

Nama : Yolanda Fazhri Gemilang

NPM : 41155050190068

Kelas : A2 / Semester 7

Bagian I

1. Regresi linier dan regresi logistik adalah dua jenis algoritma pembelajaran terawasi. Regresi linier digunakan ketika variabel dependen kontinu, dan modelnya linier. Regresi logistik digunakan ketika variabel dependen diskrit, dan modelnya adalah nonlinier.
 2. Pengertian Support Vector Machine (SVM) Support Vector Machine—selanjutnya disebut SVM—adalah metode pada machine learning yang dapat digunakan untuk menganalisis data dan mengurutkannya ke dalam salah satu dari dua kategori.
 3. Pengertian dan Cara Kerja Algoritma K-Nearest Neighbors (KNN) K-nearest neighbors atau knn adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (train data sets), yang diambil dari k tetangga terdekatnya (nearest neighbors). Dengan k merupakan banyaknya tetangga terdekat.
 4. Naïve Bayes merupakan sebuah metode penggolongan berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen). Pada klasifikasi Naïve Bayes, proses pembelajaran lebih ditekankan pada mengestimasi probabilitas. Keuntungan dari pendekatan ini yaitu pengklasifikasian akan mendapatkan nilai error yang lebih kecil ketika data set berjumlah besar
 5. Decision tree merupakan metode non parametrik yang digunakan untuk klasifikasi dan regresi. Tujuan dari decision tree adalah membuat model yang memprediksi nilai variabel target dengan mengikuti aturan keputusan sederhana dari fitur data yang tersedia
 6. 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. Klasifikasi ini dilakukan melalui penggabungan tree dalam decision tree dengan cara training dataset yang Anda miliki.
- Pada akhirnya, algoritma ini sangat berguna dalam pengklasifikasian data, terutama jika jumlahnya besar. Random Forest adalah algoritma terbaik sebab mudah digunakan dan fleksibel.

7. K-Means merupakan algoritma yang umum digunakan untuk clustering dokumen. Prinsip utama K-Means adalah menyusun k prototype atau pusat massa

(centroid) dari sekumpulan data berdimensi. Sebelum diterapkan proses algoritma

K-means, dokumen akan di preprocessing terlebih dahulu. Kemudian dokumen direpresentasikan sebagai vektor yang memiliki term dengan nilai tertentu.

8. Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data.

9. 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 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. Pada data mining, algoritma Apriori banyak digunakan untuk menemukan data yang paling sering muncul dalam sebuah database. Item data transaksi pada database membentuk itemset.

10. 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”. SOM juga merupakan metode untuk melakukan pengurangan dimensi pada sampel yang dilatih. SOM memproyeksikan data ke dalam grid berdimensi rendah untuk proses pengelompokan yang menerapkan metode unsupervised learning. Dengan SOM, Anda dapat melakukan pengurangan dimensi tanpa khawatir kehilangan data.

Bagian 2

Clustering data *Liga120192021* Menggunakan Metode K-Means

```
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved

Files
sample_data
Liga120192021.csv

+ Code + Text
#library
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

dataset = pd.read_csv('/content/Liga120192021.csv')
X = dataset.iloc[:, [3, 4]].values

from sklearn.cluster import KMeans
X = dataset.iloc[:, [3, 4]].values

from sklearn.cluster import KMeans
wcss = []
for i in range(1, 104):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 104), wcss)
plt.title('Kmeans clustering Metode Elbow')
plt.xlabel('Jumlah Cluster')
plt.ylabel('WCSS')
plt.show()

<ipython-input-8-671bcef31bb7>:16: ConvergenceWarning: Number of distinct clusters (86) found smaller than n_clusters (87). Possibly due to duplicate points
kmeans.fit(X)
<ipython-input-8-671bcef31bb7>:16: ConvergenceWarning: Number of distinct clusters (86) found smaller than n_clusters (88). Possibly due to duplicate points
kmeans.fit(X)
<ipython-input-8-671bcef31bb7>:16: ConvergenceWarning: Number of distinct clusters (86) found smaller than n_clusters (89). Possibly due to duplicate points
kmeans.fit(X)
<ipython-input-8-671bcef31bb7>:16: ConvergenceWarning: Number of distinct clusters (86) found smaller than n_clusters (90). Possibly due to duplicate points
kmeans.fit(X)
<ipython-input-8-671bcef31bb7>:16: ConvergenceWarning: Number of distinct clusters (86) found smaller than n_clusters (91). Possibly due to duplicate points
kmeans.fit(X)
41s completed at 10:48 AM
```



