

PREDIKSI HASIL PERTANDINGAN SEPAK BOLA MENGGUNAKAN METODE FNN DAN LSTM

Dimas Saputra¹, Muhammad Dawam Fakhri², Archamul Fajar Pratama³, Anggraini Puspita Sari⁴

^{1,2,3,4} Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur

{¹21081010151, ²21081010256, ³21081010178}@student.upnjatim.ac.id

⁴anggraini.puspita.if@upnjatim.ac.id

*Corresponding author email: anggraini.puspita.if@upnjatim.ac.id

Abstrak- Penelitian ini mengembangkan program menggunakan jaringan syaraf tiruan FNN untuk memprediksi persentase pemenang pertandingan sepak bola. Menggunakan data historis pertandingan, program ini menghasilkan prediksi berdasarkan faktor seperti histori permainan tim, tempat bertanding, dan tahun pertandingan. Evaluasi menunjukkan bahwa metode FNN berhasil mencapai tingkat keberhasilan sebesar 84% dalam memberikan informasi berharga bagi penggemar, pelatih, dan petaruh, sedangkan LSTM mencapai tingkat keberhasilan sebesar 76%. Potensi pengembangan meliputi mempertimbangkan faktor lain seperti tim yang bertanding, skor kedua tim, tempat diadakannya pertandingan, dan faktor-faktor lainnya yang mempengaruhi hasil pertandingan. Penggunaan jaringan syaraf tiruan FNN memberikan manfaat praktis dalam perencanaan strategi tim dan pengambilan keputusan yang efektif daripada LSTM.

Kata kunci- program, jaringan syaraf tiruan, FNN, LSTM, prediksi, sepak bola.

I. PENDAHULUAN

Sepak bola memiliki banyak sekali peminat dari berbagai kalangan. Sepak bola juga menjadi industri dengan berbagai aspek seperti bisnis, hiburan, perkembangan teknologi pengetahuan sepak bola. Dari hal ini, mulai banyak prediksi hasil pertandingan sepak bola yang dinikmati banyak kalangan penggemar dan para analis. [3]. Pada dasarnya kecerdasan buatan merupakan ilmu yang mengambil cara berpikir seperti manusia. Biasanya kecerdasan buatan digunakan untuk menyelesaikan masalah perhitungan seperti bisnis, matematika dan sebagainya. Salah satu perhitungan yang menggunakan kecerdasan buatan ialah memprediksi kemenangan pertandingan sepak bola. [4].

Penelitian ini mengembangkan program menggunakan jaringan syaraf tiruan dengan metode FNN dan LSTM untuk memprediksi persentase kemenangan dalam pertandingan sepak bola. Program ini menggabungkan kedua metode tersebut untuk meningkatkan pemahaman dan prediksi dalam dunia sepak bola. Diharapkan program ini dapat membantu

penggemar sepak bola dan pemangku kepentingan dalam membuat keputusan yang lebih cerdas dalam konteks pertandingan sepak bola.

Penelitian ini menunjukkan bahwa FNN memiliki akurasi yang lebih tinggi daripada LSTM dalam memprediksi hasil pertandingan sepak bola pada dataset sederhana dengan jumlah data normal. Namun, LSTM lebih baik dalam memodelkan hubungan temporal yang kompleks. Akurasi program juga dipengaruhi oleh faktor lain seperti kualitas data dan pemilihan fitur. Oleh karena itu, tidak ada metode yang secara pasti lebih baik, karena tergantung pada karakteristik data dan kompleksitas hubungan dalam pertandingan sepak bola. Penelitian ini memberikan implikasi penting dalam penggunaan kecerdasan buatan untuk meningkatkan pemahaman dalam dunia sepak bola.

II. PENELITIAN TERKAIT

Penelitian ini juga pernah diteliti oleh Hanan Risma dari Program Studi Informatika, Universitas Islam Indonesia. Penelitian ini berfokus pada penggunaan metode fuzzy untuk memprediksi pertandingan dalam liga sepak bola Bundesliga. Bundesliga dipilih karena memiliki basis penggemar yang luas di seluruh dunia. Tujuan penelitian ini adalah mencari nilai akurasi terbaik menggunakan metode fuzzy. Metode fuzzy dipilih karena memiliki stabilitas persentase yang cukup baik dibandingkan metode lain. Pemilihan Bundesliga sebagai kasus penelitian ini dilakukan karena masih kurangnya penelitian yang menggunakan Bundesliga sebagai fokusnya. Selain itu, Bundesliga dipilih karena persaingan yang ketat hingga akhir musim, dengan perubahan yang sering terjadi pada tim-tim, yang dapat berpengaruh pada kualitas tim di awal musim dan memunculkan kejutan dalam klasemen [5].

Penelitian ini juga pernah dibuat oleh Muhammad Rafli Pratama dari jurusan Teknologi Informasi, Universitas Teknokrat Indonesia yang membahas penggunaan algoritma Hybrid PSO-JST (Particle Swarm Optimization - Jaringan Syaraf Tiruan) untuk memprediksi hasil pertandingan sepakbola. Tujuan utama penelitian ini adalah meningkatkan pendekatan terhadap data target yang ditentukan dengan menggunakan dua metode yaitu Jaringan Syaraf Tiruan dan Particle Swarm Optimization.[14]

Penelitian yang sama juga dilakukan oleh Yonata dkk dari Universitas Prima Indonesia menggunakan metode data mining dengan algoritma C.45 untuk memprediksi klub pemenang dalam kompetisi Liga Champion. Liga Champion adalah turnamen sepak bola antar klub yang bergengsi dan diminati oleh penggemar sepak bola di seluruh dunia. Penelitian ini bertujuan untuk membantu penggemar sepak bola dalam memprediksi klub mana yang akan menjadi pemenang menggunakan teknik data mining. Metode ini menggunakan algoritma C.45 yang terkenal dalam mengklasifikasikan dan memprediksi [12].

III. METODOLOGI PENELITIAN

1. Data Mining

Dalam penelitian ini untuk memprediksi pemenang sepak bola tentunya kita memerlukan data riwayat sepak bola sebagai dasar variabel penentu untuk menentukan rasio pemenang dalam pertandingan nantinya. Oleh karena itu kita perlu melakukan proses data mining untuk memilih dataset yang baik dan berkualitas.

Data mining adalah suatu teknik yang sangat berharga dalam memperkirakan berbagai fenomena yang terkait dengan objek yang sedang diteliti. Melalui analisis data yang besar dan kompleks, metode ini mampu membantu kita dalam memprediksi berbagai hal, baik dengan memiliki informasi awal maupun tanpa pengetahuan sebelumnya mengenai kelas tertentu dari objek yang sedang dipelajari. Dengan memanfaatkan data mining, masyarakat dapat mengoptimalkan pengelolaan lingkungan mereka, sehingga menciptakan perubahan yang lebih baik dan berkelanjutan [6].

Dibandingkan dengan paper penelitian sebelumnya pada sitasi data mining di atas, penelitian ini memiliki kelebihan dalam mengolah dataset yang berupa angka menjadi hasil prediksi menggunakan metode FNN dan LSTM di mana pada penelitian sebelumnya hanya memfokuskan pada data mining saja. Saran untuk penelitian pada paper selanjutnya dalam data mining yaitu penggunaan data variabel yang detail serta juga memiliki tujuan dan makna yang jelas kedepannya ingin diolah untuk apa.

2. Neural Network

Jaringan Syaraf Tiruan (Artificial Neural Network) adalah model kompleks yang non-linear. Dibangun dari komponen mirip dengan model regresi, jaringan ini dapat divisualisasikan sebagai grafik dengan beberapa sub-grafik yang memiliki perilaku serupa. Meskipun struktur jaringan saraf dirancang sebelumnya, pengolahan jaringan ini berkembang selama proses pembelajaran dan tidak hanya untuk menghasilkan hipotesis [7].

Neural network sendiri memiliki berbagai sub-metode misalnya FNN, CNN, LSTM, dan lainnya. Pada penelitian kali ini kami menggunakan dua metode untuk dibandingkan yaitu metode FNN dan LSTM dalam memprediksi persentase pemenang dalam pertandingan sepak bola.

3. Feedforward Neural Network

FNN adalah model sederhana untuk masalah data time series. Terdiri dari berbagai layer, yaitu input layer, hidden layer, dan output layer. Input layer terhubung dengan hidden layer, yang kemudian terhubung dengan output layer. Tidak ada hubungan langsung antara input dan output layer. FNN dapat mempelajari pola kompleks dalam data time series dan memberikan prediksi atau klasifikasi yang akurat [8].

Metode FNN ini nantinya akan kami gunakan sebagai model pertama dalam penelitian ini yang nantinya akan dibandingkan dengan metode LSTM.

4. Long Short Term Memory

LSTM (Long Short Term Memory) merupakan evolusi dari RNN (Recurrent Neural Network) yang diperkenalkan oleh Hochreiter & Schmidhuber pada tahun 1997. Arsitektur LSTM terus dikembangkan dan digunakan dalam berbagai bidang seperti speech recognition dan forecasting. LSTM memiliki mekanisme "gate" yang memungkinkan pemrosesan yang lebih baik pada urutan data panjang dan kompleks [9][2].

Keunggulan menggunakan LSTM adalah LSTM dilatih menggunakan algoritma BP, meningkatkan kemampuan RNN, jalannya program lebih cepat daripada menggunakan RNN, dan memiliki cell state (C) dan gates untuk mengatur aliran informasi [1].

5. Confusion Matrix

Confusion matrix adalah metode untuk mengukur performa model machine learning. Terdiri dari matriks 2x2, menggunakan istilah True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). TP adalah data positif yang benar diklasifikasikan, FP adalah data negatif yang salah diklasifikasikan, TN adalah data negatif yang benar diklasifikasikan, dan FN adalah data positif yang salah diklasifikasikan. Confusion matrix membantu menghitung akurasi, presisi, recall, dan F1-score untuk mengevaluasi performa model dengan lebih detail [10].

TABEL 1
CONFUSION MATRIX

Aktual	Kelas	Prediksi	
		A	B
	A	TP	FP
	B	FN	TN

Dari tabel confusion matrix tersebut kita dapat menghitung akurasi dari hasil uji model yang kita latih dengan menggunakan perhitungan.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

Variabel

TP (*True Positive*), Merupakan data positif yang diprediksi benar.

TN (*True Negative*), Merupakan data negatif yang diprediksi benar.

FP (*False Positive*), Merupakan data negatif namun diprediksi sebagai data positif atau dalam artian salah memprediksi.

FN (*False Negative*), Merupakan data positif namun diprediksi sebagai data negatif atau dalam artian salah memprediksi .

IV. HASIL DAN PEMBAHASAN

1. Implementasi Data

Dataset penelitian yang kami gunakan diambil dari website *kaggle.com* (*International football results from 1872 to 2023*)[11], kami menggunakan dataset *results.csv* pada dataset *kaggle* tersebut. dengan tujuan untuk memprediksi persentase kemenangan pertandingan sepak bola antara dua tim antar negara. Dataset tersebut terdiri dari 9 kolom yaitu *date*, *home team*, *away team*, *home score*, *away score*, *tournament*, *city*, *country*, dan *neutral*. Terdapat 44.341 data pada dataset tersebut, yang nantinya masih difilter sesuai inputan tim dan *range* tahun pertandingan sepak bola tersebut. Dari keseluruhan dataset tersebut dilakukan pembagian dataset dengan rasio 60% sebagai data latih dan 40% sebagai data uji.

2. Pemrosesan Data

Karena dataset yang kami gunakan masih *unsupervised* jadi kami akan membuat variabel target baru yaitu kolom pemenang yang diperoleh dari perbandingan antara skor tim kandang dan skor tim tandang seperti berikut.

```
while(data):
    jika skor_kandang > skor_tandang:
        pemenang = tim_kandang
    selain, jika skor_tandang > skor_kandang:
        pemenang = tim_tandang
    selain,
        pemenang = "draw"
```

Setelah itu juga dilakukan filter agar dataset yang dilatih tidak terlalu banyak. Yaitu dengan cara memfilter kedua tim yang bertanding dan tahun permainan sesuai inputan seperti berikut.

```
# Filter Tim
while(data):
    jika ((tim1 == tim1_input) && (tim2 == tim2_input)):
        return valid
    selain, ((tim1 == tim2_input) && (tim2 == tim1_input)):
        return valid
    return tidak_valid

# Filter Tahun Permainan
while(data):
    jika(data.tahun >= tahun_awal_input):
    jika(data.tahun <= tahun_akhir_input):
        return valid
    return tidak_valid
```

3. Pembentukan Model

Pada penelitian kami menggunakan library tensorflow untuk menggunakan *method* *sequential* untuk membangun layer jaringan saraf tiruan dari model baik FNN dan LSTM.

Tensorflow adalah sebuah perpustakaan (library) sumber terbuka yang dikembangkan oleh tim Google Brain. Fokus utamanya adalah pada komputasi numerik dan machine learning yang berskala besar. Tensorflow menggunakan Python sebagai antarmuka pemrograman yang memungkinkan pengguna untuk membangun aplikasi dengan menggunakan berbagai framework. Salah satu keunggulan Tensorflow adalah kemampuannya dalam melatih dan menjalankan jaringan syaraf tiruan untuk berbagai tugas seperti klasifikasi tulisan tangan, pengenalan gambar, pemrosesan bahasa alami, dan bahkan simulasi berbasis Persamaan Differensial Parsial (PDE) [13].

Pada model dalam metode FNN kami menggunakan 3 layer yaitu,

1. Input layer, berisi 16 unit *Dense* dengan activation *relu*
2. hidden layer, berisi 16 unit *Dense* dengan activation *relu*
3. Output layer, berisi 3 unit *Dense* yang berarti akan menghasilkan 3 output berupa persentase tim1 menang, tim2 menang, dan draw. Mempunyai activation *sigmoid*.

Berikut ringkasan dari model FNN yang kami gunakan pada penelitian ini.

Layer (type)	Output Shape	Param #
input (Dense)	(None, 16)	112
hidden_layer (Dense)	(None, 16)	272
output (Dense)	(None, 3)	51

Pada model dalam metode LSTM kami menggunakan 4 layer yaitu,

1. Reshape layer, dengan *target_shape* (6,1) dan *input_shape* (6,none). Layer ini ditujukan untuk

- membentuk ulang *shape* data dari semula (6,none) menjadi (6,1) agar sesuai dengan konsep LSTM.
- LSTM layer, berisi 64 unit di dalam layernya.
 - hidden layer, berisi 32 unit *Dense* dengan activation *relu*
 - Output layer, berisi 3 unit *Dense* yang berarti akan menghasilkan 3 output berupa persentase tim1 menang, tim2 menang, dan draw. Mempunyai activation *sigmoid*.

Berikut ringkasan dari model LSTM yang kami gunakan pada penelitian ini.

Layer (type)	Output Shape	Param #
reshape (Reshape)	(None, 6, 1)	0
lstm (LSTM)	(None, 64)	16896
dense (Dense)	(None, 32)	2080
dense_1 (Dense)	(None, 3)	99

4. Pengujian Sistem

Dalam pengujian penelitian kedua model tersebut, kami menggunakan inputan yang telah kami sepakati untuk diuji sebelumnya, Yaitu

- tim1 = "Argentina"
- tim2 = "Brazil"
- tahun_awal = 2000
- tahun_akhir = 2020

Dari filter tersebut, maka kami memperoleh sebanyak 23 data untuk dilatih dan diuji sesuai dengan penjelasan rasio data latih dan data uji sebelumnya.

Setelah melakukan pelatihan menggunakan dua model yaitu model FNN dan LSTM, diperoleh dua buah hasil dari data uji.

Berikut perbandingan hasil prediksi data uji dari model FNN dan LSTM.

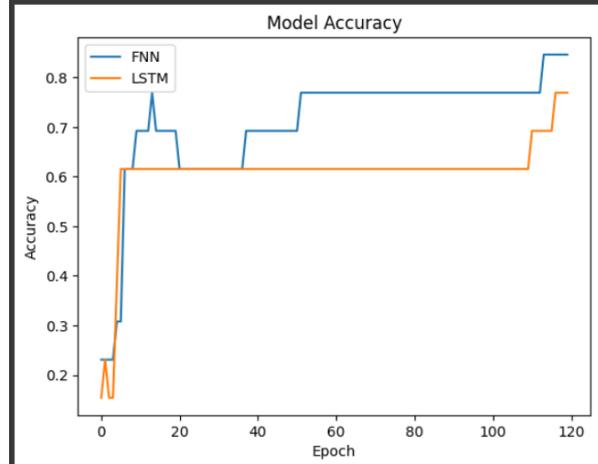
	Actual Winner (Dataset)	FNN Prediction	LSTM Prediction
0	draw	Argentina	Argentina
1	Argentina	Brazil	Brazil
2	Brazil	Brazil	Brazil
3	Brazil	Brazil	Brazil
4	Brazil	Brazil	Brazil
5	Argentina	Argentina	Argentina
6	Argentina	Brazil	Brazil
7	Argentina	Brazil	Brazil
8	Brazil	Brazil	Brazil
9	draw	Brazil	Brazil

Berikut adalah perbandingan akurasi antara model FNN dan LSTM yang diperoleh dari hasil uji data uji sebelumnya.

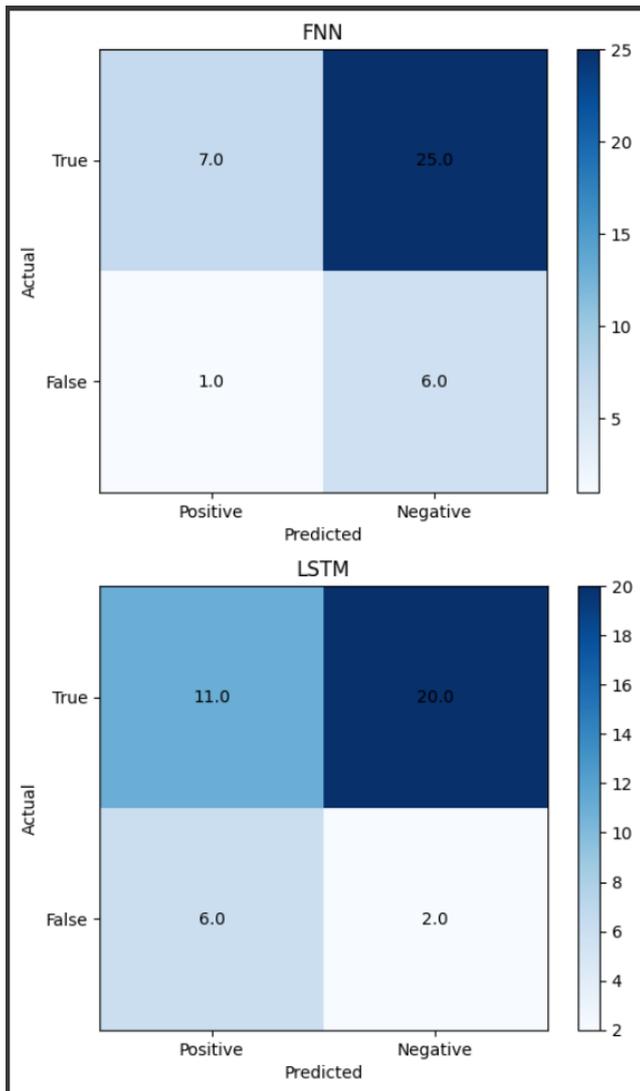
Akurasi (FNN): 84.61538553237915%
Loss (FNN): 52.53410339355469%

Akurasi (LSTM): 76.92307829856873%
Loss (LSTM): 65.39058089256287%

Selisih Persentase Akurasi antara FNN dan LSTM adalah: 7.692307233810425%
Selisih Persentase Loss antara FNN dan LSTM adalah: 12.856477499008179%

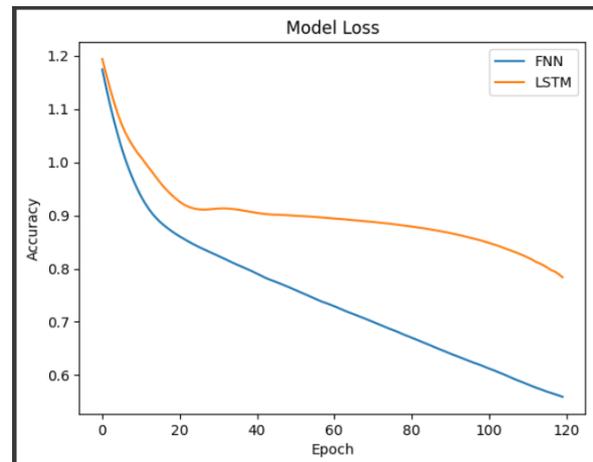


Berikut adalah tabel *confusion matrix* yang diperoleh dari hasil pengujian data uji kedua model tersebut. Terlihat dimana FNN memiliki prediksi (Label positive) lebih banyak daripada model LSTM sesuai dengan akurasi yang tertera sebelumnya.



untuk mengoptimalkan parameter model agar menghasilkan prediksi yang semakin akurat.[15]

Berikut perbandingan loss kedua model dari hasil pengujian data uji.



IV. KESIMPULAN

Penerapan FNN dan LSTM pada data pertandingan sepak bola memberikan hasil prediksi yang berbeda. FNN mencapai akurasi 84,6%, sedangkan LSTM mencapai akurasi 76,9%. FNN lebih efektif dalam dataset sederhana dengan jumlah data normal, sementara LSTM dapat lebih baik dalam memodelkan hubungan temporal yang kompleks. Akurasi program juga dipengaruhi oleh faktor lain seperti kualitas data dan pemilihan fitur. Tidak ada jawaban pasti mengenai metode yang lebih baik, karena tergantung pada karakteristik data dan kompleksitas hubungan dalam pertandingan sepak bola.

Perhitungan akurasi model menggunakan *confusion matrix*

FNN

$$\text{Accuracy} = \frac{7+25}{7+25+1+6} * 100\% \\ = 82,01\%$$

LSTM

$$\text{Accuracy} = \frac{11+20}{7+25+1+6} * 100\% \\ = 79,49\%$$

Selain melihat perbandingan kedua model melalui akurasi kita juga dapat melihat perbandingan kedua model tersebut melalui hasil loss modelnya. Loss Function atau Cost Function adalah fungsi matematis yang digunakan untuk mengukur perbedaan atau kesalahan antara nilai yang diprediksi oleh model dengan nilai yang sebenarnya dalam suatu tugas atau masalah. Fungsi ini menggambarkan kerugian yang terkait dengan semua kemungkinan output yang dihasilkan oleh model, dan bertujuan

REFERENSI

- [1] A. P. Sari, E. A. Hakim, D. A. Prasetya, R. Arifuddin, and P. Dani, "Sistem Prediksi Kecepatan dan Arah Angin Menggunakan Bidirectional Long Short-Term Memory," in Seminar Keinsinyuran Program Studi Program Profesi Insinyur, vol. 1, 2021.
- [2] A. P. Sari, D. A. Prasetya, T. Yasuno, A. N. Sihananto, M. M. Al Haromainy, and W. S. Saputra, "Forecasting Model of Wind Speed and Direction by Convolutional Neural Network-Deep Convolutional Long Short Term Memory," in 2022 IEEE 8th Information Technology International Seminar (ITIS), pp. 200-205, October 2022.
- [3] Zein and G. Gunawan, "Prediksi Hasil FIFA World Cup Qatar 2022 Menggunakan Machine Learning dengan Python," Jurnal Riset Matematika, pp. 153-162, 2022.
- [4] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang," J. Khatulistiwa Informatika, vol. 5, no. 1, pp. 490845, 2020.
- [5] H. R. Isriwanto and I. V. Papatungan, "Analisis Kebutuhan Prediksi Pertandingan Bundesliga Menggunakan Metode Fuzzy," in Proceedings of the XYZ Conference, 2021, pp. 123-128.
- [6] A. P. Natasuwarna, "Tantangan menghadapi era revolusi 4.0-big data dan data mining," SINDIMAS, vol. 1, no. 1, pp. 23-27, 2019.

- [7] N. Hadiano, H. B. Novitasari, and A. Rahmawati, "Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 163-170, 2019.
- [8] N. A. Izati, B. Warsito, and T. Widiyarih, "Prediksi Harga Emas Menggunakan Feed Forward Neural Network Dengan Metode Extreme Learning Machine," *J. Gaussian*, vol. 8, no. 2, pp. 171-183, 2019.
- [9] A. W. P. Aldi, J. Jondri, and A. Aditsania, "Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin," in *eProceedings of Engineering*, vol. 5, no. 2, 2018.
- [10] S. C. Esananda, B. Nugroho, and F. T. Anggraeny, "Implementasi Fase Boosting pada Algoritma C5.0 dalam Menentukan Prestasi Akademik Siswa," in *Prosiding Seminar Nasional Informatika Bela Negara*, vol. 2, pp. 1-6, November 2021.
- [11] J. Mart, "International Football Results from 1872 to 2017," 2017. [Online]. Available: <https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017>.
- [12] C. Tandian, Y. Laia, and A. Saputra, "Penerapan Data Mining Dalam Memprediksi Pemenang Klub Sepak Bola Pada Ajang Liga Champion Dengan Algoritma C.45," *J. Sist. Inf. dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 91-98, 2019.
- [13] R. F. Muharram, "Implementasi Artificial Intelligence Untuk Deteksi Masker Secara Realtime Dengan Tensorflow Dan SSD Mobilenet Berbasis Python," *JRKT (Jurnal Rekayasa Komputasi Terapan)*, vol. 1, no. 03, 2021.
- [14] M. R. Pratama, "Model Prediksi Hasil Pertandingan Sepakbola Menggunakan Kombinasi Particle Swarm Optimization-Neural Network Backpropagation," 2022.
- [15] K. P. Danukusumo, "Implementasi deep learning menggunakan convolutional neural network untuk klasifikasi citra candi berbasis GPU," Doctoral dissertation, UAJY, 2017.