

Nama : Satya Athaya Daniswara

NIM : 1103213152

TUGAS PERBAIKAN BAB 2

Berikut adalah parafrase dan elaborasi dari isi dokumen yang Anda berikan, yang berfokus pada konsep-konsep utama dalam pembelajaran terawasi (supervised learning):

1. Fungsi Keputusan

- Definisi: Fungsi keputusan adalah mekanisme yang digunakan oleh model klasifikasi untuk menentukan kategori mana yang sesuai untuk sebuah titik data berdasarkan fitur-fitur yang dimilikinya.

2. Memperkirakan Probabilitas

- Algoritma: Beberapa algoritma, seperti Regresi Logistik dan Naive Bayes, tidak hanya memberikan label kelas yang paling mungkin, tetapi juga memperkirakan probabilitas untuk setiap kelas. Ini sangat berguna untuk memahami tingkat keyakinan model terhadap prediksinya.

3. Ketidakpastian dalam Klasifikasi Multikelas

- Klasifikasi Multikelas: Dalam situasi di mana terdapat lebih dari dua kelas, model harus mampu memprediksi lebih dari satu kelas dan menangani ketidakpastian dengan memberikan probabilitas untuk setiap kelas.

4. Ringkasan dan Pandangan ke Depan

- Kesimpulan: Bab ini merangkum berbagai algoritma yang digunakan dalam pembelajaran terawasi, menekankan pentingnya pemilihan algoritma yang tepat serta tantangan seperti overfitting, underfitting, dan pemilihan fitur.

- Pandangan ke Depan: Meskipun banyak algoritma yang telah terbukti efektif, penelitian dan pengembangan dalam pembelajaran mesin terus berlanjut, terutama dalam teknik deep learning dan pemrosesan data besar.

5. Klasifikasi dan Regresi

- Klasifikasi: Tugas di mana model memprediksi kategori atau label diskrit, seperti mengklasifikasikan email sebagai spam atau tidak spam.

- Regresi: Tugas di mana model memprediksi nilai kontinu, seperti memprediksi harga rumah berdasarkan fitur-fitur tertentu.

6. Generalisasi, Overfitting, dan Underfitting

- Generalisasi: Kemampuan model untuk berfungsi dengan baik pada data baru yang belum pernah dilihat sebelumnya.
- Overfitting: Terjadi ketika model terlalu kompleks dan belajar dari noise dalam data pelatihan, sehingga kinerjanya buruk pada data uji.
- Underfitting: Terjadi ketika model terlalu sederhana untuk menangkap pola dalam data, sehingga tidak dapat memberikan prediksi yang akurat.

7. Hubungan Kompleksitas Model dengan Ukuran Dataset

- Kompleksitas Model: Semakin kompleks model, semakin besar risiko overfitting, terutama pada dataset kecil.
- Ukuran Dataset: Dengan dataset yang lebih besar, model yang lebih kompleks dapat lebih efektif karena lebih banyak data membantu model belajar pola yang lebih umum.

8. Algoritma Pembelajaran Mesin Terawasi

- Algoritma yang Dibahas:
 - k-Nearest Neighbors (k-NN)
 - Model Linear (misalnya, Regresi Linear)
 - Naive Bayes
 - Support Vector Machines (SVM)
 - Decision Trees
 - Random Forests

9. Evaluasi Model

- Metrik Evaluasi: Setelah model dilatih, penting untuk mengevaluasi kinerjanya menggunakan data uji. Beberapa metrik yang digunakan termasuk:
 - Akurasi: Persentase prediksi yang benar.
 - Precision dan Recall: Mengukur kinerja model, terutama dalam kasus ketidakseimbangan kelas.
 - F1-score: Rata-rata harmonis antara precision dan recall.
 - Mean Squared Error (MSE): Digunakan dalam regresi untuk mengukur perbedaan antara prediksi dan nilai sebenarnya.

10. Mengatasi Tantangan dalam Pembelajaran Terawasi

- Pembersihan Data: Data sering kali mengandung nilai yang hilang atau tidak konsisten, sehingga pembersihan data sangat penting untuk kinerja model.
- Pemilihan Fitur: Memilih fitur yang relevan dan mengurangi fitur yang tidak penting dapat meningkatkan kinerja model dan mengurangi risiko overfitting.
- Skalabilitas: Beberapa algoritma, seperti SVM, dapat memakan waktu jika dataset sangat besar, sehingga diperlukan teknik untuk meningkatkan skalabilitas.

11. Contoh Dataset

- Dataset yang Sering Digunakan:
 - Iris Dataset: Untuk klasifikasi bunga berdasarkan fitur morfologi.
 - Boston Housing Dataset: Untuk regresi, memprediksi harga rumah.
 - Digits Dataset: Untuk klasifikasi gambar digit tulisan tangan.

12. Kelebihan dan Kekurangan Algoritma

- Naive Bayes: Cepat dan efisien, tetapi asumsi independensi antar fitur sering kali tidak benar.
- Decision Trees: Mudah diinterpretasikan, tetapi rentan terhadap overfitting.
- Random Forests: Meningkatkan akurasi dengan mengurangi overfitting melalui voting dari beberapa pohon keputusan.

Dokumen ini memberikan gambaran menyeluruh tentang konsep-konsep dasar dalam pembelajaran terawasi, tantangan yang dihadapi, serta algoritma yang digunakan untuk klasifikasi dan regresi. Jika Anda memerlukan elaborasi lebih lanjut pada bagian tertentu, silakan beri tahu!