

Nama : Satya Athaya Daniswara

NIM : 1103213152

TUGAS PERBAIKAN BAB 3

1. Definisi Pembelajaran Tidak Terawasi

- Unsupervised Learning: Merupakan metode dalam pembelajaran mesin di mana algoritma bekerja dengan data yang tidak memiliki label. Tujuan utama dari pembelajaran tidak terawasi adalah untuk menemukan pola atau struktur dalam data tanpa bimbingan dari label target.

2. Tujuan Pembelajaran Tidak Terawasi

- Clustering (Pengelompokan): Salah satu tujuan utama dari pembelajaran tidak terawasi adalah mengelompokkan data ke dalam grup yang memiliki kesamaan. Misalnya, dalam analisis pelanggan, data pelanggan dapat dikelompokkan berdasarkan perilaku pembelian mereka.
- Dimensionality Reduction (Reduksi Dimensi): Mengurangi jumlah fitur dalam dataset sambil mempertahankan informasi penting. Ini berguna untuk visualisasi data dan mengurangi kompleksitas model.

3. Clustering

- K-Means: Salah satu algoritma clustering yang paling umum digunakan. Algoritma ini membagi data menjadi k cluster berdasarkan kedekatan data. Prosesnya melibatkan:
 - Menentukan jumlah cluster (k).
 - Menginisialisasi centroid untuk setiap cluster.
 - Mengelompokkan data berdasarkan jarak ke centroid.
 - Memperbarui centroid berdasarkan rata-rata data dalam cluster.
 - Mengulangi proses hingga konvergensi.

4. Metode Clustering Lainnya

- Hierarchical Clustering: Membangun struktur pohon (dendrogram) untuk mengelompokkan data. Metode ini dapat bersifat agglomerative (menggabungkan cluster) atau divisive (memecah cluster).
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Algoritma yang mengelompokkan data berdasarkan kepadatan. Ini efektif untuk menemukan cluster dengan bentuk yang tidak teratur dan dapat mengidentifikasi outlier.

5. Reduksi Dimensi

- PCA (Principal Component Analysis): Teknik yang digunakan untuk mereduksi dimensi data sambil mempertahankan varians sebanyak mungkin. PCA mengubah data ke dalam bentuk baru yang disebut komponen utama, yang merupakan kombinasi linier dari fitur asli.
- t-SNE (t-Distributed Stochastic Neighbor Embedding): Metode yang digunakan untuk visualisasi data berdimensi tinggi dengan memetakan data ke dalam ruang berdimensi rendah (biasanya 2D atau 3D) sambil mempertahankan struktur lokal.

6. Tantangan dalam Pembelajaran Tidak Terawasi

- Evaluasi Hasil: Tanpa label target, evaluasi hasil dari algoritma clustering atau reduksi dimensi menjadi lebih sulit. Metode evaluasi seperti silhouette score atau Davies-Bouldin index dapat digunakan untuk menilai kualitas cluster.
- Menentukan Jumlah Cluster: Salah satu tantangan utama dalam clustering adalah menentukan jumlah cluster yang optimal. Metode seperti elbow method atau silhouette analysis dapat membantu dalam menentukan jumlah cluster yang tepat.

7. Preprocessing dan Scaling

- Preprocessing: Langkah penting dalam pembelajaran tidak terawasi adalah membersihkan dan menyiapkan data. Ini termasuk mengatasi nilai yang hilang, menghapus outlier, dan mengubah format data jika diperlukan.
- Scaling: Menstandarisasi atau menormalisasi data untuk menyamakan skala fitur. Ini penting untuk algoritma berbasis jarak, seperti K-Means, di mana jarak antar titik data mempengaruhi hasil clustering.

8. Visualisasi

- Visualisasi dalam Clustering: Membantu memahami pola dalam data setelah clustering. Teknik visualisasi seperti scatter plot dapat digunakan untuk menampilkan hasil clustering, dan metode reduksi dimensi seperti PCA atau t-SNE dapat digunakan untuk memvisualisasikan data berdimensi tinggi.

Kesimpulan

Bab ini menjelaskan konsep dasar dari pembelajaran tidak terawasi, termasuk tujuan, metode, dan tantangan yang dihadapi. Clustering dan reduksi dimensi adalah dua teknik utama dalam pembelajaran tidak terawasi yang membantu dalam menemukan pola dan struktur dalam data tanpa label. Evaluasi hasil dan preprocessing data juga merupakan aspek penting untuk memastikan keberhasilan algoritma dalam menemukan informasi yang berguna.

Jika Anda memerlukan elaborasi lebih lanjut pada bagian tertentu atau ada pertanyaan lain, silakan beri tahu!

1. Definisi Pembelajaran Tidak Terawasi

- Unsupervised Learning: Algoritma yang bekerja tanpa label target, bertujuan untuk memahami struktur internal data.

2. Tujuan Pembelajaran Tidak Terawasi

- Clustering (Pengelompokan): Membagi data menjadi kelompok berdasarkan kesamaan.
- Dimensionality Reduction (Reduksi Dimensi): Mengurangi jumlah fitur dalam data sambil tetap menjaga informasi penting.

3. Clustering

- K-Means: Algoritma yang membagi data menjadi k cluster berdasarkan kedekatan data.

4. Reduksi Dimensi

- PCA (Principal Component Analysis): Teknik untuk mereduksi dimensi data sambil mempertahankan varians sebanyak mungkin, berguna untuk visualisasi data berdimensi tinggi.

5. Tantangan dalam Pembelajaran Tidak Terawasi

- Evaluasi Hasil: Tanpa label target, evaluasi hasil menjadi lebih sulit.
- Menentukan Jumlah Cluster: Kesulitan dalam menentukan jumlah cluster atau dimensi yang optimal.

6. Preprocessing dan Scaling

- Preprocessing: Langkah penting untuk membersihkan dan menyiapkan data.
- Scaling: Menstandarisasi atau menormalisasi data untuk menyamakan skala fitur, penting untuk algoritma berbasis jarak.

7. Visualisasi

- Visualisasi dalam Clustering: Membantu memahami pola dalam data setelah clustering, menggunakan teknik seperti PCA dan t-SNE.

