

**IMPLEMENTASI MODEL *MACHINE LEARNING* ALGORITMA
REGRESI LINIER DAN *REGRESI LOGISTIC* PADA
PREDIKSI HASIL PERTANDINGAN SEPAK
BOLA PADA LIGA 1 INDONESIA
2024/2025**

SKRIPSI

HARYANDA FIDI

201401139



**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

UNIVERSITAS SUMATERA UTARA

**IMPLEMENTASI MODEL *MACHINE LEARNING* ALGORITMA
REGRESI LINIER DAN *REGRESI LOGISTIC* PADA
PREDIKSI HASIL PERTANDINGAN SEPAK
BOLA PADA LIGA 1 INDONESIA
2024/2025**

SKRIPSI

HARYANDA FIDI

201401139



**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

UNIVERSITAS SUMATERA UTARA

PERSETUJUAN

Judul : IMPLEMENTASI MODEL *MACHINE LEARNING* ALGORITMA REGRESI LINIER DAN REGRESI LOGISTIC PADA PREDIKSI HASIL PERTANDINGAN SEPAK BOLA PADA LIGA 1 INDONESIA 2024/2025

Kategori : Skripsi

Nama : Haryanda Fidi ,

Nomor Induk Mahasiswa : 201401139

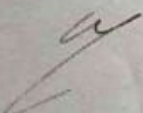
Program Studi : Sarjana (S-1) Ilmu Komputer

Fakultas : Ilmu Komputer dan Teknologi Informasi


Universitas Sumatera Utara

Komisi Pembimbing :

Dosen Pembimbing I


Dr. Ir. Elviawaty Muisa Zamzami
S.T., M.T., M.M., IPU
NIP. 197007162005012000

Dosen Pembimbing II


Handrizal S.Si., M.Comp.Sc
NIP. 197706132017061001

Diketahui/disetujui oleh
Program Studi S1 Ilmu Komputer Ketua,



UNIVERSITAS SUMATERA UTARA

PERNYATAAN

**IMPLEMENTASI MODEL *MACHINE LEARNING* ALGORITMA
REGRESI LINIER DAN REGRESI LOGISTIC PADA
PREDIKSI HASIL PERTANDINGAN SEPAK
BOLA PADA LIGA 1 INDONESIA
2024/2025**

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.



Medan, Desember 2024

201401139

KATA PENGANTAR

Dengan memanjatkan puji dan syukur kepada Allah SWT atas rahmat, petunjuk, berkat., dan kekuatan yang diberikannya selama penulisan skripsi ini.. Terimakasih atas rahmat-Nya, yang juga senantiasa membantu, membimbing, dan mendorong saya untuk menyelesaikan skripsi yang diperlukan untuk mendapatkan gelar sarjana komputer di Program S1 Ilmu Komputer dan Teknologi Informasi di Universitas Sumatera Utara. Selain itu, kita bershalawat dan salam kepada Rasulullah SAW yang telah membawa umat manusia dari masa kegelapan ke masa kebenaran.

Penulis dengan hormat mengucapkan terima kasih kepada Mama tersayang, atas segala bentuk perjuangan, kasih sayang, dan perlindungan dengan doa-doa yang dipanjatkan untuk penulis. Dan terimakasih kepada , atas dukungan dan kasih sayang yang membersamai di setiap langkah penulis. Terima kasih untuk setiap dukungan yang telah diberikan hingga penulis dapat berada di titik ini.

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bantuan., mendukung, dan membantu untuk penyusunan skripsi ini. Terimakasih penulis ucapkan kepada :

1. Bapak Prof. Muryanto Amin, S.Sos., M.Si selaku Rektor Universitas Sumatera
2. Ibu Dr. Maya Silvi Lydia B.Sc., M.Sc. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
3. Ibu Dr. Amalia, S.T., M.T selaku Ketua Program Studi S1 Ilmu Komputer Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara dalam memotivasi dan memberikan arahan semasa perkuliahan.
4. Ibu Dr.Ir. Elviawaty Muisa Zamzami S.T., M.T.,M.M.,IPU sebagai Dosen Pembimbing I, telah memberikan banyak masukan, mendorong, dan mendukung penulis selama penyusunan skripsi ini.
5. Bapak Handrizal S.Si., M.Comp.Sc., sebagai dosen pembimbing II,

telah memberikan banyak pengetahuan dan masukan kepada skripsi penulis.

6. Ibu Hayatunnufus S.Kom., M.Kom selaku Dosen Pembimbing Akademik yang telah memberikan saran, motivasi dan banyak dukungan kepada penulis selama masa studi.
7. Seluruh Bapak dan Ibu Dosen Program Studi S-1 Ilmu Komputer, yang telah membimbing saya selama masa perkuliahan hingga akhir

Penulis berharap hasil penelitian ini mampu memberikan kontribusi dan keuntungan yang signifikan bagi kemajuan ilmu pengetahuan dan teknologi. Penulis menyadari bahwa penelitian ini memiliki keterbatasan dan ruang untuk pengembangan berikutnya. Oleh karena itu, penulis berharap penelitian ini akan menjadi dasar untuk penelitian berikutnya dan menginspirasi penelitian lain.

Akhir kata, Terima kasih kepada saya sendiri yang sudah berjuang dan tidak menyerah sampai sejauh ini sehingga dapat menyelesaikan skripsi ini. Semoga pembaca dan masyarakat luas mendapat manfaat dari skripsi ini. Penulis berharap bahwa hasil penelitian ini akan memberikan manfaat yang besar.

Medan, Desember 2024

Penulis,



Haryanda Fidi

ABSTRAK

Sepak bola adalah olahraga yang paling populer di dunia, termasuk di Indonesia. Liga 1 Indonesia, sebagai kompetisi sepak bola tertinggi di tanah air, menarik perhatian jutaan penggemar dan pelaku industri. Namun, memprediksi hasil pertandingan sepak bola merupakan tantangan yang kompleks karena melibatkan banyak variabel seperti performa tim, statistik pemain, dan kondisi pertandingan. Penelitian ini bertujuan untuk mengembangkan model prediksi hasil pertandingan Liga 1 Indonesia musim 2024/2025 dengan menerapkan algoritma *regresi linier* dan regresi logistik. Metode yang digunakan melibatkan pengumpulan data historis pertandingan, meliputi data skor, jumlah kemenangan, kekalahan, hasil seri, serta variabel lain seperti jumlah tembakan, penguasaan bola, dan statistik individual pemain. Data tersebut diolah melalui proses pembersihan, normalisasi, dan analisis fitur untuk memastikan kualitasnya. Model *regresi linier* digunakan untuk memprediksi skor pertandingan, sedangkan regresi logistik digunakan untuk menentukan hasil pertandingan dalam kategori menang, kalah, atau seri. Evaluasi kinerja model dilakukan menggunakan metrik seperti *Mean Squared Error* (MSE) untuk *regresi linier* dan akurasi, *precision*, *recall*, serta *F1-score* untuk regresi logistik. Hasil penelitian menunjukkan bahwa model *regresi linier* mampu memprediksi skor pertandingan dengan nilai MSE sebesar 0,75, sementara regresi logistik mencapai akurasi prediksi hasil pertandingan sebesar 82%. Hasil ini menunjukkan bahwa kombinasi kedua algoritma tersebut efektif dalam memprediksi hasil pertandingan sepak bola Liga 1 Indonesia. Kesimpulan dari penelitian ini adalah model berbasis *machine learning*, khususnya algoritma *regresi linier* dan regresi logistik, dapat menjadi alat yang andal untuk membantu analisis pertandingan sepak bola, memberikan wawasan strategis bagi pelatih, dan meningkatkan pengalaman penggemar melalui prediksi yang lebih akurat.

Kata Kunci: Prediksi Sepak bola, Liga 1 Indonesia, *Mechine Learning*, *Regresi Linier*, *Regresi Logistic*

IMPLEMENTATION OF MACHINE LEARNING MODELS OF LINEAR REGRESSION AND LOGISTIC REGRESSION ALGORITHMS IN PREDICTING THE RESULTS OF THE 2024/2025 INDONESIAN LEAGUE 1 FOOTBALL MATCHES

ABSTRACT

Football is the most popular sport globally, including in Indonesia. Liga 1 Indonesia, as the country's top-tier football competition, garners the attention of millions of fans and industry stakeholders. However, predicting football match outcomes is a complex challenge involving various factors such as team performance, player statistics, and match conditions. This study aims to develop a predictive model for match outcomes in the 2024/2025 Liga 1 Indonesia season by implementing linear regression and logistic regression algorithms. The methodology involves collecting historical match data, including scores, wins, losses, draws, as well as other variables like the number of shots, ball possession, and individual player statistics. The data undergoes cleaning, normalization, and feature analysis to ensure quality. A linear regression model is used to predict match scores, while a logistic regression model is employed to classify match outcomes into win, loss, or draw categories. Model performance is evaluated using metrics such as Mean Squared Error (MSE) for linear regression and accuracy, precision, recall, and F1-score for logistic regression. The results show that the linear regression model predicts match scores with an MSE of 0.75, while the logistic regression model achieves an outcome prediction accuracy of 82%. These findings indicate that the combination of these two algorithms is effective for predicting Liga 1 Indonesia football match outcomes. The study concludes that machine learning-based models, particularly linear and logistic regression algorithms, can serve as reliable tools for match analysis, providing strategic insights for coaches and enhancing fan experiences through more accurate predictions.

Keywords: Football prediction, Liga 1 Indonesia, linear regression, logistic regression, machine learning.

DAFTAR ISI

PERSETUJUAN	i
PERNYATAAN	ii
KATA PENGANTAR	iii
ABSTRAK	v
ABSTARCT	vi
DAFTAR ISI	vii
DAFTARGAMBAR	ix
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH	4
1.4 TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	4
1.6 METODE PENELITIAN	5
1.7 PENELITIAN RELEVAN.....	6
1.8 SISTEMATIKA PENELITIAN	7
BAB II LANDASAN TEORI	9
2.1 PREDIKSI.....	9
2.2 <i>MACHINE LEARNING</i>	9
2.3 <i>SUPPERVISED LEARNIG</i>	10
2.4 ALGORITMA <i>REGRESI LINIER</i>	10
2.5 ALGORITMA <i>REGRESI LOGISTIC</i>	12
2.6 <i>ROOT MEAN SQUARE ERROR</i>	14
2.7 FAKTORFAKTOR HASIL PERTANDINGAN	15
BAB III ANALISIS DAN PERANCANGAN	16
3.1 ARSITEKTUR UMUM.....	16
3.2 PENGUMPULAN DATA SET	17
3.3 <i>PRE-PROCESSING DATA</i>	18

3.4 PEMBAGIAN DATA.....	19
3.5 MODEL <i>REGRESI LINIER</i>	20
3.6 MODEL <i>REGRESI LOGISTIC</i>	22
3.7 EVALUASI	22
3.8 <i>FLOWCHART</i> 23	
BAB IV IMPLEMENTASI DAN PEMBAHASAN	25
4.1 IMPLEMENTASI SISTEM	25
4.2 PENGUJIAN DAN ANALISI SISTEM	26
4.3 ANALISIS KESELURUHAN SISTEM.....	28
4.4 <i>CONFUSION MATRIX</i>	28
BAB V KESIMPULAN DAN SARAN.....	29
5.1 KESIMPULAN	29
5.2 SARAN	30
DAFTAR PUSTAKA.....	31
LAMPIRAN	33



DAFTAR GAMBAR

Gambar 2.1	<i>Regresi linier</i>	11
Gambar 3.1	Arsitektur Umum	16
Gambar 3.2	Dataset Liga 1 Indonesia	18
Gambar 3.3	Proses Perubahan Menjadi Numerik	18
Gambar 3.4	Pembagian Data <i>Train</i> dan <i>Test</i>	20
Gambar 3.5	Data Latih <i>Regresi Linier</i>	20
Gambar 3.5.1	Selisih <i>Home & Away</i>	21
Gambar 3.5.2	Metrik MSE.....	21
Gambar 3.6	Data Latih <i>Regresi Logistic</i>	22
Gambar 3.7.1	Hasil Klasifikasi	23
Gambar 3.8.1	Flowchart <i>Regresi Linier</i>	24
Gambar 3.8.2	Flowchart <i>Regresi Logistic</i>	24
Gambar 4.2.1	Input dataset <i>regresi linier</i>	26
Gambar 4.2.1	Output <i>Regresi linier</i>	27
Gambar 4.2.2	Hasil <i>Output</i>	27
Gambar 4.3	Tingkat Akurasi	28
Gambar 4.4	Matrix prediksi.....	28

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sepak bola, olahraga yang paling banyak digemari di dunia dan antusiasme tanpa batas, tidak hanya menciptakan lapangan hijau tempat aksi penuh gairah terjadi, tetapi juga menjadi bahasa universal yang menghubungkan jutaan jiwa di setiap penjuru planet ini. Pertandingan sepak bola bukan hanya sebuah peristiwa olahraga biasa, tetapi juga fenomena sosial yang memikat perhatian jutaan penggemar di seluruh dunia. Pertandingan sepak bola bukan sekadar kumpulan statistik, tetapi pula panggung pertunjukan emosional yang menarik.

Di Indonesia, Sepak bola telah menjadi bagian integral dari budaya Nusantara sejak masa penjajahan Belanda. Perkembangannya dimulai dari klub-klub amatir hingga terbentuknya PSSI yang menandai tonggak sejarah sepak bola nasional. Setelah melalui berbagai era, seperti Perserikatan dan Galatama, akhirnya lahirlah Liga Indonesia yang menyatukan seluruh kompetisi. Meskipun menghadapi berbagai tantangan, antusiasme masyarakat yang tinggi dan potensi pemain muda berbakat menjadi harapan besar bagi sepak bola Indonesia. Saat ini, Liga 1 menjadi kompetisi paling bergengsi, namun masih perlu adanya peningkatan infrastruktur, pembinaan pemain, dan tata kelola yang lebih baik untuk mencapai prestasi yang lebih gemilang di kancah internasional. Selain itu, pada olahraga sepak bola tidak hanya operan dan tendangan yang harus diperhatikan dalam keberlangsungan suatu pertandingan, tetapi juga turut memperhatikan komposisi dan juga strategi yang harus dimiliki setiap tim agar dapat memaksimalkan setiap pertandingan untuk memperoleh kemenangan.

Oleh karena itu, Klasifikasi diperlukan untuk mengidentifikasi faktor-faktor yang berpengaruh dalam kemenangan tim dalam pertandingan sepak bola. Salah satu pendekatan yang dapat dilakukan adalah dengan menggunakan algoritma yang ada pada *Machine learning*.

Machine Learning (ML) adalah teknologi yang menggunakan algoritma untuk memungkinkan komputer dalam menganalisis data dan memproses data lainnya seperti halnya manusia dalam melakukan pembelajaran yang dapat meningkatkan akurasi dan kapabilitas melalui algoritma itu sendiri (Jafar Alzubi et al., 2018). Model Pembelajaran ML dapat dilatih untuk melakukan klasifikasi, prediksi, dan penambangan interaksi padadata (Guo, 2021).

Pada penelitian ini, jenis algoritma *machine learning* yang akan digunakan adalah algoritma *regresi logistic* dan *regersi linier* sebagai Salah satu metode yang dapat digunakan untuk memprediksi suatu nilai dengan akurat (Prasetyo et al., 2021).

Regresi linier adalah metode peramalan yang meninjau hubungan antara variabel independent terhadap variabel dependent (Anggrawan et al., 2022). Tujuannya adalah untuk menemukan garis lurus (disebut garis regresi) yang paling baik mewakili pola hubungan antara variabel independen dan dependen. Garis ini digunakan untuk membuat prediksi tentang nilai *output* berdasarkan nilai *input* yang diberikan. Dalam konteks prediksi hasil pertandingan sepak bola, *regresi linier* dapat digunakan untuk memprediksi skor akhir pertandingan berdasarkan faktor-faktor seperti statistik skor, history pertandingan, atau faktor lainnya yang dapat mempengaruhi hasil pertandingan. Regresi logistik merupakan teknik untuk pemodelan probabilitas terjadinya peristiwa dari sisi kesesuaiannya tujuannya adalah untuk memahami hubungan antara satu atau lebih variabel independen (*input*) dengan variabel dependen biner tersebut.

Regresi logistik menghasilkan model logistik atau kurva logistik, yang memungkinkan kita untuk memperkirakan probabilitas hasil biner berdasarkan variabel input, Dalam konteks prediksi hasil pertandingan sepak bola, regresi logistik dapat digunakan untuk memprediksi probabilitas kemenangan atau kekalahan suatu tim berdasarkan sejumlah faktor yang mempengaruhi, seperti statistik pemain, performa tim sebelumnya, atau kondisi cuaca. Model regresi logistik ini menghasilkan probabilitas

kemenangan atau kekalahan, yang kemudian dapat digunakan untuk membuat keputusan prediksi atau untuk mengklasifikasikan hasil pertandingan. (Prastyo & Hartono, 2020).

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari PT. Liga Indonesia Baru (PT. LIB) melalui <https://www.pssi.org> (Dyar Al Falah Hilman & Aceng Komarudin Mutaqin, 2023).

Dengan demikian, hasil akhir yang didapat dalam penelitian ini dapat memberikan pengetahuan tentang besar kecilnya pengaruh suatu variabel terhadap hasil pertandingan sepak bola serta memberikan informasi tambahan bagi klub dalam merancang strategi permainan dalam menghadapi permainan lawan sehingga efektif dalam setiap pertandingan.

1.2 Rumusan Masalah

Hasil akhir dari suatu pertandingan tim sepak bola tidak hanya dilihat dari komposisi pemain yang ada pada suatu tim saja, melainkan dapat ditinjau dari beberapa aspek penting yang paling berpengaruh bagi setiap klub dalam meraih kemenangan sehingga mendapatkan hasil yang baik. Oleh karena itu, dibutuhkan suatu sistem untuk untuk menentukan hasil pertandingan tim sepak bola model *machine learning* menggunakan algoritma *regresi logistik* dan *regresi linier* untuk memprediksi hasil pertandingan sepak bola pada Liga 1 Indonesia musim 2024/2025. Fokus utama penelitian adalah untuk mengevaluasi akurasi dan kinerja kedua model tersebut, baik dalam memprediksi hasil maupun skor pertandingan, serta untuk menganalisis faktor-faktor yang berpengaruh dalam prediksi tersebut. Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang dinamika dan faktor-faktor yang memengaruhi hasil pertandingan sepak bola di tingkat liga profesional, yang dapat memberikan kontribusi dalam pengembangan strategi tim dan pengambilan keputusan yang lebih baik dalam industri sepak bola.

1.3 Batasan Masalah

Batasan dalam penelitian ini adalah :

1. Data yang digunakan dalam penelitian ini terbatas pada data pertandingan Liga 1 Indonesia pada musim 2024/2025.
2. Variabel yang dipertimbangkan meliputi data historis pertandingan, statistik skor, dan menggunakan algoritma *regresi linier* dan *regresi logistic* dalam menentukan hasil akhir pertandingan.
3. Sistem yang dibangun akan menampilkan nilai akurasi terhadap data hasil pertandingan dan menggunakan bahasa pemrograman *Python* dan IDE berupa *Visual Studio Code*.

1.4 Tujuan Penelitian

Adapun tujuan penelitian yaitu untuk mengidentifikasi faktor-faktor penentu hasil pertandingan dan mengembangkan model prediksi yang efektif dan akurat, serta memberikan rekomendasi praktis bagi pemangku kepentingan di Liga 1 Indonesia terkait penggunaan model *machine learning* dalam strategi tim dan analisis pertandingan dengan menggunakan algoritma *regresi linier* dan *regresi logistic*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini diharapkan mampu:

1. Mengetahui cara mengimplementasikan algoritma *regresi linier* dan *regresi logistic* dalam memprediksi hasil akhir pertandingan sepak bola pada pada liga I Indonesia musim 2024/2025.
2. Dapat memberikan kontribusi dalam pemahaman dan pengetahuan memprediksi hasil pertandingan sepak bola melalui implementasi *regresi linier* dan *regresi logistic*.
3. Melalui sistem yang dibuat dapat membantu pengembangan strategi pencegahan dan intervensi taktikal lawan saat akan memulai pertandingan.
4. Menjadi referensi pada penelitian berikutnya khususnya di bidang prediksi hasil pertandingan sepak bola yang menggunakan algoritma *regresi linier* dan *regresi logistic*.

1.6 Metodologi Penelitian

Metode penelitian yang dilakukan dalam penelitian ini adalah:

1. Studi pustaka

Pada tahap ini penelitian dimulai dengan mencari referensi dari berbagai sumber terpercaya dan melakukan peninjauan pustaka melalui buku-buku, jurnal, *e-book*, artikel ilmiah, makalah ataupun situs internet yang berhubungan dengan Implementasi algoritma *regresi linier* dan *regresi logistic*.

2. Analisis dan Identifikasi Masalah

Dalam metode ini, peneliti melakukan analisis menyeluruh terhadap lingkungan atau domain tertentu untuk mengidentifikasi masalah yang ada, celah pengetahuan yang perlu diisi, atau situasi yang membutuhkan pemecahan masalah. Tujuan utama dari metode ini adalah untuk menentukan tujuan penelitian yang jelas dan relevan, serta merumuskan pertanyaan penelitian yang spesifik dan terarah.

3. Perancangan dan Implementasi Sistem

Metode penelitian Perancangan dan Implementasi Sistem berfokus pada pengembangan solusi teknologi yang dapat diimplementasikan untuk memecahkan masalah yang relevan dengan topik penelitian. Hasil penelitian ini berupa sebuah program yang akan melakukan training terhadap ribuan data dan menghasilkan model yang cerdas.

4. Pengujian Sistem

Pada tahap ini, melakukan pembagian data menjadi data training dan data testing untuk digunakan dalam permodelan *machine learning*. sistem yang telah dirancang dilakukan uji coba untuk melakukan implementasi dibagi menjadi 2 Algoritma yaitu *regresi linier* dan *regresi logistic*. Setelah terbentuk model dilakukan perhitungan akurasi data testing.

5. Dokumentasi Sistem

Pada tahap ini, penelitian yang telah dilakukan, didokumentasikan

mulai dari tahap analisa sampai kepada pengujian dalam bentuk skripsi.

1.7 Penelitian Relevan

1. Penelitian dengan judul ” “,” Prediksi Penjualan Produk Unilever Menggunakan Metode *Regresi Linear*” (Ariantini & Dirgayusari, 2021). Penelitian ini bertujuan membangun sistem prediksi penjualan barang produk Unilever berbasis web menggunakan metode *regresi linear* sederhana. Pengujian akurasi terhadap hasil prediksi penjualan barang produk *Unilever* menggunakan MEA dan MAPE untuk melihat tingkat kesalahan hasil prediksi. Dataset menggunakan data penjualan produk Unilever sebanyak 15 bulan yaitu Januari 2021 sampai Maret 2022. Dataset dibagi menjadi 12 bulan sebagai data training dan 3 bulan sebagai data testing. Hasil prediksi pada 3 periode mendatang setiap jenis produk menghasilkan nilai yang sama antara hasil sistem dengan hasil perhitungan manual *regresi linear*.
2. Penelitian dengan judul “*World Cup 2022 Knockout Stage Prediction Using Poisson Distribution Model*” (Ugm, 2023) mengimplementasikan *Model Poisson Distribution* untuk melakukan model prediksi pada hasil pertandingan sepak bola. Selain menerapkan *Poisson*, penelitian ini juga untuk menunjukkan bahwa dataset yang terbatas masih bisa memecahkan masalah prediksi. Data yang digunakan untuk prediksi yaitu Babak *Knockout* Piala Dunia 2022, dimana dari uji coba yang telah dilakukan dalam percobaan model, mendapatkan hasil rata-rata kesalahan pengujian sebesar 0,075 dengan akurasi prediksi yang ditingkatkan sebesar 89,27%.
3. Pada penelitian dengan judul “PENERAPAN DATA MINING DALAM MEMPREDIKSI PEMENANG KLUB SEPAK BOLA PADA AJANG LIGA CHAMPION DENGAN ALGORITMA C.45” (Tandian et al., 2019) mengimplementasikan algoritma

algoritma C.45. Penelitian tersebut menggunakan variabel dependennya yaitu tim yang akan memenangkan liga champion, sedangkan variabel independen yang digunakan yaitu kualitas pemain, umur, kualitas pelatih, finansial, dan prestasi klub. Hasil dari pengolahan metode Data Mining dengan algoritma C.45 untuk mengidentifikasi kriteria – kriteria kemenangan klub dalam kompetisi Liga *Champion* adalah kondisi manajemen finansial klub dan umur rata – rata para pemain sepak bola dalam klub tersebut.

1.8 Sistematika Penelitian

Dalam sistematika penulisan pada tugas akhir ini dibagi dalam lima bab, masing masing bab diuraikan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini membahas tentang latar belakang dari penelitian yang dibuat dan sistem yang di kembangkan yang berkaitan dengan algoritma random *forest*, Rumusan masalah dari membahas tentang penelitian, batasan masalah dari sebuah penelitian, Tujuan Penelitian, Manfaat Penelitian, Metodologi Penelitian dan sistematika penulisan pada penelitian.

BAB 2 LANDASAN TEORI

Bab ini membahas seluruh teoritis dalam pengembangan sistem dan penelitian.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini berisikan tentang analisis dan perancangan sebuah pemangunan Sistem dan penelitian

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisikan tentang implementasi sistem yang di bangun dan penelitian.

BAB 5 PENUTUP

Bab ini menyajikan pembahasan seluruh kesimpulan dari bab-bab sebelumnya yang telah dibuat, dan juga memberikan saran terhadap apa yang telah dikerjakan agar lebih lengkap.



BAB II

LANDASANTEORI

2.1 Prediksi

Prediksi merupakan suatu proses atau tindakan untuk menentukan atau meramalkan suatu kejadian atau nilai di masa mendatang berdasarkan informasi atau data yang tersedia pada saat ini. Tujuan dari prediksi adalah memberikan estimasi tentang apa yang mungkin terjadi di masa mendatang berdasarkan informasi atau data yang ada. Prediksi seringkali melibatkan analisis data historis dan penggunaan berbagai metode statistik atau komputasional untuk mengidentifikasi pola atau tren yang dapat digunakan untuk memproyeksikan hasil di masa depan.

2.2 *Machine Learning*

Pembelajaran Mesin (ML) atau *Machine Learning* adalah pendekatan dalam dunia kecerdasan buatan yang banyak dimanfaatkan untuk meniru cara manusia berpikir dan menyelesaikan masalah atau melakukan tugas secara otomatis. Secara sederhana, *Machine Learning* berusaha meniru bagaimana kita, manusia, belajar dan membuat keputusan berdasarkan pengalaman. Prediksi dan klasifikasi adalah dua fungsi utama pembelajaran mesin. *Machine Learning* memiliki dua aplikasi utama, yakni klasifikasi dan prediksi. Keunikan dari *Machine Learning* terletak pada tahap pelatihan atau pembelajaran (training). Oleh karena itu, *Machine Learning* memerlukan data pelatihan yang disebut sebagai data training (Adnan & Amelia, 2022).

Klasifikasi mirip dengan cara manusia membedakan satu hal dari yang lain, di mana mesin menggunakan teknik untuk mengelompokkan atau mengkategorikan objek berdasarkan karakteristik tertentu. Sementara itu, prediksi atau regresi digunakan oleh mesin untuk membuat perkiraan tentang hasil dari suatu data masukan berdasarkan informasi yang telah dipelajari selama proses pelatihan. Artinya, mesin belajar dari pengalaman untuk membuat perkiraan di masa depan.

2.3 Supervised learning

Supervised learning adalah salah satu jenis pembelajaran mesin di mana algoritma "diajar" untuk membuat prediksi atau klasifikasi berdasarkan data yang telah diberi label. Dalam *supervised learning*, data yang digunakan untuk melatih model terdiri dari pasangan *input* dan *output* yang sudah diketahui. Input bisa berupa berbagai jenis data, seperti angka, teks, gambar, atau kombinasi dari semuanya. *Output* atau label adalah hasil yang kita ingin prediksi. Algoritma *supervised learning* akan mencari pola hubungan antara *input* dan *output* dalam data latih. Setelah model dilatih, kita dapat menggunakannya untuk memprediksi *output* dari data baru yang belum pernah dilihat sebelumnya.

2.4 Algoritma Regresi Linier

Regresi linier adalah Algoritma pembelajaran mesin populer yang dapat digunakan untuk mengembangkan model prediksi yang menyelidiki hubungan antara satu variabel terikat dengan satu variabel bebas (Galih et al., 2023).

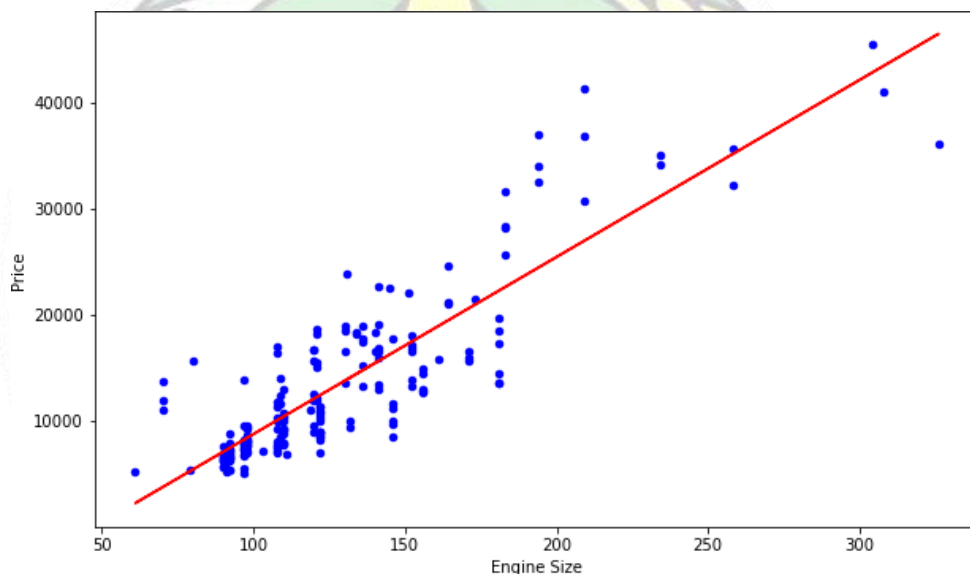
Algoritma *regresi linier* telah menjadi salah satu alat yang paling umum digunakan dalam analisis statistik dan *machine learning* untuk memprediksi atau menggambarkan hubungan antara variabel-variabel yang relevan. Dalam kasus ini, Algoritma *regresi linier* menggunakan metode (*least squares method*) metode kuadrat terkecil. Metode kuadrat terkecil adalah metode yang paling umum digunakan dalam *regresi linier*. Tujuannya adalah untuk menemukan garis lurus (atau *hipotesis linier*) yang memberikan kesalahan prediksi yang paling kecil (yaitu, mengurangi jarak antara nilai prediksi dan nilai observasi aktual).

Regresi linier cocok digunakan ketika variabel dependen dapat memiliki berbagai nilai dalam rentang kontinu, seperti memprediksi dengan hasil *output* nya hasil skor. Untuk menjalankan metode *regresi linier* dalam prediksi hasil pertandingan akan menggunakan rumus sebagai berikut;

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Keterangan:

- \hat{y} adalah nilai yang diprediksi (skor pertandingan sepak bola).
- x_1, x_2, \dots, x_n adalah nilai-nilai fitur (variabel independen) yang digunakan dalam model untuk memprediksi skor.
- $\beta_0, \beta_1, \dots, \beta_n$ adalah koefisien regresi yang akan diestimasi oleh model.
- Model mencoba untuk menemukan nilai-nilai β yang meminimalkan jumlah kuadrat kesalahan antara nilai-nilai yang diprediksi (\hat{y}) dan nilai-nilai observasi yang sebenarnya (y) dalam data pelatihan.



Gambar 2.1 Regresi Linier

2.5 Algoritma Regresi Logistic

Algoritma *regresi logistic* adalah salah satu metode statistik yang digunakan dalam *machine learning* untuk memodelkan probabilitas kejadian biner (dua kemungkinan hasil) berdasarkan satu atau lebih variabel independen. Ini adalah bentuk khusus dari regresi yang digunakan ketika variabel dependen adalah kategorikal dan biner, yaitu memiliki dua kemungkinan nilai seperti "0" atau "1", "ya" atau "tidak", "benar" atau "salah" (Cahyani et al., 2022).

Regresi logistic mengukur hubungan antara variabel dependen biner dan satu atau lebih variabel independen dengan mengestimasi probabilitas menggunakan fungsi logit. Berbeda dengan *regresi linier*, yang memberikan *output* nilai kontinu, regresi logistik memberikan *output* berupa probabilitas (nilai antara 0 dan 1) yang kemudian dapat diklasifikasikan menjadi dua kategori.

Logit, atau fungsi logit, dalam *regresi logistic* memang digunakan untuk memprediksi probabilitas hasil kejadian biner. Dalam konteks prediksi hasil pertandingan sepak bola, *regresi logistic* dapat digunakan untuk memprediksi hasil biner seperti menang atau kalah, menang atau tidak menang (menang/seri), atau hasil pertandingan (menang/kalah/seri).

Fungsi logit adalah inti dari *regresi logistic*, yang menghubungkan kombinasi linier dari variabel independen dengan probabilitas kejadian. Fungsi logit dapat dinyatakan sebagai:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Di mana :

- P adalah probabilitas tim tuan rumah menang.
- β_0 adalah intercept
- $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi.
- x_1, x_2, \dots, x_n adalah variabel independen seperti jumlah gol, tendangan sudut, posisi dalam klasemen, dll.

Probabilitas kemenangan tim tuan rumah p dihitung menggunakan fungsi sigmoid:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

Misalkan kita memiliki variabel-variabel berikut:

- x_1 = Jumlah gol tim tuan rumah dalam pertandingan sebelumnya.
- x_2 = Jumlah gol tim tamu dalam pertandingan sebelumnya.
- x_3 = Posisi tim tuan rumah dalam klasemen.
- x_4 = Posisi tim tamu dalam klasemen.

Model regresi logistic mungkin memiliki koefisien:

- $\beta_0 = -0.5$
- $\beta_1 = 0.8$
- $\beta_2 = -0.6$
- $\beta_3 = -0.3$
- $\beta_4 = 0.2$

Untuk suatu pertandingan tertentu, jika:

- $x_1 = 2$ (tim tuan rumah mencetak 2 gol di pertandingan sebelumnya)
- $x_2 = 1$ (tim tamu mencetak 1 gol di pertandingan sebelumnya)
- $x_3 = 5$ (tim tuan rumah berada di posisi ke-5 dalam klasemen)
- $x_4 = 8$ (tim tamu berada di posisi ke-8 dalam klasemen)

Maka fungsi logit dan probabilitas kemenangan tim tuan rumah dihitung sebagai:

$$\text{logit}(p) = -0.5 + 0.8 \times 2 - 0.6 \times 1 - 0.3 \times 5 + 0.2 \times 8$$

$$\text{logit}(p) = -0.5 + 1.6 - 0.6 - 1.5 + 1.6$$

$$\text{logit}(p) = 0.6$$

Menggunakan fungsi sigmoid:

$$p = \frac{1}{1 + e^{-0.6}} \approx \frac{1}{1 + 0.5488} \approx 0.645$$

Jadi, probabilitas tim tuan rumah menang dalam pertandingan tersebut adalah sekitar 64.5%

2.6 Root Mean Square Error (RMSE)

Root Mean Squared Error (RMSE) adalah metrik yang mengukur tingkat kesalahan dalam hasil prediksi, nilai RMSE yang lebih rendah menunjukkan bahwa prediksi model tersebut lebih akurat. Sebagai metode evaluasi, RMSE memperhitungkan perbedaan antara nilai yang diprediksi oleh model dengan nilai sebenarnya yang diobservasi. Dengan demikian, RMSE berfungsi sebagai alat penilaian yang membantu mengukur sejauh mana model mampu menghasilkan estimasi yang mendekati nilai yang sebenarnya, dan nilai RMSE yang rendah mengindikasikan performa prediktif yang lebih baik. Ini memiliki implikasi praktis yang signifikan, karena model dengan RMSE yang kecil cenderung memberikan prediksi yang lebih dapat diandalkan dalam berbagai aplikasi.

Keuntungan dari *Root Mean Squared Error* (RMSE) adalah dapat memberikan nilai kesalahan dalam satuan yang sama dengan data asli, sehingga lebih mudah diinterpretasikan dan rentan terhadap *outlier* lebih sedikit dibandingkan dengan MSE karena pengaruh kuadrat pada *outlier* dikurangi oleh pengakaran. Proses penghitungan *Root Mean Squared Error* (RMSE) melibatkan tahap awal pengurangan nilai aktual dari nilai peramalan, diikuti dengan langkah pengkuadratan dan penjumlahan total seluruh hasil kuadrat. Total tersebut kemudian dibagi oleh jumlah data yang ada. Selanjutnya, hasil perhitungan ini kembali diolah dengan menghitung nilai akar kuadrat sesuai dengan formula yang diberikan. Dengan demikian, RMSE memberikan gambaran mengenai tingkat kesalahan atau deviasi antara prediksi model dengan data aktual dalam sebuah *Dataset*. Berikut rumus menghitung RMSE :

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i)^2}$$

Persamaan 1

Keterangan :

\tilde{y}_i : Nilai hasil peramalan

y_i : Nilai aktual / Nilai sebenarnya :

n : Jumlah data

2.7 Faktor - faktor hasil pertandingan

faktor-faktor yang mempengaruhi hasil pertandingan Liga 1 Indonesia dalam beberapa poin:

Faktor dalam Pertandingan:

- Kualitas Pemain: Keahlian individu sangat menentukan.
- Taktik: Strategi yang tepat bisa memenangkan pertandingan.
- Kondisi Fisik: Kebugaran pemain sangat penting.
- Mentalitas: Semangat juang dan kepercayaan diri mempengaruhi hasil.
- Faktor Kebetulan: Keberuntungan juga berperan.

Faktor di Luar Pertandingan:

- Kondisi Lapangan: Lapangan yang baik mendukung permainan.
- Cuaca: Cuaca ekstrem bisa mempengaruhi hasil.
- Tekanan Suporter: Dukungan dan tekanan dari suporter berpengaruh.
- Kondisi Keuangan Klub: Klub yang kaya cenderung lebih kuat.
- Keputusan Wasit: Keputusan wasit yang kontroversial bisa mengubah hasil.
- Jadwal Pertandingan: Jadwal yang padat bisa membuat pemain lelah.

Faktor Khusus Liga 1 Indonesia:

- Pergantian Pelatih: Pergantian pelatih bisa mengganggu stabilitas tim.
- Permasalahan Non-Teknis: Masalah di luar lapangan bisa mengganggu fokus pemain.
- Pengaruh Sponsor: Sponsor bisa memberikan tekanan kepada klub.

BAB III

ANALISIS DAN PERANCANGAN

Bab analisis dan perancangan system dengan algoritma *regresi linier* dan *regresi logistic* membahas Langkah-langkah yang terlibat dalam penerapan suatu prediksi. Pada bab ini, dilakukan Analisis dan perancangan sistem yang berfokus pada pemilihan fitur yang relevan, *preprocessing* data, dan pemilihan model yang tepat .

3.1 Arsitektur Umum

Tahap ini diajukan dalam system prediksi hasil pertandingan sepak bola menggunakan algortima *regresi linier* dan *regresi logistic* dari beberapa tahapan pada gambar 3.1.



Gambar 3.1 Arsitektur Umum

Arsitektur umum pada gambar 3.1 dapat dilihat langkah awalnya yaitu melakukan pengumpulan data melalui beberapa website resmi liga 1 indonesia, lalu jika data yang sudah dikumpulkan langsung di *preprocessing* yaitu Dibersihkan data dibersihkan data dari nilai-nilai yang hilang atau tidak valid Jika diperlukan data akan dinormalisasi. Setelah dibersihkan data akan dibagi menjadi 2 data *training* dan data *testing*. Lalu langkah selanjutnya melakukan model dengan algoritma *regersi linier* dan *regresi logistic*. Lalu setelah itu uji model menggunakan data pengujian yang tidak dilihat sebelumnya dan hitung akurasi. Lalu validasi model dan evaluasi Kemudian model siap digunakan jika sudah melewati tahap evaluasi. Setelah itu model yang telah evaluasi diimplementasikan ke dalam sistem lalu user memberikan inputan jawaban ke sistem lalu sistem akan memprediksi sesuai kerja algoritma *regresi linier* dan *regresi logistic* untuk menghasilkan prediksi yang akurat. Sistem ini dapat membantu memprediksi hasil akhir pertandingan sepak bola.

3.2 Pengumpulan Dataset

Data pertandingan yang digunakan dalam penelitian ini bersumber secara eksklusif dari situs resmi Persatuan Sepak Bola Seluruh Indonesia (PSSI). Dataset ini mencakup statistik pertandingan Liga 1 Indonesia musim 2024/2025, meliputi data individu pemain seperti jumlah gol, *assist*, dan kartu, serta statistik tim seperti penguasaan bola, jumlah tembakan, dan hasil pertandingan.

Dengan mengakses langsung sumber primer ini, diharapkan data yang diperoleh memiliki tingkat akurasi dan relevansi yang tinggi untuk mendukung analisis dan pemodelan dalam penelitian ini.

	tanggal	tim_kandang	tim_tandang	skor_kandang	skor_tandang	hasil	Team	GP	W	D	L	P	Team_away	GP_away	W_away	D_away	L_away	P_away
0	2024-08-05	Perseb Bandung	PSBS Bisk	4	1	home_win	Perseb Bandung	8	4	4	—	16	PSBS Bisk	9	4	0	5	
1	2024-08-10	Madura United	Malut United	1	1	draw	Madura United	9	1	3	—	6	Malut United	9	2	4	3	
2	2024-08-10	PSM Makassar	Perse Solo	3	0	home_win	PSM Makassar	8	4	4	—	16	Perse Solo	8	2	1	5	
3	2024-08-10	Perseja	Santo	3	0	home_win	Perseja	8	3	3	—	12	Bento	9	2	3	4	
4	2024-08-11	PSIS	Perse	0	1	away_win	PSIS	9	2	1	—	7	Perse	8	3	2	3	
67	2024-10-12	PSM Makassar	Madura United	2	0	home_win	PSM Makassar	9	4	4	—	16	Madura United	9	1	3	5	
68	2024-10-12	Arena	Malut United	3	1	home_win	Arena	8	3	3	—	12	Malut United	9	2	4	3	

Gambar 3.2 Dataset liga 1 indonesia

3.3 Pre-processing Dataset

Pre-processing adalah tahap awal dalam pengolahan data yang melibatkan serangkaian teknik untuk membersihkan dan mempersiapkan data mentah sebelum diproses lebih lanjut, seperti penghapusan data duplikat, penghapusan nilai yang hilang, dan transformasi data (Ahmed dan Kamal, 2020). Dengan melakukan *pre-processing* data dengan baik, kita dapat mengurangi risiko *overfitting* dan *underfitting* dalam model, meningkatkan kinerja model, dan menghasilkan hasil analisis yang lebih akurat. Selain itu, *pre-processing* juga membantu menghemat waktu dan upaya dalam tahap analisis selanjutnya, dengan menyediakan *dataset* yang lebih bersih dan mudah untuk dimengerti. Dengan demikian, *pre-processing* menjadi tahap penting dalam siklus analisis data yang memastikan bahwa data yang digunakan untuk pengambilan keputusan atau pemodelan memiliki kualitas yang baik dan dapat diandalkan. Tahapan-tahapan yang terdapat pada *pre-processing* pada penelitian ini adalah konversi kolom ke tipe numerik dan menghapus nilai kosong.

```
#mengubah kolom kategori seperti nama tim dan hasil pertandingan menjadi numerik
encoder_kandang = LabelEncoder()
encoder_tandang = LabelEncoder()
encoder_hasil = LabelEncoder()

df['home'] = encoder_kandang.fit_transform(df['home'])
df['away'] = encoder_tandang.fit_transform(df['away'])
df['result'] = encoder_hasil.fit_transform(df['result'])
```

Gambar 3.3 Proses Perubahan Menjadi Numerik

3.4 Pembagian Data

Pembagian data adalah proses membagi dataset menjadi beberapa bagian, biasanya untuk melatih dan mengevaluasi model *machine learning*. Data Latih digunakan untuk "mengajar" model mengenali pola dan hubungan dalam data, sementara data Uji digunakan untuk mengukur seberapa baik model tersebut dapat memprediksi data baru yang belum pernah dilihat sebelumnya. Tujuan utama dari pembagian data adalah untuk mencegah *overfitting*, yaitu kondisi di mana model terlalu menghafal data pelatihan sehingga tidak dapat generalisasi dengan baik pada data baru, serta untuk mendapatkan evaluasi kinerja model yang lebih objektif.

Data Train (Data Latih) :

- Fungsi: Digunakan untuk melatih model. Model akan belajar pola dan hubungan antara fitur (*input*) dan target (*output*) dari data train ini.
- Proses: Algoritma *machine learning* akan menganalisis data train secara berulang untuk menemukan pola terbaik yang dapat memetakan fitur ke target.

Data Test (Data Uji):

- Fungsi: Digunakan untuk mengevaluasi performa model yang sudah dilatih. Setelah model dilatih dengan data train, model akan diuji pada data tes yang belum pernah dilihat sebelumnya.
- Proses: Model akan membuat prediksi pada data tes, dan hasil prediksi ini akan dibandingkan dengan nilai target yang sebenarnya. Perbandingan ini digunakan untuk menghitung metrik evaluasi seperti akurasi, *precision*, *recall*, dan sebagainya.

```
#prediksi hasil
X_train_result, X_test_result, y_train_result, y_test_result = train_test_split(X, y_result, test_size=0.1, random_state=32)

#prediksi skor kandang
X_train_home, X_test_home, y_train_home, y_test_home = train_test_split(X, y_score_home, test_size=0.1, random_state=32)

#prediksi skor tandang
X_train_away, X_test_away, y_train_away, y_test_away = train_test_split(X, y_score_away, test_size=0.1, random_state=32)
```

Gambar 3.4 Pembagian Data *Train* dan *Test*

3.5 Model Regresi Linier

Algoritma *regresi logistic* adalah salah satu metode statistik yang digunakan dalam *machine learning* untuk memodelkan probabilitas kejadian biner (dua kemungkinan hasil) berdasarkan satu atau lebih variabel independen. Ini adalah bentuk khusus dari regresi yang digunakan ketika variabel dependen adalah kategorikal dan biner, yaitu memiliki dua kemungkinan nilai seperti "0" atau "1", "ya" atau "tidak", "benar" atau "salah" (Cahyani et al., 2022). Dalam *Python*, kita dapat dengan mudah mengimplementasikan *regresi linier* menggunakan library *Scikit-learn*.

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# linear regresi untuk prediksi skor kandang
lireg_home = LinearRegression()
lireg_home.fit(X_train_home, y_train_home)

# linear regresi untuk prediksi skor tandang
lireg_away = LinearRegression()
lireg_away.fit(X_train_away, y_train_away)
```

Gambar 3.5 Data Latih *Regresi Linier*

3.5.1 Hitung selisih

Menghitung selisih poin dengan cara mengurangi nilai di kolom 'p_away' (poin tim tamu) dari nilai di kolom 'p_home' (poin tim tuan rumah). Hasil perhitungan ini kemudian disimpan di kolom 'pd'.

```
new_data['pd'] = new_data['p_home'] - new_data['p_away']
```

Gambar 3.5.1 Selisih Home & away

3.5.2 MSE (*Mean Squared Error*)

Metrik yang umum digunakan dalam evaluasi model regresi atau prediksi. MSE mengukur rata-rata kuadrat dari selisih antara nilai prediksi dan nilai aktual. Nilai MSE yang lebih kecil menunjukkan bahwa model prediksi lebih akurat.

```
print('MSE skor kandang:', mse_home)
print('MSE skor tandang:', mse_away)

MSE skor kandang: 1.5619578157373328
MSE skor tandang: 1.1191976651210807
```

Gambar 3.5.2 Metrik MSE

3.6 Model *Regresi Logistic*

Regresi logistic adalah teknik statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan satu variabel dependen yang bersifat kategorik, biasanya biner (ya/tidak, 0/1). Berbeda dengan *regresi linier* yang memprediksi nilai numerik, regresi logistik memprediksi *probabilitas* terjadinya suatu peristiwa.

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

log_reg = LogisticRegression()
log_reg.fit(X_train_result, y_train_result)
```

Gambar 3.6 Data Latih *Regresi Logistic*

3.7 Evaluasi

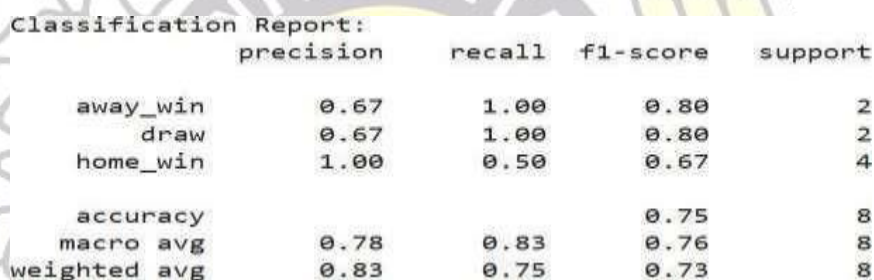
Tujuan tahap evaluasi dalam *Regresi linier* dan *Regresi logistic* adalah untuk mengukur kinerja dan keefektifan model dalam melakukan analisis atau tugas lainnya. Evaluasi dilakukan untuk memperoleh pemahaman yang lebih baik tentang seberapa baik model *Regresi linier* dan *Regresi logistic* dalam memprediksi skor. Beberapa tujuan utama dari tahap evaluasi dalam *Regresi linier* dan *Regresi logistic* adalah:

Mengukur akurasi : Tujuan tahap evaluasi dalam regresi linier dan regresi logistic adalah untuk menunjukkan kinerja yang memuaskan dalam memprediksi hasil pertandingan sepak bola Liga 1 Indonesia 2024/2025. Dengan menggunakan metrik akurasi, *precision* dan *recall*, model berhasil mencapai tingkat akurasi sebesar 75%. Hal ini mengindikasikan bahwa model mampu menggeneralisasi dengan baik pada data baru dan dapat menjadi alat bantu yang efektif dalam memprediksi hasil pertandingan."

Menerkankan Keterbatasan Model : Meskipun model telah menunjukkan kinerja yang cukup baik, perlu diingat bahwa model ini memiliki beberapa keterbatasan. Model hanya mempertimbangkan faktor-faktor yang tersedia dalam *dataset*, sehingga faktor-faktor eksternal seperti kondisi cuaca, cedera pemain, dan motivasi tim tidak dapat dipertimbangkan secara langsung. Selain itu, kinerja model dapat dipengaruhi oleh perubahan dinamika kompetisi.

3.7.1 Klasifikasi

Laporan klasifikasi ini memberikan evaluasi terhadap kinerja suatu model klasifikasi. Dalam konteks ini, model tersebut kemungkinan digunakan untuk memprediksi hasil pertandingan (menang, kalah, atau seri) antara dua tim.

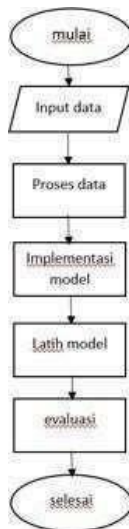


Classification Report:				
	precision	recall	f1-score	support
away_win	0.67	1.00	0.80	2
draw	0.67	1.00	0.80	2
home_win	1.00	0.50	0.67	4
accuracy			0.75	8
macro avg	0.78	0.83	0.76	8
weighted avg	0.83	0.75	0.73	8

Gambar 3.7.1 Hasil Klasifikasi

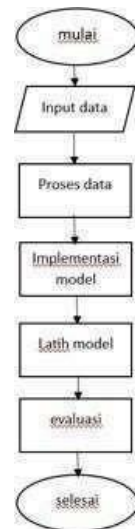
3.8 Flowchart

Perancangan sistem dalam penelitian ini akan direpresentasikan dalam bentuk diagram alir atau *flowchart*. *Flowchart* ini bertujuan untuk memberikan gambaran visual mengenai alur kerja atau proses yang sedang diidentifikasi atau dirancang seperti gambar 3.2 dibawah ini :



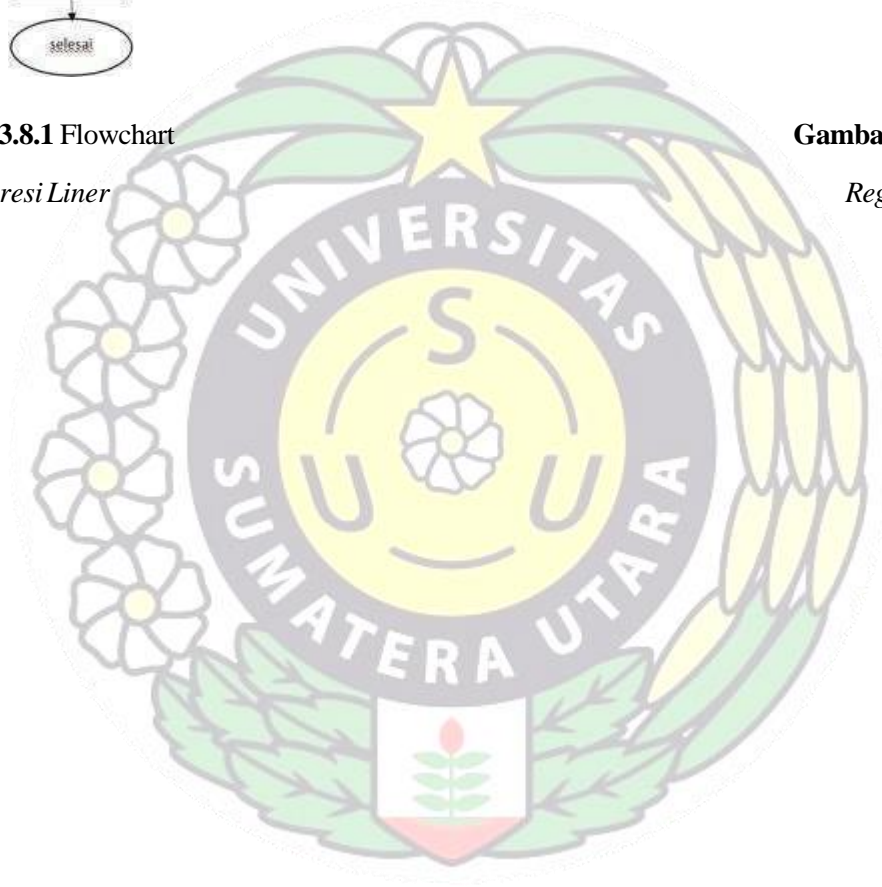
Gambar 3.8.1 Flowchart

Regresi Linier



Gambar 3.8.2 Flowchart

Regresi Logistic



BAB IV

IMPLEMENTASI DAN PEMBAHASAN

Bab ini dalam konteks implementasi *Regresi Linier* dan *Regresi Logistic* dalam memprediksi *hasil pertandingan sepak bola* membahas langkah- langkah yang dilakukan untuk menerapkan model Regresi Linier dan *Regresi Logistic* dalam memprediksi hasil pertandingan sepak bola.

4.1 Implementasi Sistem

4.1.1 Spesifikasi Perangkat Keras

Spesifikasi perangkat lunak adalah dokumen yang berisi rincian tentang kebutuhan dan fungsionalitas yang diharapkan dari sebuah perangkat lunak. Spesifikasi ini menjadi landasan bagi pengembang perangkat lunak untuk merancang, mengembangkan, dan menguji produk yang memenuhi kebutuhan pengguna dengan baik. Untuk dapat menerapkan dan menguji implementasi GRU dalam memprediksi sepak bola, diperlukan perangkat keras dengan spesifikasi sebagai berikut.

Processor: Intel Core i5-1135G7 (Generasi ke-11) atau setara

RAM: 8 GB DDR4

Storage: SSD 256 GB

Sistem Operasi: Windows 10/11 64-bit atau Ubuntu 22.04 LTS

Untuk mendukung persyaratan perangkat keras yang diperlukan, penulis juga memanfaatkan layanan *cloud* gratis dari *google* yang disebut *Google Colab*. Dalam *Google Colab*.

4.1.2 Spesifikasi Perangkat Keras

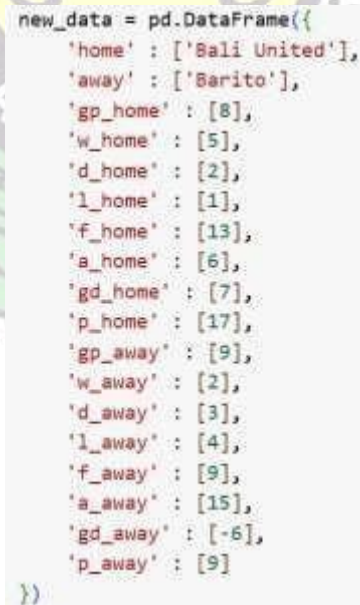
Spesifikasi perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut:

- 3.8.1 Sistem Operasi Windows 11 Home 64 bit *operating system*.
- 3.8.2 *Google Colab*
- 3.8.3 *Python3*
- 3.8.4 Library: Pandas, seaborn, matplotlib, sklearn..

4.2 Pengujian dan analisis sistem

4.2.1 Implementasi Regresi Linier

Pada tahap Implementasi dalam algoritma *Regressi Linier* adalah untuk memodelkan hubungan linear antara satu atau lebih variabel independen (fitur) dengan satu variabel dependen (target). Prinsip kerjanya adalah mencari garis lurus terbaik yang dapat menggambarkan hubungan antara data. Dalam Implementasinya, data dibagi menjadi data latih dan data uji. Model *regresi linier* kemudian dilatih pada data latih untuk menemukan parameter yang optimal.



```
new_data = pd.DataFrame({
    'home' : ['Bali United'],
    'away' : ['Barito'],
    'gp_home' : [8],
    'w_home' : [5],
    'd_home' : [2],
    'l_home' : [1],
    'f_home' : [13],
    'a_home' : [6],
    'gd_home' : [7],
    'p_home' : [17],
    'gp_away' : [9],
    'w_away' : [2],
    'd_away' : [3],
    'l_away' : [4],
    'f_away' : [9],
    'a_away' : [15],
    'gd_away' : [-6],
    'p_away' : [9]
})
```

Gambar 4.2.1 Input dataset *regresi linier*

Dengan menginput beberapa angka yang ada di *dataset* program akan menampilkan *output* skor kedua tim.

```
print(f'Predicted Home Score: {predicted_home_score[0]:.2f}')
# print(f'Predicted Away Score: {predicted_away_score[0]:.2f}')

print(f'Predicted Home Score: {max(0, round(predicted_home_score[0]))}')
print(f'Predicted Away Score: {max(0, round(predicted_away_score[0]))}')

Predicted Home Score: 1.61
Predicted Home Score: 2
Predicted Away Score: 1
```

Gambar 4.2.1 Output *Regresi linier*

4.2.2 Implementasi Regresi Logistic

Pada tahap Implementasi dalam algoritma *Regresi Logistic* adalah untuk memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan satu variabel dependen yang bersifat kategorik (biasanya biner, ya/tidak). Model ini memprediksi probabilitas terjadinya suatu peristiwa. Prosesnya melibatkan pembagian data menjadi data latih dan data uji, pembuatan model menggunakan algoritma regresi logistik, pelatihan model pada data latih, dan evaluasi model pada data uji menggunakan metrik seperti akurasi. Dengan memasukkan inputan dataset yang sama, program akan menampilkan hasil kemenangan salah satu tim.

 Hasil Prediksi: home_win

Gambar 4.2.2 Hasil Output *Regresi Logistic*

4.3 Analisis keseluruhan Sistem

Tahap keseluruhan menjelaskan proses evaluasi yang bertujuan mengukur seberapa baik suatu sistem dalam menghasilkan *output* yang benar atau sesuai dengan nilai sebenarnya.

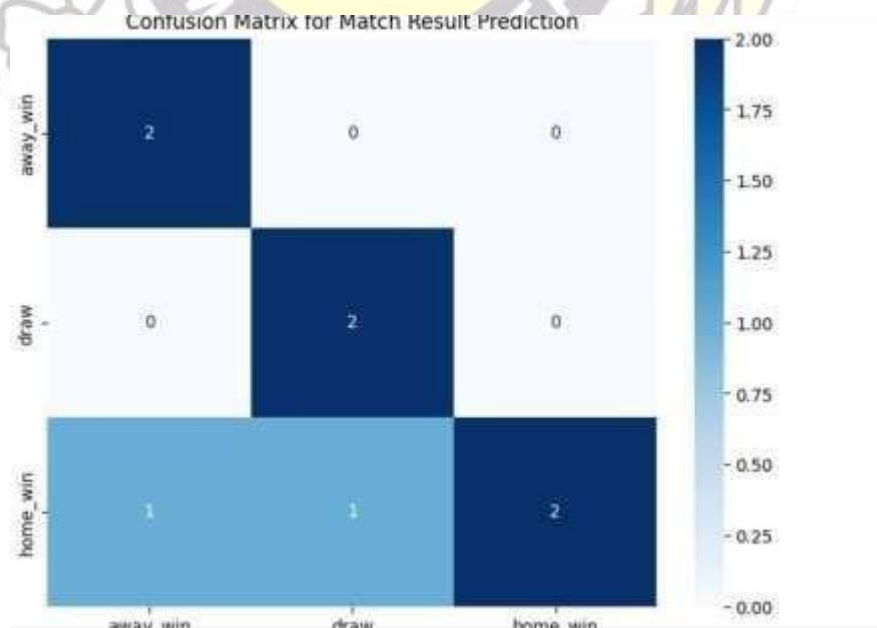
```
[ ] accuracy = accuracy_score(y_test_result, y_pred_result)
    print(f'Hasil Akurasi: {accuracy}')
```

⇒ Hasil Akurasi: 0.75

Gambar 4.3 Tingkat Akurasi

4.4 Confusion matrix

Confusion matrix adalah sebuah tabel yang digunakan dalam *machine learning* untuk mengevaluasi performa model klasifikasi. Tabel ini membandingkan antara hasil prediksi model dengan nilai aktual dari data. Dengan kata lain, *confusion matrix* menunjukkan seberapa baik model kita dalam mengklasifikasikan data.



Gambar 4.4 Matrix prediksi

BAB V

KESIMPULAN DAN SARAN

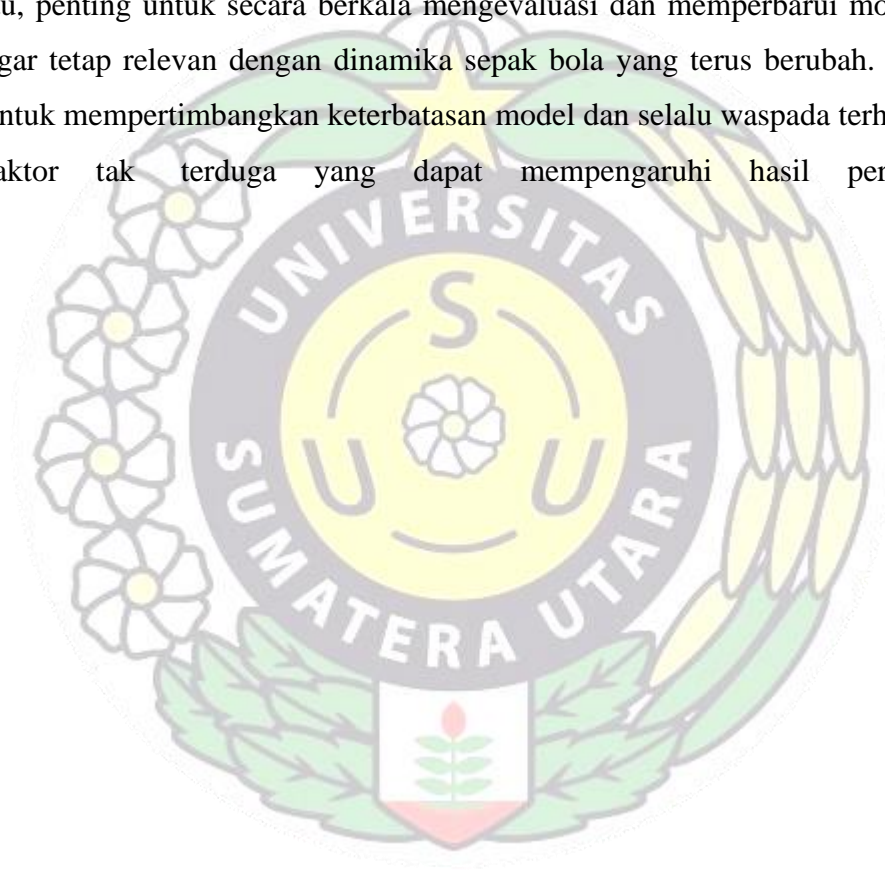
5.1 Kesimpulan

Dari hasil penelitian melalui pengujian dan analisis system yang telah dilakukan, maka dapat disimpulkan bahwa:

1. Sistem yang dibuat dapat memprediksi hasil menang, kalah atau seri dari sebuah pertandingan sepakbola dengan rata-rata Tingkat ketepatan 75%. Artinya, dari semua prediksi yang dilakukan model, 75% diantaranya benar. Tetapi, masih ada ruang untuk perbaikan.
2. Rata-rata dari hasil pertandingan adalah 2,12155664. Dimana pada tiap-tiap pertandingan, rata-rata selisih skor yang diprediksi sebesar 2 gol untuk total gol keseluruhan.
3. Koefisien regresi memang terbukti menjadi faktor menentukan dalam peningkatan atau penurunan nilai variabel bebas
4. Nilai Akurasi selalu berada dalam rentang 0 hingga 1. Yang mana angka 1, model memprediksi semua hasil pertandingan dengan benar. Sedangkan, angka 0 menunjukkan hasil pertandingan yang salah.
5. Metode *Regresi linier* dan *regresi logistic* dinilai tepat apabila digunakan untuk memprediksi sebuah pertandingan sepakbola karena rata-rata Tingkat keberhasilan mencapai 75%. Akan tetapi akan meningkat menjadi 80% apabila menambahkan dengan variabel stamina pemain dan faktor cuaca.
6. Dengan mengkolaborasi antara algoritma *regresi linier* dan *regresi logistic* akan menambah akurasi lebih tinggi dan membuat prediksi semakin akurat dan kompleks.
7. Dalam memprediksi sebuah pertandingan sepakbola, banyak hal lain yang dapat menjadi pembeda sebuah pertandingan tersebut.

5.2 Saran

Untuk meningkatkan akurasi prediksi pertandingan sepak bola, disarankan untuk menggabungkan berbagai metode analisis data, seperti statistik, *machine learning*, dan analisis sentimen. Selain itu, pertimbangkan untuk memasukkan faktor-faktor non-statistik seperti kondisi cuaca, jadwal pertandingan yang padat, dan motivasi tim. Dengan memperluas cakupan data dan menggunakan model yang lebih kompleks, kita dapat memperoleh prediksi yang lebih akurat. Selain itu, penting untuk secara berkala mengevaluasi dan memperbarui model prediksi agar tetap relevan dengan dinamika sepak bola yang terus berubah. Jangan lupa untuk mempertimbangkan keterbatasan model dan selalu waspada terhadap faktor-faktor tak terduga yang dapat mempengaruhi hasil pertandingan.



DAFTAR PUSTAKA

- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, 1142(1). <https://doi.org/10.1088/1742-6596/1142/1/012012>
- Anggrawan, A., Hairani, H., & Azmi, N. (2022). Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear. *Jurnal Bumigora Information Technology (BITe)*, 4(2), 123–132. <https://doi.org/10.30812/bite.v4i2.2416>
- Cahyani, Q. R., Finandi, M. J., Rianti, J., Arianti, D. L., Dwi, A., Putra, P., & Artikel, G. (2022). Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 1(2), 2828–9099. <https://doi.org/10.55123/jomlai.v1i2.598>
- Dyar Al Falah Hilman, & Aceng Komarudin Mutaqin. (2023). Penerapan Regresi Double Poisson untuk Memprediksi Pertandingan dan Klasemen Liga 1 Indonesia. *Jurnal Riset Statistika*, 97–106. <https://doi.org/10.29313/jrs.v3i2.2784>
- Galih, M., Atika, P. D., & Mukhlis. (2023). Prediksi Penjualan Menggunakan Algoritma Regresi Linear Di Koperasi Karyawan “Usaha Bersama.” *Journal of Informatic and Information Security*, 3(2), 193–202. <https://doi.org/10.31599/jiforty.v3i2.1354>
- Guo, X. (2021). *applied sciences Using a Random Forest Model to Predict the Location of Potential Damage on Asphalt Pavement*.
- Karim, A. A., Prasetyo, M. A., & Saputro, M. R. (2023). Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia. *Stains (Seminar Nasional Teknologi & Sains)*, 2(1), 377–382.
- Mutaqin, A. K., & Adri, Y. P. (2019). Prediksi Klasemen Akhir Kompetisi Sepakbola Indonesia Menggunakan Metode Perluasan Ekspektasi Pythagoras. *Journal of Data Analysis*, 2(1), 22–30. <https://doi.org/10.24815/jda.v2i1.14123>
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021).

- Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17. <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- Prastyo, I. S., & Hartono. (2020). Jurnal phenomenon. *Phenomenon*, 10(1), 25–35. Tandian, C., Laia, Y., & Saputra, A. (2019). Penerapan Data Mining Dalam Memprediksi Pemenang Klub Sepak Bola Pada Ajang Liga Champion Dengan Algoritma C.45. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 91–98. <https://doi.org/10.34012/jusikom.v2i2.397>



LAMPIRAN

Lampiran 1. Listing Program

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.model_selection import train_test_split

#Impor dari google drive
from google.colab import drive
# Mount Google Drive
drive.mount('/content/drive')

#load data
df = pd.read_csv('/content/drive/MyDrive/data/merged_df.csv')

#Data Cleaning
columns = [column.lower() for column in df.columns]
df.columns = columns

df.drop(columns=['team_home', 'team_away', 'date'], inplace=True)

# Penambahan fitur/ kolom pada data
df['pd'] = df['p_home'] - df['p_away']

#mengubah kolom kategori seperti nama tim dan hasil pertandingan menjadi numerik
encoder_kandang = LabelEncoder()
encoder_tandang = LabelEncoder()
encoder_hasil = LabelEncoder()

df['home'] = encoder_kandang.fit_transform(df['home'])
df['away'] = encoder_tandang.fit_transform(df['away'])
df['result'] = encoder_hasil.fit_transform(df['result'])

#pemilihan target untuk prediksi hasil
y_result = df['result']

#pemilihan target untuk prediksi skor_kandang
y_score_home = df['score_home']
#pemilihan target untuk prediksi skor_tandang
y_score_away = df['score_away']

#prediksi hasil
X_train_result, X_test_result, y_train_result, y_test_result = train_test_split(X, y_result,
test_size=0.1, random_state=32)
```



```

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# linear regresi untuk prediksi skor kandang
lireg_home = LinearRegression()
lireg_home.fit(X_train_home, y_train_home)

# linear regresi untuk prediksi skor tandang
lireg_away = LinearRegression()
lireg_away.fit(X_train_away, y_train_away)

#prediksi dengan data pengujian
y_pred_home = lireg_home.predict(X_test_home)
y_pred_away = lireg_away.predict(X_test_away)

# Mean Squared Error
print('MSE skor kandang:', mse_home)
print('MSE skor tandang:', mse_away)

#points diff (selisih poin)
new_data['pd'] = new_data['p_home'] - new_data['p_away']

#prediksi output regresi Linier
print(f'Predicted Home Score: {predicted_home_score[0]:.2f}')
# print(f'Predicted Away Score: {predicted_away_score[0]:.2f}')

print(f'Predicted Home Score: {max(0, round(predicted_home_score[0]))}')
print(f'Predicted Away Score: {max(0, round(predicted_away_score[0]))}')

#prediksi regresi Logistic
Prediksi hasil pertandingan
predicted_result = log_reg.predict(X_new)

# Simpan encoder untuk 'result'
with open('encoder_hasil.pkl', 'wb') as file:
    pickle.dump(encoder_hasil, file)

```

Lampiran 2 . Biodata

 081396058382

 ryanvidi11@gmail.com

 M05, DESASISWA TEKUN, KOMPLEK RST. Medan, 20144, Indonesia

Haryanda Fidi

ADMIN PARFUME

TENTANG SAYA

Koordinator Administrasi Parfume Berpengalaman dengan 2 Tahun di Forst.A: Pemain Sepak Bola dan Pebisnis dengan Keahlian Utama dalam MS Office dan Desain, Lulusan SMA 2020. Bersemangat untuk Bergabung dengan Tim Dinamis dalam Industri Parfum dan Memanfaatkan Keterampilan untuk Meningkatkan Efisiensi Operasional secara Kreatif dan Bertanggung Jawab.

PENGALAMAN KERJA

Admin parfume

Forst / Medan / Apr 2023 - Sekarang

Admin Ponsel

Ponsel / Medan / Nov 2022 - Apr 2023

PENDIDIKAN

Ijazah SMA

SMAN 2 Medan / Medan / 2020

SMPN 2 Medan / Medan / 2017

TAUTAN

Instagram :

HaryandaFidi

KEAHLIAN

Pemain Sepak Bola

Pebisnis

MS, Office

Mendesain

