

# *Prediksi Hasil Pertandingan NBA Berdasarkan Statistik Pemain Menggunakan Decision Tree dan Random Forest*

Ammar Ghazy Tanumijaya  
5025231203  
Teknik Informatika  
Institut Teknologi Sepuluh Nopember  
Surabaya, Indonesia  
ammartanumijaya.04@gmail.com

**Abstract**— Paper ini membahas penerapan algoritma data mining untuk memprediksi hasil pertandingan NBA (menang atau kalah) berdasarkan data statistik pemain dan tim lawan. Dua algoritma, yaitu Decision Tree dan Random Forest, diuji dan dibandingkan. Dataset yang digunakan berasal dari data pertandingan NBA historis, yang diproses dengan encoding dan scaling. Model Random Forest menunjukkan performa terbaik dengan akurasi mencapai 87.5%. Model tersebut kemudian di deploy dalam bentuk aplikasi web interaktif menggunakan Streamlit. Hasil penelitian menunjukkan bahwa statistik seperti poin, assist, dan team lawan memiliki kontribusi signifikan dalam menentukan hasil pertandingan.

**Keywords**— NBA, prediksi pertandingan, data mining, Random Forest, Streamlit

## I. INTRODUCTION

Dalam dunia olahraga profesional seperti NBA, prediksi hasil pertandingan memiliki nilai strategis yang sangat tinggi, baik untuk pelatih, manajer tim, analis data, maupun penggemar. Keberhasilan dalam memprediksi hasil pertandingan dapat digunakan untuk merancang strategi permainan, menentukan formasi optimal, dan membuat keputusan manajerial yang lebih berbasis data. Di sisi lain, industri taruhan dan penyiaran juga sangat bergantung pada kemampuan untuk memproyeksikan hasil pertandingan secara akurat.

Data mining menjadi pendekatan yang semakin populer dalam menganalisis dan mengambil keputusan dari data olahraga yang sangat besar dan kompleks. Setiap pertandingan NBA menghasilkan ribuan data statistik dari berbagai aspek permainan. Dengan memanfaatkan teknik data mining dan machine learning, kita dapat menemukan pola tersembunyi dan membuat prediksi yang andal.

Penelitian ini bertujuan untuk membangun model klasifikasi biner untuk memprediksi hasil pertandingan NBA berdasarkan statistik pemain dan identitas lawan tanding. Fokus utama adalah membandingkan dua algoritma populer yaitu Decision Tree dan Random Forest dalam hal akurasi dan efektivitas prediksi.

## II. DATASET DAN PREPROCESSING

### A. Dataset

Dataset yang digunakan dalam penelitian ini berasal dari data historis pertandingan NBA yang dikumpulkan dari sumber terbuka seperti Basketball Reference. Dataset mencakup ratusan pertandingan dengan berbagai fitur statistik pemain, antara lain: menit bermain (MP), jumlah field goal (FG), poin (PTS), assist (AST), steal (STL), dan blok (BLK), serta nama tim ('Tm') dan lawan ('Opp').

Target dari klasifikasi ini adalah variabel 'Res' (Result), yang didefinisikan sebagai 1 jika tim menang, dan 0 jika tim

kalah. Label ini dihasilkan dari kolom 'Result' yang berisi nilai 'W' atau 'L'.

### B. Preprocessing

Sebelum membangun model, data terlebih dahulu diproses. Data kategorikal seperti nama tim dan lawan diubah menjadi numerik dengan menggunakan Label Encoding. Sementara itu, seluruh fitur numerik dinormalisasi menggunakan MinMaxScaler untuk memastikan skala data seragam dan mempercepat proses pembelajaran model.

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk memastikan evaluasi model dilakukan secara objektif pada data yang belum pernah dilihat saat pelatihan.

## III. METODE DAN IMPLEMENTASI

### A. Metode

Dua algoritma klasifikasi digunakan dalam penelitian ini, yaitu Decision Tree dan Random Forest. Decision Tree adalah algoritma yang mudah dipahami dan divisualisasikan, namun rentan terhadap overfitting pada data pelatihan. Sementara itu, Random Forest adalah metode ensemble yang menggabungkan banyak pohon keputusan dan mengambil suara mayoritas untuk menghasilkan prediksi akhir, yang umumnya lebih akurat dan stabil.

### B. Implementasi

Implementasi algoritma dilakukan menggunakan pustaka scikit-learn. Model dilatih pada data yang telah diproses sebelumnya. Evaluasi dilakukan dengan menggunakan metrik akurasi, confusion matrix, precision, recall, dan F1-score untuk memberikan gambaran yang menyeluruh mengenai kinerja masing-masing model.

Random Forest dipilih sebagai kandidat utama untuk deployment karena memiliki kinerja yang lebih konsisten dan akurasi lebih tinggi dibandingkan Decision Tree.

## IV. HASIL DAN EVALUASI

### A. Hasil

Setelah pelatihan, Decision Tree menghasilkan akurasi sebesar 81.3%, sedangkan Random Forest mencapai akurasi sebesar 87.5% pada data uji. Selain itu, hasil confusion matrix menunjukkan bahwa Random Forest memiliki false positive dan false negative yang lebih rendah dibandingkan Decision Tree.

Analisis feature importance yang dilakukan pada model Random Forest mengungkap bahwa fitur paling berpengaruh dalam menentukan hasil pertandingan adalah PTS (Points), AST (Assists), dan STL (Steals). Hal ini sejalan dengan

pemahaman umum bahwa kontribusi skor dan peran distribusi bola sangat menentukan keberhasilan tim.

### B. Evaluasi

Dengan evaluasi metrik yang komprehensif, Random Forest terbukti sebagai pilihan yang unggul dalam kasus prediksi hasil pertandingan ini.

## V. DEPLOYMENT

Model Random Forest yang telah dilatih kemudian di deploy dalam bentuk aplikasi web interaktif menggunakan framework Streamlit. Aplikasi memungkinkan pengguna untuk memasukkan data statistik pemain secara manual dan menerima prediksi hasil pertandingan secara real-time.

Untuk mendukung deployment, objek model, scaler, dan label encoder disimpan dalam bentuk file `.pkl` menggunakan pustaka `pickle`. File ini kemudian dimuat kembali dalam aplikasi Streamlit untuk memastikan konsistensi hasil prediksi dengan model yang telah dilatih.

Interface pengguna dibangun dengan mempertimbangkan kemudahan input dan interpretasi hasil, dengan komponen seperti slider dan dropdown. Aplikasi ini dapat dijalankan secara lokal maupun di hosting di cloud seperti Streamlit Community Cloud.

## VI. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa model Random Forest sangat efektif dalam memprediksi hasil pertandingan NBA berdasarkan data statistik pemain. Dengan akurasi mencapai 87.5%, model ini memberikan hasil yang memuaskan dan layak digunakan dalam konteks analisis olahraga.

Beberapa fitur statistik memiliki pengaruh signifikan terhadap hasil pertandingan, seperti jumlah poin, assist, dan steal. Hal ini memberikan insight yang berguna bagi pelatih dan analis dalam mengevaluasi performa pemain dan strategi permainan.

Untuk pengembangan selanjutnya, disarankan untuk menambahkan fitur kontekstual lain seperti status home/away, klasemen liga, kondisi pemain, dan rotasi tim. Penggunaan teknik ensemble lain atau model deep learning juga bisa dieksplorasi untuk meningkatkan akurasi prediksi.

## REFERENCES

- [1] Eduardo Palmieri, "NBA Player Stats Season 24-25," Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/eduardopalmieri/nba-player-stats-season-2425>
- [2] Basketball Reference, "NBA Player Stats," [Online]. Available: <https://www.basketball-reference.com>