

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Sepak bola merupakan salah satu olahraga paling populer di dunia yang telah bertransformasi menjadi sebuah industri yang terintegrasi dengan hiburan, wisata, dan gaya hidup (Syahputra, 2016). Kompetisi profesional seperti *English Premier League* (EPL) menjadi salah satu ajang paling bergengsi dengan basis data pertandingan yang sangat kaya. Faktor-faktor statistik, seperti jumlah tembakan, penguasaan bola, pelanggaran, dan kemenangan beruntun, dapat menjadi indikator penting dalam menentukan hasil pertandingan (Fardani dkk., 2025). Namun, analisis performa tim kerap masih didominasi pendekatan subjektif oleh pelatih dan pengamat sehingga potensi data statistik yang melimpah belum sepenuhnya dimanfaatkan untuk mendukung pengambilan keputusan berbasis bukti (Sedi dkk., 2023). Dengan adanya perkembangan teknologi, pemanfaatan *machine learning* dalam menganalisis data pertandingan semakin relevan untuk melakukan klasifikasi hasil pertandingan sepak bola.

Sejumlah penelitian terdahulu telah membuktikan potensi model *machine learning* dalam mengklasifikasikan hasil pertandingan sepak bola. Pada penelitian terdahulu menunjukkan bahwa model *Random Forest* dapat mengolah data statistik untuk memprediksi hasil pertandingan dengan akurasi sebesar 68,55% (Alfredo & Isa, 2019). Penelitian lain menggunakan algoritma *Naive Bayes* dan *K-Nearest Neighbor* (KNN) juga menunjukkan kemampuan klasifikasi dengan akurasi tertinggi pada *Naive Bayes* sebesar 63,5%. Terdapat keterbatasan pada penelitian sebelumnya, yaitu permasalahan dalam menangani data sekuensial serta pola kompleks antar atribut. Keterbatasan ini menjadi dasar penting untuk mengeksplorasi algoritma yang lebih baik dalam menangkap dinamika data pertandingan, terutama dalam kasus satu klub seperti *Manchester United* (Prabowo, 2020).

Seiring dengan kemajuan teknologi *machine learning*, pendekatan *deep learning* telah menjadi fondasi penting dalam pengembangan model analisis data berskala besar dan kompleks. *Deep learning* bekerja dengan memanfaatkan lapisan jaringan saraf yang dalam untuk mengekstraksi representasi fitur secara hierarkis sehingga mampu menangkap pola-pola nonlinier yang sulit dipelajari oleh algoritma pembelajaran tradisional (Maszeri dkk., 2025). Salah satu arsitektur dalam *deep learning* yang secara khusus dirancang untuk memproses data berurutan adalah *Recurrent Neural Network* (RNN). RNN memiliki kemampuan untuk menyimpan informasi dari langkah waktu sebelumnya dan menggunakannya untuk memengaruhi keluaran berikutnya, sehingga cocok diterapkan pada data yang bersifat sekuensial seperti deret waktu dan teks. Meskipun demikian, RNN konvensional masih memiliki keterbatasan dalam mempertahankan informasi jangka panjang akibat permasalahan hilangnya gradien, yang menghambat proses pembelajaran terhadap ketergantungan antar waktu (Schmidt, 2019).

Long Short-Term Memory (LSTM) dikembangkan sebagai solusi terhadap permasalahan hilangnya gradien pada RNN konvensional. Secara struktur, LSTM menyerupai RNN, namun sel *rekurens* pada RNN digantikan dengan sel LSTM yang mampu mempertahankan informasi dalam periode waktu yang lebih panjang. Sama halnya dengan sel pada RNN biasa, sel LSTM tetap memproses keluaran sebelumnya, tetapi memiliki tambahan kemampuan untuk memantau *cell state*, yaitu sebuah vektor yang berfungsi sebagai penyimpan memori jangka panjang (Vennerød dkk., 2021). Hal ini menjadikan LSTM sangat cocok digunakan dalam klasifikasi hasil pertandingan sepak bola yang secara alami berbentuk deret waktu, misalnya tren kemenangan beruntun, pola tembakan per laga, atau konsistensi penguasaan bola. Penggunaan model yang mampu menangani data berurutan diyakini dapat menghasilkan prediksi lebih akurat dibandingkan metode tradisional (Wakhid dkk., 2022). Oleh karena itu, penerapan LSTM berpotensi menjadi solusi yang efektif untuk mengatasi keterbatasan penelitian sebelumnya yang hanya menggunakan algoritma berbasis klasifikasi statis.

Penerapan LSTM sudah dilakukan dalam penelitian sebelumnya dan terbukti memberikan hasil yang baik. Cahuantzi dkk. (2023) membandingkan kinerja LSTM dan *Gated Recurrent Unit* (GRU) dalam mempelajari urutan simbolik dengan tingkat kompleksitas berbeda, menggunakan metrik kesamaan teks *Damerau-Levenshtein* (DL) dan *Jaro-Winkler* (JW). Hasil penelitian menunjukkan bahwa pada urutan dengan kompleksitas tinggi, LSTM memperoleh skor kesamaan teks rata-rata 1,0 dengan waktu pelatihan median 12,53 detik, sedangkan GRU membutuhkan waktu lebih lama yaitu 22,84 detik untuk tingkat akurasi serupa. Temuan ini menegaskan bahwa LSTM lebih unggul dalam menangani urutan kompleks karena kemampuannya menangkap ketergantungan jangka panjang secara lebih efektif. Pada penelitian lain Khumaidi dkk. (2021) berhasil mengimplementasikan model LSTM dengan 1000 *epoch* untuk memprediksi kualitas udara dan suhu di Kota Bandung berdasarkan parameter *Particulate Matter* (PM10), Indeks Standar Pencemaran Udara (ISPU), suhu, dan kelembaban. Hasil pengujian menunjukkan bahwa LSTM menghasilkan performa prediksi yang baik pada tiga parameter utama, yakni suhu, kelembaban, dan ISPU, dengan nilai *Root Mean Squared Error* (RMSE) masing-masing sebesar 3,15; 6,97; dan 1,85, yang seluruhnya lebih kecil dari standar deviasi *dataset* uji (4,40; 9,32; dan 1,90). Temuan ini menunjukkan bahwa model LSTM efektif dalam menangani data runtun waktu serta mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi.

Berdasarkan hal tersebut, penelitian ini akan berfokus pada analisis kinerja model LSTM untuk mengklasifikasikan hasil pertandingan sepak bola klub *Manchester United*. Model LSTM akan dilatih dan diuji menggunakan data statistik pertandingan klub *Manchester United* pada ajang EPL dari tahun 2000 hingga 2025. Data historis tersebut menyediakan cakupan waktu yang panjang dan beragam sehingga memungkinkan algoritma LSTM untuk belajar dari variasi performa tim dalam berbagai kondisi kompetisi. Dengan mengolah atribut seperti jumlah tembakan, jumlah pelanggaran, hingga rekor kemenangan, diharapkan model dapat mengenali pola signifikan yang menentukan hasil pertandingan. Penekanan pada satu klub diharapkan

menghasilkan model prediktif yang lebih akurat dibandingkan pelatihan dengan data seluruh pertandingan liga, seperti yang dilakukan dalam penelitian sebelumnya (Alfredo & Isa, 2019). Pendekatan ini memungkinkan analisis yang lebih spesifik dan kontekstual.

Dengan demikian, penelitian ini bertujuan untuk menganalisis kinerja algoritma LSTM dalam mengklasifikasikan hasil akhir pertandingan berdasarkan data statistik pada saat pertandingan berlangsung. Hasil dari penelitian ini diharapkan tidak hanya memberikan kontribusi berupa model prediksi berbasis data yang lebih akurat, tetapi juga menjadi referensi untuk penelitian lanjutan dalam bidang analisis olahraga ataupun teknologi. Selain itu, penerapan LSTM dapat memperluas wawasan mengenai efektivitas model *deep learning* dalam konteks klasifikasi hasil pertandingan, sekaligus menjadi perbandingan dengan metode baseline yang lebih sederhana seperti KNN.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah disebutkan di atas, maka didapatkan permasalahan yang dapat diidentifikasi yaitu:

- 1) Klasifikasi hasil pertandingan sepak bola yang masih menggunakan metode konvensional belum mampu menangkap pola data pertandingan yang kompleks dan sekuensial secara optimal.
- 2) Model prediksi yang telah digunakan dalam penelitian terdahulu masih memiliki keterbatasan dalam menangkap hubungan sekuensial antar atribut statistik pertandingan.
- 3) RNN konvensional memiliki kelemahan berupa hilangnya gradien, yang menghambat kemampuannya dalam mempelajari hubungan jangka panjang pada data pertandingan.

1.3 Pembatasan Masalah

Berdasarkan latar belakang dan identifikasi masalah yang telah disampaikan, maka didapatkan pembatasan masalah, yaitu:

- 1) Penelitian ini menggunakan algoritma LSTM untuk tugas klasifikasi hasil akhir pertandingan klub *Manchester United* menggunakan data statistik ketika pertandingan berlangsung.
- 2) Sumber data yang dianalisis adalah statistik pertandingan dari ajang EPL dalam rentang waktu pada tahun 2000 hingga 2025 melalui situs web www.football-data.uk.
- 3) Evaluasi penelitian hanya berfokus pada pengukuran kinerja model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-Score*, bukan pada implementasi praktisnya di lapangan.

1.4 Rumusan Masalah

Berdasarkan latar belakang, identifikasi masalah, dan pembatasan masalah tersebut, maka didapatkan sebuah rumusan masalah yaitu “Bagaimana kinerja model LSTM dalam mengklasifikasikan hasil akhir pertandingan klub *Manchester United* berdasarkan data statistik historis?”

1.5 Tujuan Penelitian

Berdasarkan uraian di atas, maka tujuan dari penelitian ini adalah untuk menganalisis dan mengukur sejauh mana kinerja model LSTM dalam melakukan klasifikasi hasil akhir pertandingan klub *Manchester United*.

1.6 Manfaat Penelitian

Dari penelitian ini maka didapatkan manfaat dari penelitian yaitu:

1.6.1 Manfaat Teoritis

- 1) Memberikan kontribusi dalam literatur *machine learning* melalui penerapan algoritma LSTM pada klasifikasi hasil pertandingan sepak bola.
- 2) Menambah pemahaman mengenai efektivitas LSTM dibandingkan metode tradisional seperti *Naive Bayes* dan KNN dalam mengolah data sekuensial.

- 3) Menjadi dasar pengembangan kajian akademik mengenai penggunaan *deep learning* pada analisis data olahraga.

1.6.2 Manfaat Praktis

- 1) Menyediakan model prediksi berbasis data statistik yang dapat digunakan sebagai masukan dalam strategi dan persiapan pertandingan.
- 2) Menawarkan acuan prediksi yang lebih objektif dan efektif untuk menganalisis peluang kemenangan klub favorit.
- 3) Menjadi referensi praktis bagi penelitian dan pengembangan sistem prediksi hasil pertandingan sepak bola di masa mendatang.

