**JAWABAN MACHINE LEARNING**

**UJIAN AKHIR SEMESTER**



Agnes Fadilah Damayanti

41155050190020

Informatika A1

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS LANGLANGBUANA**

**BANDUNG**

**2023**

**Bagian 1 :**

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

**Linear Regresion** adalah teknik analisis data yang memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui. Secara matematis memodelkan variabel yang tidak diketahui atau tergantung dan variabel yang dikenal atau independen sebagai persamaan linier. Misalnya, anggaplah Anda memiliki data tentang pengeluaran dan pendapatan Anda untuk tahun lalu. Teknik regresi linier menganalisis data ini dan menentukan bahwa pengeluaran Anda adalah setengah dari penghasilan Anda. Mereka kemudian menghitung biaya masa depan yang tidak diketahui dengan mengurangi separuh pendapatan yang diketahui di masa depan.

**Linear regresion** digunakan untuk prediksi dan modelling berdasarkan data-data yang ada. Regresi linear sederhana merupakan pemodelan regresi berdasarkan 1 variabel prediktor.

**Logistic Regresion**adalah teknik analisis data yang menggunakan matematika untuk menemukan hubungan antara dua faktor data. Kemudian menggunakan hubungan ini untuk memprediksi nilai dari salah satu faktor tersebut berdasarkan faktor yang lain. Prediksi biasanya memiliki jumlah hasil yang terbatas, seperti ya atau tidak.

**Logistic Regresion**, digunakan untuk model regresi yang variable dependennya merupakan variable dummy. Apa itu variable dummy? Variable dummy adalah dikotomus variable, yaitu variable yang hanya menggunakan dua kemungkinan nilai. Biasa dilambangkan dengan nilai 0 dan 1.

1. **Apa itu**[**Support Vector Machine**](http://elearning.unla.ac.id/mod/resource/view.php?id=6297)**dan apa gunanya?**

**Support Vector Machine (SVM)** merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear.

**SVM** digunakan untuk mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas. Hyperplane adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas. Dalam 2-D fungsi yang digunakan untuk klasifikasi antar kelas disebut sebagai line whereas, fungsi yang digunakan untuk klasifikasi antas kelas dalam 3-D disebut plane similarly, sedangan fungsi yang digunakan untuk klasifikasi di dalam ruang kelas dimensi yang lebih tinggi di sebut hyperplane.

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

Algoritma ***K-Nearest Neighbor***merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah K data terdekat (tetangganya) sebagai acuan untuk menentukan kelas dari data baru. Algoritma ini mengklasifikasikan data berdasarkan similarityatau kemiripan atau kedekatannya terhadap data lainnya. **K-Nearest Neighbor** digunakan untuk menentukan nilai bobot setiap kriteria dengan mengklasifikasikan dengan baik atau buruk.

1. **Apa itu**[**Naive Bayes**](http://elearning.unla.ac.id/mod/resource/view.php?id=6298)**dan apa gunanya?**

**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. Beberapa contohnya adalah penyaringan spam, analisis sentimental, dan klasifikasi artikel berita.

1. **Apa itu**[**Decision Tree**](http://elearning.unla.ac.id/mod/resource/view.php?id=6282)**dan apa gunanya?**

**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.

**Decision tree** adalah suatu struktur yang digunakan untuk mengambil suatu keputusan. Untuk bisa mengambil keputusan dengan struktur ini, Anda perlu menyiapkan beberapa variabel yang nantinya berfungsi sebagai bahan pertimbangan keputusan yang diambil.

1. **Apa itu**[**Random Forest**](http://elearning.unla.ac.id/mod/resource/view.php?id=6283)**dan apa gunanya?**

**Random Forest** adalah kumpulan dari decision tree atau pohon keputusan. Algoritma ini merupakan kombinasi masing-masing tree dari decision tree yang kemudian digabungkan menjadi satu model. Biasanya, Random Forest dipakai untuk masalah regresi dan klasifikasi dengan kumpulan data yang berukuran besar.

**Random Forest** adalah algoritma dalam machine learning yang digunakan untuk pengklasifikasian data set dalam jumlah besar. Karena fungsinya bisa digunakan untuk banyak dimensi dengan berbagai skala dan performa yang tinggi.

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

**K-Means** adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode k-means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

**K-Means** merupakan algoritma yang umum digunakan untuk clustering doku- men. Prinsip utama K-Means adalah menyusun k prototype atau pusat masa (centroid) dari sekumpulan data berdimensi. Sebelum diterapkan proses algoritma K-means, dokumen akan di preprocessing terlebih dahulu.

1. **Apa itu Agglomerate**[**Clustering**](http://elearning.unla.ac.id/mod/resource/view.php?id=6100)**dan apa gunanya?**

**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.

**Metode agglomerative hierarchical clustering** merupakan metode analisis cluster yang bertujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya, yang dimulai dengan objek-objek individual sampai objek-objek tersebut bergabung menjadi satu cluster tunggal.

1. **Apa itu Apriori Algorithm dan apa gunanya?**

**Algoritma Apriori** merupakan salah satu algoritma unsupervised pada data mining, khususnya pada tugas asosiasi. Asosiasi adalah proses menyelidiki hubungan antara satu objek data dengan objek lainnya dalam dataset.

**Algoritma Apriori** adalah [algoritma](https://www.trivusi.web.id/2022/06/mengenal-algoritma.html) yang digunakan untuk menghitung aturan asosiasi antar objek. Aturan asosiasi menjelaskan bagaimana dua atau lebih objek terkait satu sama lain.

1. **Apa itu Self Organizing Map dan apa gunanya?**

**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 Map (SOM) merupakan metode untuk membagi pola masukan ke dalam beberapa kelompok cluster. Penelitian ini dilakukan untuk melakukan klastering/pengelompokan jurusan sekolah terhadap siswa yang mengalami kebingunan untuk memutuskan.**

**Bagian 2:**

Metode Logistic Regression



