungsi untuk mengelompokkan data kedalam data cluster. Algoritma ini dapat menerima data tanpa ada label kategori. K-Means Clustering Algoritma juga merupakan metode non-hierarchy. Metode Clustering Algoritma adalah mengelompokkan beberapa data ke dalam kelompok yang menjelaskan data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain.

**K-Means Clustering** adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan unsupervised learning dan menggunakan metode yang mengelompokkan data berbagai partisi.

K means clustering merupakan metode algoritma dasar,yang diterapkan sebagai berikut

- Menentukan jumlah cluster
- Secara acak mendistribusikan data cluster
- Menghitung rata rata dari data yang ada di cluster.

8. Apa itu Agglomerate [\*Clustering\*](#) dan apa gunanya?

Jawaban : **Agglomerative** (metode penggabungan) adalah strategi pengelompokan hirarki yang dimulai dengan setiap objek dalam satu cluster yang terpisah kemudian membentuk cluster yang semakin membesar. Jadi, banyaknya cluster awal adalah sama dengan banyaknya objek. Dalam agglomerative method, teknik pengelompokan yang paling dikenal adalah:

**a. Single linkage (jarak terdekat atau tautan tunggal)**

Teknik yang menggabungkan cluster-cluster menurut jarak antara anggota-anggota terdekat di antara dua cluster.

**b. Average linkage (jarak rata-rata atau tautan rata-rata)**

Teknik yang menggabungkan cluster-cluster menurut jarak rata-rata pasangan anggota masing-masing pada himpunan antara dua cluster.

**c. Complete linkage (jarak terjauh atau tautan lengkap)**

Teknik yang menggabungkan cluster-cluster menurut jarak antara anggota-anggota terjauh di antara dua cluster.

9. Apa itu Apriori Algorithm dan apa gunanya?

Jawaban : Algoritma Apriori adalah [algoritma](#) yang digunakan untuk menghitung aturan asosiasi antar objek. Aturan asosiasi menjelaskan bagaimana dua atau lebih objek terkait satu sama lain.

Dengan kata lain, algoritma apriori adalah algoritma berbasis aturan asosiasi yang menganalisis apakah orang yang membeli produk A juga membeli produk B.

Algoritma ini dikemukakan oleh ilmuwan R. Agrawal dan Srikant pada tahun 1994. Algoritma ini kebanyakan digunakan untuk analisis keranjang belanja dan membantu menemukan produk yang dapat dibeli bersama.

Apriori juga dapat digunakan di bidang perawatan kesehatan untuk menemukan reaksi obat untuk pasien.

10. Apa itu Self Organizing Map dan apa gunanya?

Jawaban : *Self-organizing maps* (SOM) merupakan suatu jenis *artificial neural network* yang dilatih dengan metode *unsupervised learning*. Jaringan ini mampu menghasilkan sebuah representasi terpisah atas ruang *input* sampel pelatihan dengan dimensi rendah (biasanya dua dimensi). Representasi tersebut kemudian disebut sebagai “*map*”.

*Self-organizing maps* (SOM) mampu mempertahankan informasi struktural dari data pelatihan. Selain itu, SOM juga menghasilkan data yang tidak linier secara inheren. Jika Anda memiliki data berdimensi tinggi dan menerapkan metode lain—misalnya, *principal component analysis*—risiko data menghilang saat dimensi direduksi akan jauh lebih besar.

## Clustering data *Liga120192021* Menggunakan Metode K-Means





