

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ultimate Fighting Championship (UFC) merupakan salah satu organisasi seni bela diri campuran (Mixed Martial Arts/MMA) terbesar di dunia yang menyajikan pertandingan antara petarung dari berbagai disiplin bela diri seperti tinju, gulat, jiu-jitsu Brasil, Muay Thai, dan lainnya. Seiring dengan meningkatnya popularitas UFC, analisis data dan prediksi hasil pertandingan menjadi topik yang semakin menarik, baik untuk kepentingan penelitian akademik maupun industri taruhan olahraga (Kirkpatrick, 2020).

Prediksi hasil pertandingan olahraga telah berkembang dengan pesat berkat kemajuan dalam bidang AI (machine learning). Dalam beberapa tahun terakhir, algoritma pembelajaran mesin seperti Random Forest dan XGBoost banyak digunakan untuk menganalisis data olahraga dan membuat prediksi yang lebih akurat dibandingkan dengan pendekatan tradisional seperti analisis statistik konvensional atau intuisi manusia (Bunker & Thabtah, 2019).

Beberapa penelitian telah menerapkan algoritma Random Forest dan XGBoost untuk memprediksi hasil pertandingan olahraga. Sebuah studi oleh Fandi, R. S., Arifiyanti, A. A., & Wati, S. F. A. membandingkan kinerja kedua algoritma dalam memprediksi hasil pertandingan Dota 2, di mana XGBoost dengan skenario pembagian data 80:20 menunjukkan akurasi terbaik sebesar 52%. Penelitian lain yang dilakukan oleh Liputra, Edbert Anthony menggunakan Random Forest, Naive

Bayes, dan Logistic Regression untuk memprediksi hasil pertandingan bulu tangkis dengan tujuan mengidentifikasi faktor-faktor kunci yang mempengaruhi kemenangan. Selain itu, metode Random Forest juga telah diterapkan dalam prediksi hasil pertandingan sepak bola, sebagaimana dibahas dalam studi dari Martua, M Sukron, yang membandingkan kinerja model menggunakan data asli dan data yang telah di-oversampling untuk menangani ketidakseimbangan data. Studi-studi ini menunjukkan bahwa penggunaan algoritma machine learning, terutama Random Forest dan XGBoost, dapat membantu meningkatkan akurasi prediksi dalam berbagai cabang olahraga.

Random Forest merupakan algoritma ensemble learning berbasis *bagging* yang membangun sejumlah besar *decision tree* secara paralel. Setiap pohon keputusan dilatih menggunakan subset acak dari data pelatihan dan fitur, dengan tujuan untuk meningkatkan generalisasi model dan mengurangi *variance*. Pendekatan ini menjadikan Random Forest tahan terhadap *overfitting*, terutama pada data yang bersifat kompleks dan memiliki dimensi tinggi. Keunggulan lainnya adalah kemampuannya dalam mengukur pentingnya fitur (*feature importance*) serta kestabilan prediksi melalui rata-rata atau voting mayoritas dari hasil masing-masing pohon.

Sementara itu, XGBoost (Extreme Gradient Boosting) merupakan salah satu algoritma *boosting* yang menggabungkan sejumlah model lemah (*weak learners*) dalam bentuk *decision tree* secara sekuensial. Setiap model baru dibangun untuk memperbaiki kesalahan dari model sebelumnya melalui pendekatan *gradient descent* terhadap fungsi kerugian (*loss function*). XGBoost memiliki keunggulan

dalam hal efisiensi komputasi, regularisasi (L1 dan L2) untuk mencegah *overfitting*, serta kemampuan menangani data yang tidak seimbang (*imbalanced data*). Selain itu, XGBoost mendukung *parallel processing*, *tree pruning*, dan *sparsity-aware algorithm*, yang secara keseluruhan meningkatkan performa dan akurasi prediksi. (Chen & Guestrin, 2016).

Dalam konteks UFC, beberapa faktor statistik yang dapat digunakan untuk memprediksi hasil pertandingan meliputi jumlah kemenangan dan kekalahan sebelumnya, akurasi serangan, jumlah serangan per menit, efektivitas takedown, pertahanan terhadap serangan lawan, serta waktu pertarungan rata-rata. Dengan memanfaatkan data-data ini, model pembelajaran mesin dapat dibuat untuk menentukan probabilitas kemenangan seorang petarung berdasarkan performa historisnya (Magnusson, 2024).

Beberapa penelitian terdahulu telah menunjukkan efektivitas pendekatan machine learning dalam memprediksi hasil pertandingan olahraga. Misalnya, penelitian yang menggunakan Random Forest untuk memprediksi hasil pertandingan tenis menunjukkan akurasi hingga 80% (arXiv:1910.03203). Selain itu, studi lain yang diterapkan pada UFC menggunakan Random Forest mencapai akurasi sebesar 67,56% dalam memprediksi pemenang pertandingan (Hadi, 2022). Namun, masih terdapat peluang untuk meningkatkan akurasi prediksi dengan menggabungkan algoritma lain seperti XGBoost serta melakukan optimasi pada pemilihan fitur dan parameter model.

Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma Random Forest dan XGBoost dalam memprediksi hasil pertandingan UFC berdasarkan

statistik pertarungan. Dengan membandingkan performa kedua algoritma, diharapkan dapat diperoleh model yang lebih akurat dan andal dalam memprediksi pemenang suatu pertandingan. Penelitian ini juga diharapkan dapat memberikan wawasan bagi petarung, pelatih, analis olahraga, serta industri taruhan dalam memahami faktor-faktor utama yang mempengaruhi hasil pertandingan UFC.

Dengan semakin berkembangnya teknologi kecerdasan buatan, pendekatan berbasis machine learning dalam analisis olahraga akan terus mengalami peningkatan. Studi ini tidak hanya memberikan kontribusi dalam bidang prediksi hasil pertandingan UFC, tetapi juga membuka peluang untuk penerapan lebih lanjut dalam berbagai disiplin olahraga lainnya.

1.2 Rumusan Masalah

Adapun rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

- a. Bagaimana cara mengidentifikasi fitur yang paling berpengaruh dalam menentukan hasil pertandingan UFC?
- b. Apa saja variabel atau faktor-faktor yang mempengaruhi kemenangan atau kekalahan petarung dalam pertarungan UFC?

1.3 Batasan Masalah

Dalam penelitian ini, ditetapkan sejumlah batasan yang berfungsi sebagai pedoman untuk menjaga fokus dan arah pengembangan penelitian. Adapun batasan-batasan yang diterapkan dalam studi ini adalah sebagai berikut:

- a. Penelitian ini hanya menggunakan data statistik pertarungan tahun 1994-2022 dari Kaggle yang tersedia secara publik dan tidak mempertimbangkan faktor non-statistik seperti kondisi mental petarung, strategi yang diterapkan, atau faktor eksternal lainnya.
- b. Model prediksi yang digunakan dalam penelitian ini hanya berbasis algoritma XGBoost dan Random Forest tanpa melakukan perbandingan dengan algoritma machine learning lainnya.

1.4 Tujuan Penelitian

- a. Mengevaluasi performa kedua model menggunakan metrik evaluasi seperti accuracy, precision, recall, F1-score, dan ROC-AUC, serta membandingkan efektivitas masing-masing algoritma dalam melakukan prediksi.
- b. Mengembangkan fitur-fitur statistik yang relevan dari data pertarungan UFC guna digunakan sebagai input dalam model prediksi.
- c. Membangun dan melatih model machine learning menggunakan algoritma Random Forest dan XGBoost untuk mengklasifikasikan hasil pertandingan UFC (menang atau kalah) berdasarkan data statistik petarung.

1.5 Metodologi

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa statistik pertandingan UFC, seperti:

- a. Statistik petarung (tinggi, berat, jangkauan, umur, dll.)

- b. Rekor kemenangan dan kekalahan
- c. Statistik pertarungan (jumlah pukulan signifikan, takedown, submission, dll.)
- d. Hasil pertandingan (menang/kalah)

2. Preprocessing Data

I. Pembersihan Data:

- a. Menghapus data duplikat dan mengisi nilai yang hilang.

II. Feature Engineering:

- a. Normalisasi dan transformasi data numerik.
- b. Pembuatan fitur tambahan, misalnya rasio kemenangan, rata-rata strike per menit, dll.

III. Encoding Data Kategorikal:

- a. Mengubah variabel kategorikal seperti metode kemenangan (KO/TKO, submission, decision) menjadi format numerik.

3. Eksplorasi dan Analisis Data

- a. Visualisasi distribusi data untuk melihat pola tertentu.
- b. Analisis korelasi antara fitur dan hasil pertarungan.

4. Implementasi Algoritma Machine Learning

I. Pemilihan Model:

- a. Menggunakan **XGBoost** dan **Random Forest** sebagai algoritma utama untuk klasifikasi hasil pertandingan (menang/kalah).

II. Pembagian Dataset:

- a. Split data menjadi **training (80%)** dan **testing (20%)**.

III. **Pelatihan Model:**

- a. Melatih model XGBoost dan Random Forest dengan dataset yang telah diproses.

IV. **Tuning Hyperparameter:**

- a. Menggunakan Grid Search atau Random Search untuk mendapatkan parameter terbaik.

5. **Evaluasi Model**

Menggunakan metrik evaluasi seperti:

- a. **Accuracy:** Persentase prediksi yang benar.
- b. **Precision, Recall, dan F1-Score:** Untuk melihat keseimbangan antara positif dan negatif.
- c. **ROC-AUC Score:** Untuk melihat performa model dalam klasifikasi.

6. **Perbandingan Hasil dan Analisis**

- a. Membandingkan performa **XGBoost** dan **Random Forest** berdasarkan metrik evaluasi.
- b. Menganalisis faktor-faktor yang paling berpengaruh dalam menentukan hasil pertandingan.

7. **Kesimpulan dan Saran**

- a. Menyimpulkan hasil penelitian dan model mana yang memiliki performa terbaik.
- b. Memberikan rekomendasi pengembangan model lebih lanjut, seperti penggunaan deep learning atau tambahan data fitur lainnya.

1.6 Sistematika Penulisan

Laporan tugas akhir ini disusun dengan menggunakan sistematika penulisan yang dapat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini dimulai dengan penjelasan mengenai latar belakang penelitian yang berjudul “PREDIKSI HASIL PERTANDINGAN ULTIMATE FIGHTING CHAMPIONSHIP BERDASARKAN STATISTIK PERTARUNGAN MENGGUNAKAN ALGORITMA XGBOOST & RANDOM FOREST”. Kemudian dalam bab ini juga dibahas penentuan rumusan dan batasan masalah serta penjelasan tujuan penelitian dan metodologi yang digunakan pada penelitian ini. Pada akhir bab ini dijelaskan mengenai sistematika penulisan yang digunakan.

BAB II LANDASAN TEORI

Bab ini berisi tentang teori-teori dan penelitian terdahulu yang menjadi dasar dalam merancang dan mengembangkan tugas akhir. Teori yang dibahas meliputi penjelasan mengenai UFC dan statistik pertarungan, konsep machine learning, algoritma Random Forest, algoritma XGBoost, serta metrik evaluasi model seperti accuracy, precision, recall, F1-score, dan ROC-AUC. Selain itu, juga dibahas beberapa studi terdahulu yang relevan dengan topik penelitian ini.

BAB III METODOLOGI

Bab ini menjelaskan tahapan dan proses dalam pelaksanaan penelitian. Penjelasan dimulai dari pengumpulan data statistik UFC, proses preprocessing data, feature engineering untuk menghitung skor keterampilan petarung, perhitungan form score, hingga proses implementasi dan pelatihan model menggunakan algoritma Random Forest dan XGBoost. Selain itu, dijelaskan pula proses evaluasi model berdasarkan metrik evaluasi klasifikasi.

BAB IV HASIL DAN ANALISIS

Bab ini berisi hasil implementasi dari model prediksi yang telah dibangun, termasuk visualisasi dan analisis dari hasil pengujian model. Model Random Forest dan XGBoost dievaluasi dan dibandingkan berdasarkan akurasi, precision, recall, F1-score, dan ROC-AUC. Analisis dilakukan untuk mengetahui kelebihan dan kekurangan masing-masing model serta faktor-faktor statistik yang paling berpengaruh terhadap hasil pertandingan UFC.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab penutup yang berisi kesimpulan dari hasil penelitian, yaitu model mana yang memberikan performa lebih baik dalam memprediksi hasil pertandingan UFC. Selain itu, juga disampaikan saran-saran untuk pengembangan penelitian lebih

lanjut, seperti penggunaan metode deep learning atau integrasi data non-statistik yang lebih kompleks untuk meningkatkan akurasi model di masa depan.

