

Nama : Maria Nagita Tria Vanessa
Nim : 231011400228
Kelas : TPLE005
Mata Kuliah : Machine Learning (UAS)

Implementasi Decision Tree pada Dataset Titanic

1. Teori Singkat

Decision Tree merupakan salah satu algoritma supervised learning yang digunakan untuk menyelesaikan permasalahan klasifikasi dan regresi. Algoritma ini bekerja dengan membentuk struktur pohon keputusan yang terdiri dari node, cabang, dan leaf. Setiap node merepresentasikan atribut tertentu yang digunakan untuk membagi data, sedangkan leaf menunjukkan hasil akhir berupa kelas atau nilai prediksi.

Pada proses pembentukan Decision Tree, pemilihan atribut dilakukan berdasarkan kriteria tertentu seperti Gini Index atau Entropy, yang bertujuan untuk menghasilkan pembagian data paling homogen. Salah satu keunggulan utama Decision Tree adalah kemampuannya dalam menghasilkan model yang mudah dipahami dan diinterpretasikan, karena proses pengambilan keputusan dapat divisualisasikan dalam bentuk pohon.

Dalam penelitian ini, Decision Tree digunakan untuk memprediksi kelangsungan hidup penumpang Titanic, dengan target klasifikasi berupa dua kelas, yaitu penumpang yang selamat dan tidak selamat.

2. Metodologi

Metodologi penelitian dimulai dengan pemilihan dataset Titanic yang diperoleh dari Kaggle dan pustaka Seaborn. Dataset ini berisi informasi penumpang seperti kelas penumpang (pclass), jenis kelamin (sex), usia (age), tarif (fare), dan pelabuhan keberangkatan (embarked).

Tahap selanjutnya adalah preprocessing data, yang meliputi penanganan missing value pada atribut usia dengan menggunakan nilai median, serta pengisian nilai kosong pada atribut embarked menggunakan nilai yang paling sering muncul. Data kategorikal kemudian diubah menjadi bentuk numerik menggunakan teknik label encoding agar dapat diproses oleh algoritma Decision Tree. Setelah preprocessing selesai, dataset dibagi menjadi data latih (training set) dan data uji (testing set) dengan perbandingan 80% data latih dan 20% data uji. Model Decision Tree kemudian dibangun menggunakan kriteria Gini Index dengan pembatasan kedalaman pohon (max_depth) untuk menghindari terjadinya overfitting.

Evaluasi model dilakukan menggunakan beberapa metrik klasifikasi, yaitu Accuracy, Precision, Recall, dan F1-score, untuk mengukur kinerja model dalam memprediksi data uji.

3. Hasil dan Analisis

Hasil pengujian menunjukkan bahwa model Decision Tree mampu memprediksi kelangsungan hidup penumpang Titanic dengan tingkat akurasi yang cukup baik. Nilai precision dan recall menunjukkan bahwa model dapat mengenali penumpang yang selamat maupun tidak selamat secara relatif seimbang.

Visualisasi pohon keputusan memperlihatkan bahwa atribut jenis kelamin (sex) dan kelas penumpang (pclass) menjadi faktor yang paling berpengaruh dalam menentukan keselamatan penumpang. Penumpang perempuan dan penumpang dengan kelas lebih tinggi cenderung memiliki peluang selamat yang lebih besar. Pembatasan kedalaman pohon terbukti membantu dalam mengurangi kompleksitas model dan meningkatkan kemampuan generalisasi terhadap data uji. Secara keseluruhan, hasil eksperimen menunjukkan bahwa Decision Tree cukup efektif digunakan untuk permasalahan klasifikasi pada dataset Titanic, terutama ketika interpretasi model menjadi kebutuhan utama.

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Decision Tree mampu memberikan performa yang baik dalam memprediksi kelangsungan hidup penumpang Titanic. Keunggulan utama metode ini terletak pada kemudahan interpretasi dan visualisasi hasil keputusan. Meskipun Decision Tree memiliki kelemahan berupa kecenderungan terhadap overfitting,

penerapan teknik pembatasan kedalaman pohon dapat meningkatkan performa model secara signifikan.

Dengan demikian, Decision Tree dapat dijadikan metode yang efektif dan relevan untuk menyelesaikan permasalahan klasifikasi, khususnya pada dataset dengan karakteristik campuran antara data numerik dan kategorikal seperti dataset Titanic.