



APAKAH REGRESI POLINOMIAL BERGANDA KUADRATIK TEPAT UNTUK ANALISIS PERFORMA TIM SEPAK BOLA? STUDI KASUS LIGA INGGRIS 2022-2023

LAPORAN TUGAS AKHIR

Oleh:

Rifqi Julian Hasyari

105218037



**FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
UNIVERSITAS PERTAMINA
FEBRUARI 2023**

1. Dilarang mengutip karya tulis ini, kecuali:
 - a. menyebutkan sumber sesuai kaidah kecendekiaan;
 - b. pengutipan hanya untuk keperluan pendidikan, penulisan karya ilmiah atau penelitian;
 - c. pengutipan tidak merugikan Universitas Pertamina.
2. Dilarang mempublikasikan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa adanya izin dari Universitas Pertamina.



APAKAH REGRESI POLINOMIAL BERGANDA KUADRATIK TEPAT UNTUK ANALISIS PERFORMA TIM SEPAK BOLA? STUDI KASUS LIGA INGGRIS 2022-2023

LAPORAN TUGAS AKHIR

Oleh:

Rifqi Julian Hasyari

105218037



**FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
UNIVERSITAS PERTAMINA
FEBRUARI 2023**

1. Dilarang mengutip karya tulis ini, kecuali:
 - a. menyebutkan sumber sesuai kaidah kecendekiaan;
 - b. pengutipan hanya untuk keperluan pendidikan, penulisan karya ilmiah atau penelitian;
 - c. pengutipan tidak merugikan Universitas Pertamina.
2. Dilarang mempublikasikan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa adanya izin dari Universitas Pertamina.





LEMBAR PENGESAHAN

Judul Tugas Akhir

: APAKAH REGRESI POLINOMIAL BERGANDA
KUADRATIK TEPAT UNTUK ANALISIS PER-
FORMA TIM SEPAK BOLA? STUDI KASUS
LIGA INGGRIS 2022-2023

Nama Mahasiswa

: Rifqi Julian Hasyari

Nomor Induk Mahasiswa

: 105218037

Program Studi

: Ilmu Komputer

Fakultas

: Sains dan Ilmu Komputer


Tanggal Lulus Sidang Tugas Akhir

: 15 Februari 2023

Jakarta, 13 Maret 2023


MENGESAHKAN

Pembimbing I



Rangga Ganzar Noegraha, Ph.D
NIP.119005

Pembimbing II



Ade Ifawan, Ph.D
NIP.116130

MENGETAHUI,

Ketua Program Studi



Muhamad Koyimatu Ph.D

NIP. 116108

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa Tugas Akhir berjudul **APAKAH REGRESI POLINOMIAL BERGANDA KUADRATIK TEPAT UNTUK ANALISIS PERFORMA TIM SEPAK BOLA? STUDI KASUS LIGA INGGRIS 2022-2023** ini adalah benar-benar merupakan hasil karya saya sendiri dan tidak mengandung materi yang ditulis oleh orang lain kecuali telah dikutip sebagai referensi yang sumbernya telah dituliskan secara jelas sesuai dengan kaidah penulisan karya ilmiah.

Apabila dikemudian hari ditemukan adanya kecurangan dalam karya ini, saya bersedia menerima sanksi dari Universitas Pertamina sesuai dengan peraturan yang berlaku.

Demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Pertamina hak bebas royalti noneksklusif (*non-exclusive royalty-free right*) atas Tugas Akhir ini beserta perangkat yang ada. Dengan hak bebas royalti noneksklusif ini Universitas Pertamina berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya

Jakarta, 10 Maret 2023

Yang membuat pernyataan,



Rifqi Julian Hasyari



ABSTRAK

Rifqi Julian Hasyari. 105218037. APAKAH REGRESI POLINOMIAL BERGANDA KUADRATIK TEPAT UNTUK ANALISIS PERFORMA TIM SEPAK BOLA? STUDI KASUS LIGA INGGRIIS 2022-2023.

Sepak bola kini tidak hanya menjadi sebuah hobi namun sudah berkembang menjadi sebuah industri yang besar. Liga Inggris atau *English Premier League* (EPL) merupakan salah satu liga sepak bola yang memiliki pasar terbesar di antara liga-liga sepak bola lainnya. Dikarenakan sepak bola telah berkembang menjadi industri, setiap tim berusaha untuk meningkatkan performa mereka. Salah satu caranya adalah dengan memanfaatkan data statistik yang mereka punya menggunakan *Machine Learning*. Tujuan dari penelitian ini adalah untuk mendapatkan fitur yang sesuai serta mendapatkan hasil performa model yang dibangun menggunakan regresi polinomial berganda kuadrat. Penelitian ini menggunakan data statistik yang diambil dari situs fbref.com dengan cara *web scrapping*. Untuk memenuhi keperluan pembuatan model regresi polinomial berganda kuadrat, data diproses dan dilakukan perkalian fitur. Selain itu, penelitian ini menggunakan korelasi *Pearson* untuk melihat nilai korelasi dari setiap fitur. Performa model yang dibangun diuji dengan menggunakan metode R-square dan *mean absolute error* (MAE). Penelitian ini menghasilkan fitur-fitur yang cocok untuk melakukan prediksi terjadinya goal, yaitu fitur banyaknya operan berhasil yang dilakukan oleh pemain ke-5 (*passComplete5*) dikalikan fitur upaya melakukan serangan yang menghasilkan tendangan mengarah ke gawang yang dilakukan oleh pemain ke-6 (*shotCreation6*), upaya percobaan melakukan operan yang dilakukan oleh pemain ke-5 (*passAttempt5*) dikalikan fitur *shotCreation6*, fitur banyaknya bola yang didapatkan pemain ke-5 (*touches5*) dikalikan fitur *shotCreation6*, fitur operan berhasil yang dilakukan oleh pemain ke-5 (*passComplete5*) dikalikan fitur menit bermain dari pemain ke-9 (*menit9*) dan fitur pemain ke-5 melakukan *dribble* bola (*carries5*) dikalikan fitur *shotCreation6*. Lalu fitur upaya tackle yang dilakukan oleh pemain ke-1 (*tackle1*) dikalikan fitur dimana pertandingan dilaksanakan (*HomeAway1*), fitur upaya pemain ke-2 menahan tendangan dari musuh (*blocks2*) dikalikan fitur kartu kuning yang didapatkan pemain ke-8 (*yellowCard8*), fitur upaya pemain ke-2 melakukan pressing (*press2*) dikalikan fitur upaya pemain ke-9 menahan tendangan dari musuh (*blocks9*), fitur tackle1 dikalikan fitur *HomeAway9*, dan fitur tackle1 dikalikan fitur *HomeAway8* digunakan untuk memprediksi goal kemasukan yang dialami oleh tim tersebut. Model ini juga dibandingkan dengan model regresi linear sebagai *baseline*. Hasil evaluasi performa model menunjukkan bahwa performa model regresi polinomial berganda kuadrat memiliki akurasi sebesar 0.13 lebih tinggi dibandingkan dengan model regresi linear untuk dataset yang tidak menggunakan fitur *goalCreation*. Namun, performa model regresi linear 0.09 lebih besar dibandingkan model regresi polinomial berganda kuadrat untuk pemakaian dataset yang menggunakan fitur *goalCreation*.

Kata kunci: EPL, statistik, *Machine Learning*, korelasi *Pearson*, regresi polinomial berganda kuadrat.

ABSTRACT

Rifqi Julian Hasyari. 105218037. IS MULTIVARIATE SQUARED REGRESSION APPROPRIATE FOR ANALYZING PERFORMANCE OF ENGLISH PREMIER LEAGUE FOOTBALL TEAMS.

Football is no longer just a hobby, it has grown into a big industry. The English Premier League (EPL) is the one of the football leagues with the largest market among other football leagues. Ever since football became an industry, every team has worked to improve their form. One way is to use machine learning to utilize the statistical data they have. The purpose of this study is to identify influential features and performance results for models built using multivariate polynomial regression. This study uses statistical data web collected from the fbref.com website. Data are processed and feature multiplied to meet the requirement of building multivariate polynomial regression models. In addition, this study uses Pearson's correlation to show the correlation value for each features. The performance of the constructed model was tested using the R-squared method and mean absolute error (MAE). This study produces features suitable for predicting goal occurrences number of successful passes made by 5th player (passComplete5) multiplied by feature attempted attack resulting in a shot on goal taken by player 6 (shotCreation6), attempted passes made by player 5 (passAttempt5) multiplied shotCreation6 feature, feature of the number of balls that the 5th player gets (touches5) multiplied by the shotCreation6 feature, feature of the successful pass made by the 5th player (passComplete5) multiplied by the minutes played feature of the player 9th (menit9) and the 5th player features dribble the ball (carries5) multiplied by the shotCreation6 feature. feature of tackle attempts made by the 1st player (tackle1) multiplied by the feature where the match was held (HomeAway1), feature of the 2nd player's attempts to blocking enemy shots (blocks2) multiplied by the yellow card feature obtained by the 8th player (yellowCard8), feature the 2nd player's attempts to press (press2) are multiplied by the 9th player's blocking enemy shots (blocks9), the tackle1 feature is multiplied by the HomeAway9 feature, and the tackle1 feature is multiplied by the HomeAway8 feature are used to predict the teams conceded. This model is also compared with a linear regression model as a baseline. The model performance evaluation results show that the multivariate polynomial regression model outperforms the linear regression model by 0.13 accuracy for the dataset without the goalCreation function. However, the performance of the 0.09 linear regression model outperforms his second-order multivariate polynomial regression model on the dataset using the goalCreation function.

Keywords: *EPL, statistic, machine learning, Pearson corelation, multivariate polynomial regression.*



KATA PENGANTAR

Segala puji bagi Allah SWT atas rahmat dan nikmat yang diberikan sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul Analisis Data Performa Tim Sepak bola Liga Inggris Dengan Menggunakan Regresi Polinomial Berganda Kuadratik.

Laporan Tugas Akhir dapat selesai karena bantuan dan dukungan dari berbagai pihak. Penulis ucapkan terima kasih kepada Bapak Rangga Ganzar Noegraha, Ph. D dan Bapak Ade Irawan, Ph.D selaku pembimbing penelitian tugas akhir yang tidak lelah memberikan saran dan masukan kepada penulis. Penulis juga ucapkan terimakasih kepada Bapak Muhamad Koyimatu, Ph.D dan Ibu Ariana Yunita, M.I.T., MBA atas saran dan masukan yang membangun dalam pengerjaan Tugas Akhir. Selanjutnya, penulis juga ingin mengucapkan terima kasih kepada orang tua serta keluarga berkat dukungan dan doa sehingga penulis mampu menyelesaikan laporan Tugas Akhir ini. Selain itu, Penulis juga mengucapkan terima kasih kepada seluruh dosen Program Studi Ilmu Komputer Universitas Pertamina atas ilmu yang sudah diberikan kepada penulis selama menempuh bangku kuliah. Ucapan terima kasih juga diberikan kepada teman-teman Mahasiswa Ilmu Komputer Universitas Pertamina yang telah membantu dan memberi dukungan kepada penulis selama di bangku kuliah.

Penulis sadar bahwa karya tulis ini masih jauh dari kata sempurna. Penulis berharap dapat memperoleh kritik dan saran membangun untuk penelitian mendatang. Semoga tulisan ini dapat bermanfaat, baik untuk keperluan penelitian selanjutnya maupun untuk menambah ilmu pengetahuan bagi pembaca.

Jakarta, 11 Februari 2023

Rifqi Julian Hasyari



DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
BAB I PENDAHULUAN	2
1.1 Latar Belakang	2
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Machine Learning</i>	6
2.2 Data Penelitian	8
2.3 Seleksi Fitur	8
2.4 <i>Split Data</i>	8
2.5 Metode Regresi Polinomial Berganda	9
2.6 Penelitian Terkait	9
BAB III METODE PENELITIAN	12
3.1 Pendekatan dan Metode Penelitian	12
BAB IV HASIL DAN PEMBAHASAN	15
4.1 Pengumpulan Data	15
4.2 <i>Pre-processing</i> data	17
4.3 Korelasi terhadap fitur goal dan Kemasukan	19
4.3.1 Korelasi untuk dataset 1 yang memiliki fitur goalCreation	19



4.3.2	Korelasi untuk dataset 2 tanpa fitur goalCreation	20
4.4	Hasil Uji Model	22
4.4.1	Dataset 1	22
4.4.2	Dataset 2	22
4.5	Klasemen Berdasarkan Performa Model	22
4.6	Prediksi Tahun 2022-2023	23
2.	BAB V KESIMPULAN DAN SARAN	26
5.1	Kesimpulan	26
5.2	Saran	27
	DAFTAR PUSTAKA	29
	LAMPIRAN A Pengumpulan Data	32



DAFTAR TABEL

Tabel 2.1	Analisis beberapa Hasil Penelitian Terkait	10
Tabel 4.1	Fitur yang ada pada dataset beserta keterangannya	15
Tabel 4.2	Fitur yang digunakan yang sudah melalui tahapan <i>pre-processing</i>	17
Tabel 4.3	Lima contoh fitur yang telah dikelompokkan dan dikalikan tiap fitur.	19
Tabel 4.4	Korelasi positif tertinggi terhadap fitur goal dengan menggunakan dataset 1	19
Tabel 4.5	Korelasi negatif tertinggi terhadap fitur goal dengan menggunakan dataset 1	19
Tabel 4.6	Korelasi positif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 1	20
Tabel 4.7	Korelasi negatif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 1	20
Tabel 4.8	Korelasi positif tertinggi terhadap fitur goal dengan menggunakan dataset 2	21
Tabel 4.9	Korelasi negatif tertinggi terhadap fitur goal dengan menggunakan dataset 2	21
Tabel 4.10	Korelasi positif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 2	21
Tabel 4.11	Korelasi negatif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 2	21
Tabel 4.12	R2 Score dan Mean Absolute Error Model 1	22
Tabel 4.13	Hasil performa model menggunakan kolom goal creation	22
Tabel 4.14	R2 Score dan Mean Absolute Error Model 2	23
Tabel 4.15	Hasil Performa model tanpa kolom goal creation	23
Tabel 4.16	Hasil prediksi goal dan kemasukan di Liga Inggris tahun 2022-2023	24
Tabel 4.17	Hasil prediksi Klasemen 2023	24

DAFTAR GAMBAR

Gambar 3.1	Diagram alir penelitian.	12
Gambar 4.1	Klasemen Berdasarkan uji Performa Model 1 dan Klasemen Asli Liga Inggris 2021-2022	23



1. Dilarang mengutip karya tulis ini, kecuali:
 - a. menyebutkan sumber sesuai kaidah keceandekiaan;
 - b. pengutipan hanya untuk keperluan pendidikan, penulisan karya ilmiah atau penelitian;
 - c. pengutipan tidak merugikan Universitas Pertamina.
2. Dilarang mempublikasikan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa adanya izin dari Universitas Pertamina.





BAB I PENDAHULUAN

1.1 Latar Belakang

Hasil penelitian *Statista Research Department* tahun 2022 menunjukkan bahwa olahraga bukan hanya menjadi sebuah hobi, namun juga telah berkembang menjadi sebuah industri. (Gough, 2022) Industri olahraga global diproyeksikan tumbuh pada *Compound Annual Growth Rate (CAGR)* 41,3% terhitung sejak tahun 2021 hingga tahun 2022. Selain itu, pasar olahraga diperkirakan akan melebihi USD 700 miliar pada tahun 2026. Dari sekian banyak cabang olahraga, sepak bola adalah salah satu olahraga paling populer di dunia, dengan lebih dari 3,5 miliar penggemar di seluruh dunia (Ver-outsos, 2022). Sepak bola menjadi sebuah pasar bisnis yang sangat menjanjikan dan sangat besar, dikarenakan terdapat banyak sponsor yang berkontribusi, tidak hanya untuk sebuah liga yang berjalan, namun juga untuk sebuah tim, bahkan untuk pemain itu sendiri. Selain itu, terdapat pemasukan dari penjualan tiket penonton yang ingin datang ke stadion. Perjanjian untuk penayangan pertandingan oleh perusahaan stasiun TV juga ikut serta memberikan andil dalam perputaran uang yang ada pada industri sepak bola (Chakraborty and Niladri, 2022).

Salah satu liga profesional sepak bola yang memiliki pasar terbesar di dunia adalah liga Inggris atau dikenal sebagai *English Premier League (EPL)*. Hal ini dapat dilihat berdasarkan beberapa aspek yang ada seperti media yang meliput, hadiah kompetisi yang besar, persaingan kompetisi yang ketat dan rasio goal per-game yang tertinggi dibandingkan liga lain. Aspek-aspek tersebut menggambarkan liga Inggris masih menjadi salah satu liga terbaik di dunia (Spacey, 2022). Sejak tahun 1992, *Premier League* menjadi liga yang paling banyak mendapat perhatian. Tidak hanya warga Inggris yang menjadi penikmat liga, namun jutaan orang di seluruh dunia turut menikmati kompetisi liga Inggris. Hal ini menjadikan liga Inggris sebagai olahraga yang paling banyak ditonton selama lebih dari satu dekade. Dalam 10 tahun terakhir, terdapat total 5 tim yang berbeda telah memenangkan liga Inggris. Hal ini menunjukkan bahwa persaingan perebutan juara tidak didominasi tim manapun (Smith, 2022). Oleh karena itu, setiap tim berusaha untuk meningkatkan performa mereka. Salah satu caranya adalah dengan memanfaatkan data statistik yang mereka punya.

Penggunaan data statistik dalam dunia olahraga telah dilakukan pada cabang olahraga kasti oleh tim profesional Oakland Athletics di Amerika Serikat. (Cho, 2018) Tim profesional Oakland Athletics adalah salah satu tim yang menerima pendapatan terendah. Oleh karena itu, Oakland Athletics menggunakan analisis untuk memaksimalkan apa yang mereka miliki agar dapat bersaing dengan tim yang memiliki pendapatan lebih besar seperti New York Yankees dan Boston Red Sox. Dengan menggunakan konsep *Moneyball*, mereka mencari cara untuk dapat menekan anggaran namun tetap harus meningkatkan efektivitas dan efisiensi kualitas tim. Cabang olahraga kasti telah menggunakan teknologi yang didukung oleh ilmu pembelajaran mesin untuk memprediksi di mana pemukul akan bermanuver, seberapa keras bola akan dipukul, dan arah pemukul (*batter*), serta lemparan pelempar (*pitcher*) (Nourse and Garret, 2021).

Teknologi pembelajaran mesin tidak hanya digunakan dalam dunia kasti, namun juga digunakan di liga Inggris oleh klub yang telah dipromosikan ke EPL pada musim 2021-2022 lalu, yaitu Brentford FC (Ahmad et al., 2019). Brentford FC menggunakan ilmu pembelajaran mesin untuk

merekruit pemain baru, dengan memanfaatkan data statistik pemain untuk mengevaluasi dan memprediksi kinerjanya. Hal ini dianggap sebagai metode yang sangat baik bagi Brentford FC untuk diterapkan, karena dapat membantu menemukan bakat terbaik dan juga strategi bisnis yang tepat untuk perusahaan mereka. Teknologi pembelajaran mesin dapat digunakan juga untuk memprediksi skor pertandingan sepak bola. Penelitian serupa telah dilakukan oleh penulis di (Prabowo, 2020), namun belum ada yang mengusulkan untuk menggunakan metode regresi polinomial berganda kuadratik untuk melakukan prediksi hasil pertandingan sepak bola. Metode ini berhasil digunakan untuk menguji hipotesis bahwa latihan berkelanjutan dapat membantu mengurangi dampak penurunan performa atletik yang diakibatkan oleh penuaan di cabang olahraga lari atletik (Young and Starkes, 2005).

Oleh karena itu, penelitian ini mengusulkan penggunaan regresi polinomial berganda kuadratik untuk cabang olahraga sepakbola. Penelitian ini melakukan prediksi performa tim liga Inggris di tahun 2022 sampai dengan tahun 2023 berdasarkan statistik setiap pemain yang memainkan pertandingan di musim sebelumnya. Performa tim yang dimaksud pada penelitian ini adalah berapa banyak tim tersebut mencetak gol, dan berapa banyak lawannya mencetak gol pada setiap pertandingan. Sedangkan data-data statistik dari setiap pemain di setiap pertandingan diambil dari suatu situs publik.

1.2 Rumusan Masalah

Metode regresi polinomial berganda kuadratik belum pernah diusulkan sebelumnya untuk memprediksi hasil pertandingan sepak bola sehingga belum diketahui fitur apa saja yang sesuai untuk memprediksi skor pertandingan untuk metode tersebut. Sehingga masalah yang diangkat dalam penelitian ini adalah:

1. Fitur apa saja yang sesuai untuk digunakan dalam memprediksi skor pertandingan setiap tim di EPL?
2. Bagaimana performa dari regresi polinomial berganda kuadratik untuk memprediksi hasil pertandingan sepak bola?

1.3 Batasan Masalah

Batasan masalah yang ingin diimplementasikan untuk masalah-masalah yang dirumuskan antara lain:

1. Studi kasus yang dilakukan hanya pada tim yang berkompetisi di liga Inggris (EPL) pada musim 2022-2023.
2. Pengujian performa hanya dilakukan terhadap data pertandingan hingga *gameweek* 19 musim 2022-2023.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mendapatkan fitur yang sesuai serta mendapatkan hasil performa dari penggunaan regresi polinomial berganda kuadratik.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat menjadi sebuah referensi untuk menyatakan apakah metode regresi polinomial berganda kuadratik bisa digunakan atau tidak dapat digunakan untuk melakukan

prediksi pada cabang olahraga sepakbola. Manfaat lain dari penelitian ini diharapkan dapat menjadi sebuah acuan bagi perusahaan-perusahaan yang ingin menyelenggarakan kebutuhan material sebagai sponsor tim dengan melihat prediksi performa tim. Selain itu, penelitian ini juga diharapkan dapat menjadi bahan evaluasi dari setiap tim untuk mengamati performa mereka di musim yang akan datang.



Universitas
Pertamina

Sub-Cipta milik Universitas Pertamina

Diindungi Undang-Undang

Copyright of Universitas Pertamina

1. Dilarang mengutip karya tulis ini, kecuali:
 - a. menyebutkan sumber sesuai kaidah kecependekiaan;
 - b. pengutipan hanya untuk keperluan pendidikan, penulisan karya ilmiah atau penelitian;
 - c. pengutipan tidak merugikan Universitas Pertamina.
2. Dilarang mempublikasikan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa adanya izin dari Universitas Pertamina.





BAB II

TINJAUAN PUSTAKA

Tinjauan Pustaka penelitian ini dimulai dengan pembahasan mengenai *machine learning*. Setelah itu, pembahasan berikutnya tentang data pemain yang diambil dari *open source web* yaitu situs <https://fbref.com/en/> sebagai sumber data penelitian. Untuk selanjutnya di dokumen ini, kata FBREF mengacu ke situs tersebut. Kemudian, pembahasan dilanjutkan dengan seleksi fitur, *split* data dan metode penelitian yang digunakan, yaitu regresi polinomial berganda kuadratik lalu ditutup dengan penelitian terkait.

2.1 *Machine Learning*

Machine learning atau pembelajaran mesin adalah studi tentang algoritme komputer yang mampu mengenali pola dalam data dengan mengubah berbagai informasi menjadi tindakan nyata tanpa dibantu manusia sedikit mungkin. Pembelajaran mesin memungkinkan kita membuat mesin (komputer) yang dapat belajar dari informasi yang ada, memungkinkan mereka membuat keputusan sendiri tanpa harus diprogram ulang (Kurniawan, 2022). *Machine learning* dengan Python memiliki tiga proses penting dalam pembelajaran mesin, yaitu:

1. Pengumpulan data, proses yang bertujuan untuk mengumpulkan data mentah dari sumber-sumber tertentu dan dikumpulkan menjadi satu.
2. Abstraksi, proses mengubah data mentah yang dikumpulkan menjadi model yang lebih umum.
3. Generalisasi, proses yang menggunakan model hasil abstrak sebagai dasar pengambilan keputusan.

Pembelajaran mesin dapat memberikan wawasan tentang struktur dan pola dalam kumpulan data besar. Hal ini juga digunakan untuk membangun model dengan belajar dari kumpulan data yang ada untuk memprediksi hasil atau perilaku (Rebala et al., 2019). Salah satu permasalahan yang diselesaikan dalam pembelajaran mesin adalah prediksi, didasari olah data historis dan menggunakannya untuk memprediksi nilai masa depan atau dengan kata lain melakukan prediksi terhadap nilai yang akan muncul.

Berdasarkan buku (Hiran et al., 2021) terdapat 7 langkah untuk melakukan pembelajaran mesin, yaitu:

1. **Gathering Data**

Langkah ini sangat penting untuk kualitas dan kuantitas data yang akan digunakan untuk membuat model. Salah satu langkahnya adalah mengumpulkan informasi relevan yang dapat digunakan untuk pelatihan model. Fase pengumpulan data adalah dasar dari *machine learning*. Kesalahan seperti memilih fungsi yang salah atau fokus pada entri terbatas dalam kumpulan data pemodelan sama sekali tidak efektif.

2. **Preparing The Data**

Setelah data yang dibutuhkan terkumpul, melalui proses ini kita mengumpulkan semua data menjadi satu. Setelah itu kita memproses data-data yang telah dikumpulkan agar siap untuk digunakan dan dibuat modelnya. Pada proses ini kita dapat membersihkan data-data yang duplikat, menghapus data-data yang kosong, dan masih banyak yang lainnya sampai data telah siap digunakan. Selain itu dalam proses ini kita juga melakukan seleksi fitur untuk melihat hubungan dari setiap atribut yang ada.

3. **Choosing a Model**

Ada beberapa model yang dikembangkan oleh ilmuwan data yang dapat digunakan untuk berbagai keperluan dan mencapai tujuan. Berbagai kategori model bagus untuk pemodelan model yang mendasari berbagai jenis kumpulan data. Model ini dirancang untuk tujuan yang berbeda, misalnya beberapa model lebih cocok untuk pengolahan kata, sedangkan model lainnya mungkin lebih siap untuk pemrosesan gambar bahkan untuk pengolahan angka.

4. **Training**

Proses pembelajaran mesin adalah pelatihan dari model. Hal ini juga berguna untuk mengetahui bidang dimana model akan diimplementasikan. Proses pelatihan melibatkan beberapa inisialisasi nilai acak misalnya X dan Y dari model, kemudian memprediksi hasilnya dengan nilai ini dan kemudian membandingkannya dengan prediksi model untuk menyesuaikan nilainya dengan prediksi sebelumnya. Proses ini kemudian diulangi dan setiap siklus pembaruan disebut *train*. Tahapan ini mirip dengan ketika seorang anak belajar naik sepeda, dimana ketika pertama kali belajar mereka mungkin jatuh berkali-kali, tetapi setelah beberapa saat mereka berkembang menjadi memahami proses dengan lebih baik dan mampu bereaksi lebih baik terhadap perbedaan situasi saat bersepeda.

5. **Evaluation** Setelah pelatihan, model tersebut perlu diuji untuk melihat apakah dapat beroperasi dengan baik dalam situasi dunia nyata. Oleh karena itu, kita menggunakan kumpulan data yang berbeda dari data *train* untuk melakukan evaluasi. Hal ini menyebabkan model akan berusaha menyelesaikan permasalahan yang berbeda dari situasi saat *training*.

6. **Hyperparameter Tuning** Ada perbaikan lebih lanjut setelah proses *training* dan *evaluation* yaitu adanya kemungkinan untuk memperbaiki atau menyesuaikan parameter. Penyesuaian parameter ini dilakukan dengan harapan meningkatkan akurasi model *training* yang telah dibuat.

7. **Prediction** Langkah terakhir dalam pembelajaran mesin adalah melakukan prediksi. Pada tahapan ini model telah dianggap selesai sehingga dapat digunakan untuk memprediksi hasil dari permasalahan yang diambil.

2.2 Data Penelitian

Perkembangan internet memberikan dampak yang positif dalam memudahkan kita untuk mengumpulkan dan mendapatkan informasi. Penelitian ini menggunakan data statistik yang dikumpulkan melalui situs FBREF. Situs ini dibuat pada Juni 2018. Awalnya FBREF hanya mencakup 6 liga nasional seperti liga Inggris, Prancis, Spanyol, Italia, Jerman, dan Amerika Serikat. Sejak itu, FBREF telah berkembang dengan lebih dari 45 negara dan 140 kompetisi. Melalui situs FBREF tersebut kita dapat menemukan berbagai informasi berupa data olahraga sepak bola.

Penelitian ini menggunakan data statistik *match logs* atau statistik dari setiap pemain di kompetisi liga Inggris pada tahun 2021-2022 yang dikumpulkan melalui situs FBREF. Data tersebut berisikan 41 kolom yang merupakan detail dari statistik pemain, kolom-kolom tersebut dipilih untuk menjadi *feature* dalam melakukan prediksi.

2.3 Seleksi Fitur

Data-data yang telah dikumpulkan dianalisis untuk dilihat hubungan dari setiap fitur yang ada. Salah satu metode untuk menganalisis kualitas dari setiap fitur ialah dengan menggunakan analisis korelasi. Analisis korelasi menentukan kekuatan hubungan antara dua fitur, yang mungkin satu variabel dependen dan satu variabel independen atau bahkan dua variabel independen. Secara numerik, keterkaitan ini biasanya ditentukan oleh nilai desimal yang dikenal sebagai koefisien korelasi. Biasanya koefisien yang lebih tinggi menunjukkan bahwa dua fitur berkorelasi kuat dan nilai yang lebih rendah menunjukkan sebaliknya.

Penelitian ini menggunakan korelasi *Pearson*. Korelasi *Pearson* menentukan nilai linear dari fitur yang digunakan, sehingga baik digunakan untuk model yang linear juga. Koefisien korelasi *Pearson* dapat dihitung dengan menggunakan rumus berikut

$$C_{A,B} = \frac{\text{Covariance}(A, B)}{\sigma_A \sigma_B} \quad (2.1)$$

dimana $C_{A,B}$ merupakan koefisien korelasi dengan rentang nilai dari -1 hingga 1, lalu $\text{Covariance}(A, B)$ merupakan kovarians dan $\sigma_A \sigma_B$ merupakan standar deviasi untuk masing-masing fitur A dan fitur B. (Kumar and Chong, 2018).

2.4 Split Data

Data perlu dipisah antara masukan dan luaran untuk melakukan prediksi. Namun, pada penggunaan *machine learning*, kualitas data dan strategi penggunaan data berdampak besar pada seberapa baik kinerjanya. Data dibagi menjadi dua set terpisah, yaitu untuk *training* dan *test*. Banyak peneliti menyarankan untuk menggunakan rasio 70/30, 80/20, dan 90/10 (*training/test*) untuk menghasilkan kumpulan data menggunakan algoritma *machine learning*. (Joseph, 2022) Mengenai rasio apa yang ideal atau optimal untuk digunakan dalam kumpulan data tertentu, tampaknya tidak dinyatakan secara eksplisit di mana pun. Set *training* ini kemudian digunakan agar sesuai dengan model statistik dan pembelajaran mesin, yang akhirnya diuji pada set pengujian. Umumnya, rasio yang sering digunakan adalah 80:20, yang berarti 80% data digunakan untuk *training* dan 20% data digunakan untuk *test*. Tidak ada rasio yang ideal atau optimal untuk digunakan (Joseph, 2022).

2.5 Metode Regresi Polinomial Berganda

Regresi polinomial berganda digunakan untuk memodelkan hubungan kompleks dengan banyak variabel. Hubungan kompleks ini biasanya non-linear dan berdimensi tinggi. Setelah persamaan (model) yang akurat dibuat atau ditemukan, persamaan tersebut kemudian dapat digunakan untuk membuat prediksi yang akurat (Kaplan, 2022).

Model regresi polinomial berganda dapat dirumuskan menjadi (Dette and Studden, 1997)

$$Y = \alpha_0 + \sum_{i=1}^1 \alpha_i x_i + \sum \alpha_{i_1 i_2} x_{i_1} x_{i_2} + \epsilon \quad (2.2)$$

dimana α menunjukkan nilai intersep

2.6 Penelitian Terkait

Beberapa peneliti sebelumnya telah melakukan penelitian terkait EPL, prediksi, dan regresi polinomial berganda dengan hasil penelitian yang berbeda-beda, antara lain:

1. (Prabowo, 2020) melakukan Prediksi Hasil Pertandingan Sepakbola English Premier League Dengan Menggunakan Algoritma *K-Nearest Neighbors* dan *Naive Bayes Classifier*. Penelitian tersebut bertujuan untuk memprediksi hasil pertandingan *English Premier League* yang berupa Home Win (H), Away Win (A), dan Draw (D).
2. (Imran et al., 2022) mengambil permasalahan bahwa beton sebagai polutan telah menyebabkan penipisan sumber daya, energi, dan efek rumah kaca emisi gas dalam penggunaannya. Oleh karena itu, Imran et al. (2022) ingin mencari solusi melalui permasalahan tersebut dengan merancang model prediksi menggunakan Multivariat Regresi Polinomial (MPR) yang bertujuan memprediksi kekuatan tekanan beton.

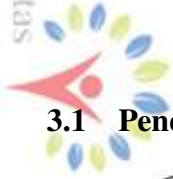
Melalui penelitian-penelitian tersebut, diperoleh berbagai macam hasil dan dikaji lebih lanjut dengan menggunakan teknik 3C2S. Teknik itu sendiri berisikan *compare* yaitu menemukan persamaan dari penelitian tersebut terhadap penelitian yang sedang dilakukan, lalu *contrast* yaitu perbedaan antar penelitian dan *criticize* yang merupakan pendapat mengenai isi penelitian tersebut. Lalu terdapat *synthesize* membandingkan serta menggabungkan beberapa penelitian menjadi sebuah ide dan *summarize* yaitu rangkuman dari penelitian tersebut. Hasil literature review yang telah dilakukan dengan menggunakan teknik 3C2S dapat dilihat pada pada Tabel 2.1 (Hasibuan and Zainal, 2007).

Tabel 2.1. Analisis beberapa Hasil Penelitian Terkait

Makalah	<i>Compare</i>	<i>Contrast</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
Makalah 1: Prediksi Hasil Pertandingan Sepakbola English Premier League Dengan Menggunakan Algoritma <i>K-Nearest Neighbors</i> dan <i>Naive Bayes Classifier</i>	Prediksi yang dilakukan adalah untuk mengetahui hasil prediksi pertandingan EPL.	Menggunakan algoritma <i>K-Nearest Neighbors</i> dan <i>Naive Bayes Classifier</i> , <i>feature</i> yang akan digunakan untuk melakukan prediksi berbeda dengan penelitian ini.	Penelitian tersebut tidak menampilkan fitur yang lebih mempengaruhi output	Algoritma <i>K-Nearest Neighbors</i> dan <i>Naive Bayes Classifier</i> dapat digunakan untuk memprediksi hasil pertandingan sepak bola.	Untuk memprediksi hasil pertandingan sepak bola berdasarkan model yang dibuat oleh peneliti, menggunakan algoritma <i>Naive Bayes Classifier</i> memiliki kinerja yang lebih bagus dibandingkan dengan algoritma <i>K-Nearest Neighbors</i>
Makalah 2: <i>Development of Prediction Model to Predict the Compressive Strength of Eco-Friendly Concrete Using Multivariate Polynomial Regression Combined with Stepwise Method.</i>	Menggunakan regresi polinomial berganda untuk membangun model prediksi.	Regresi polinomial berganda digunakan sebagai solusi untuk memprediksi kekuatan tekanan campuran beton.	Penjelasan mengenai dataset yang digunakan sangat terbatas.	Hasil penelitian dapat dimanfaatkan untuk menjaga dan meningkatkan kelangsungan hidup keberlanjutan jangka panjang dengan mengurangi konsekuensi lingkungan dalam penggunaan beton.	Hasil penelitian memperlihatkan bahwa nilai prediksi kuat tekan beton ramah lingkungan untuk dataset pengujian mendekati nilai eksperimen menggunakan model Multivariat Regresi Polinomial (MPR).

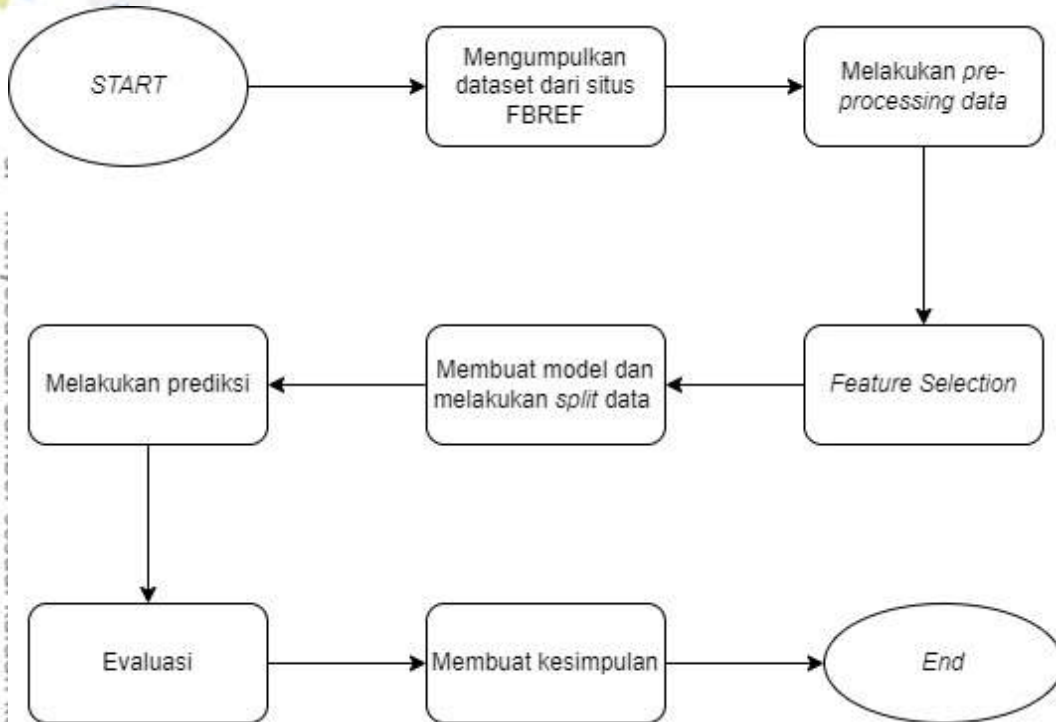
Setelah meninjau beberapa penelitian terkait diatas, terdapat beberapa perbedaan dengan penelitian yang dilakukan pada tugas akhir ini, antara lain prediksi skor pertandingan setiap tim pada EPL, prediksi klasemen akhir liga Inggris pada musim 2022/2023, dan mengetahui keakuratan hasil prediksi model yang dibangun dengan menggunakan regresi polinomial berganda kuadratik.





BAB III METODE PENELITIAN

3.1 Pendekatan dan Metode Penelitian



Gambar 3.1. Diagram alir penelitian.

Penelitian ini adalah jenis penelitian kuantitatif dengan menggunakan data statistik pemain liga Inggris. Langkah-langkah yang dilakukan pada penelitian ini disusun seperti pada Gambar 3.1. Teknik pengumpulan data penelitian ini adalah observasi. Data yang dibutuhkan adalah kumpulan data statistik setiap pemain pada setiap pertandingan liga Inggris terhitung sejak musim 2021-2022. Pengumpulan data dilakukan menggunakan metode *web scrapping* dengan memanfaatkan bahasa pemrograman Python terhadap situs FBREF. Perintah dari kode tersebut dijalankan berulang-ulang dengan mengganti *link* setiap pemain di situs tersebut sampai seluruh pemain liga Inggris pada musim 2021-2022 telah diambil datanya. Kode tersebut dapat dilihat pada Lampiran A.

Setelah data yang dibutuhkan untuk penelitian terkumpul, data diolah sesuai dengan kebutuhan penelitian. Sebelum melakukan tahapan analisis, data dibersihkan dan dipersiapkan dengan melakukan beberapa hal seperti penghapusan terhadap fitur yang tidak dibutuhkan, menambahkan fitur baru jika dibutuhkan, dan menangani *missing value*. Penanganan *missing value* dilakukan dengan dua cara. Cara pertama adalah dengan menghapus *missing value* karena tidak memiliki keterkaitan dengan penelitian yang dilakukan. Lalu cara kedua adalah dengan mengganti *missing value* berjenis data *categorical*. Selain itu, terdapat data yang diganti menjadi numerik, dengan tujuan memudahkan proses pengolahan data. Hal lainnya adalah mengganti nama fitur agar penamaan menjadi lebih jelas.

Tahapan selanjutnya adalah *feature selection*. Tahapan ini dilakukan untuk memilih fitur yang akan digunakan. Dengan metode analisis korelasi menggunakan korelasi *Pearson*. Setiap fitur diko-

relasikan terhadap fitur goal dan fitur kemasukan dengan tujuan melihat dan mengetahui besarnya nilai korelasi mereka. Setelah proses pemilihan fitur selesai selanjutnya melakukan *split* data menjadi data *train* dan data *test*, yang kemudian digunakan untuk membuat model. *Split* data dilakukan dengan tujuan untuk memvalidasi model yang telah dibuat untuk melihat eror model, dan melihat persentase performa dari model yang telah dibuat. Penelitian ini menggunakan metode regresi polinomial berganda yang hanya memproses data angka atau numerik.

Penelitian ini melakukan prediksi goal dan kemasukan yang dialami setiap tim pada setiap pertandingan. Tim yang akan diprediksi pada penelitian ini hanya tim yang terdapat pada liga Inggris di musim 2021-2022 yang masih bertahan atau tidak terdegradasi di musim 2022-2023. Penelitian ini tidak memprediksi tim yang baru saja naik ke liga Inggris di musim ini.

Model dan prediksi yang telah dibuat selanjutnya dievaluasi. Evaluasi yang dilakukan adalah mengukur besar eror. Nilai eror dilakukan dengan membandingkan prediksi hasil pertandingan dan hasil pertandingan yang sebenarnya. Eror terjadi jika prediksi dengan hasil sebenarnya tidak tepat sama. Berdasarkan nilai eror yang telah didapatkan melalui proses evaluasi, maka selanjutnya menyusun kesimpulan apakah metode ini cukup baik atau kurang baik dalam memprediksi performa tim sepak bola di liga Inggris di musim 2022-2023.







BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data mentah penelitian ini diambil melalui FBREF yang berisikan data statistik pemain liga Inggris pada musim 2021-2022. Menggunakan bahasa pemrograman Python, pengumpulan data mentah penelitian ini dilakukan dengan mengambil data pemain yang dilakukan secara berulang dengan menggantikan *link* pada perintah yang ada. Perintah tersebut diulang hingga seluruh data pemain yang bermain di liga Inggris di musim 2021-2022 terkumpul. Fitur dari data mentah yang telah berhasil dikumpulkan dapat dilihat pada Tabel 4.1. Data mentah yang terkumpul sebanyak 13.482 baris dengan 39 kolom. Kumpulan data mentah penelitian ini dapat dilihat dan diakses melalui tautan berikut ini <https://github.com/RifqiJH/FbrefEPL2021-2022/blob/main/allPlayer.csv>.

Tabel 4.1. Fitur yang ada pada dataset beserta keterangannya

Nama Fitur	Keterangan
Date	Tanggal dilaksanakannya pertandingan.
Day	Hari dilaksanakannya pertandingan.
Comp	Turnamen untuk pertandingan yang dilaksanakan.
Round	Babak atau ronde untuk pertandingan yang dilaksanakan.
Venue	Tempat dilaksanakannya pertandingan. Home untuk pertandingan yang dilaksanakan di markas sendiri sedangkan away untuk pertandingan yang dilaksanakan di kandang musuh.
Result	Skor pertandingan.
Squad	Tim yang dibela pemain.
Opponent	Tim yang dilawan pemain.
Start	Apakah pemain tersebut menjadi <i>starting line up</i> atau tidak.
Pos	Posisi dari pemain pada pertandingan tersebut.
Min	Total menit bermain dari setiap pemain pada pertandingan tersebut.
Gls	Goal yang diciptakan pemain.
Ast	Asisst yang diciptakan pemain.
PK	Total tendangan penalti yang berhasil dilakukan pemain pada pertandingan tersebut.
PKatt	Total percobaan tendangan penalti yang dilakukan pemain pada pertandingan tersebut.
Sh	Total tendangan yang dilakukan pemain pada pertandingan tersebut.

Nama Fitur	Keterangan
SoT	Total tendangan yang berhasil ke arah gawang yang dilakukan pemain pada pertandingan tersebut.
CrdY	Banyaknya kartu kuning yang didapatkan pemain pada pertandingan tersebut.
CrdR	Banyaknya kartu merah yang didapatkan pemain pada pertandingan tersebut.
Touches	Banyaknya bola yang didapatkan pemain pada pertandingan tersebut.
Press	Banyaknya pressing yang dilakukan setiap pemain terhadap pemain lawan yang memegang bola pada pertandingan tersebut.
Tkl	Banyaknya tackle yang dilakukan setiap pemain terhadap pemain lawan yang memegang bola pada pertandingan tersebut.
Int	Banyaknya kejadian pemain tersebut memotong operan yang dilakukan oleh musuh pada pertandingan tersebut.
Blocks	Banyaknya kejadian pemain tersebut menahan tembakan atau umpan yang dilakukan oleh musuh pada pertandingan tersebut.
xG	Probabilitas dari pemain tersebut mencetak goal berdasarkan peluang yang ada, termasuk dengan tendangan pinalti.
npG	Probabilitas dari pemain tersebut mencetak goal berdasarkan peluang yang ada, tidak termasuk dengan tendangan pinalti.
xA	Probabilitas <i>assist</i> yang dilakukan pemain.
SCA	Total dua tindakan ofensif/menyerang yang langsung menghasilkan tembakan ke gawang.
GCA	Total tindakan menyerang yang berhasil dijadikan goal seperti operan, dribble dan mendapatkan pelanggaran dari musuh.
Cmp	Operan berhasil yang dilakukan pemain pada pertandingan tersebut.
Att	Percobaan operan yang dilakukan pemain pada pertandingan tersebut.
Cmp%	Persentase operan berhasil yang dilakukan pemain pada pertandingan tersebut.
Prog	Operan yang berhasil dikirimkan terhadap rekan tim yang berada di sekitar kotak pertahanan musuh.
Carries	Total upaya pemain melakukan dribble bola dalam satuan yard.
Prog	Total upaya pemain melakukan dribble bola yang mengancam pertahanan musuh dalam satuan yard.
Succ (Dribble)	Total upaya pemain melakukan dribble bola yang berhasil pada pertandingan tersebut.
Att (Dribble)	Total upaya pemain melakukan dribble bola pada pertandingan tersebut.
Match Report	Fitur yang ada pada situs fbref untuk pergi ke halaman lain.
Nama	Nama pemain.

4.2 Pre-processing data

Tahapan berikutnya yang dilakukan adalah *pre-processing* data mentah. Melalui tahapan *pre-processing* pada data mentah, kegiatan yang dilakukan adalah menghilangkan atau menghapus *missing value*. *Missing value* dihapuskan pada data mentah karena dianggap tidak berkaitan dengan penelitian yang dilakukan. Selain itu melalui tahapan *pre-processing* juga dilakukan penghapusan fitur yang tidak digunakan, kemudian menambahkan fitur yang dibutuhkan, seperti fitur Goal dan fitur Kemasukan. Terakhir dilakukan penghapusan *outlier* pada kumpulan data mentah. *Outlier* pada data mentah penelitian ini adalah kumpulan pemain yang memiliki waktu bermain dibawah 45 menit. Perubahan yang terjadi setelah proses *pre-processing* diperlihatkan pada Tabel 4.2.

Tabel 4.2. Fitur yang digunakan yang sudah melalui tahapan *pre-processing*

Nama Fitur	Keterangan
tanggal	Tanggal dilaksanakannya pertandingan.
Team	Tim yang dibela pemain.
position	Posisi dari pemain pada pertandingan tersebut.
menit	Total menit bermain dari setiap pemain pada pertandingan tersebut.
shotTotal	Total tendangan yang dilakukan pemain pada pertandingan tersebut.
shotOnTarget	Total tendangan yang berhasil ke arah gawang yang dilakukan pemain pada pertandingan tersebut.
yellowCard	Banyaknya kartu kuning yang didapatkan pemain pada pertandingan tersebut.
redCard	Banyaknya kartu merah yang didapatkan pemain pada pertandingan tersebut.
touches	Banyaknya bola yang didapatkan pemain pada pertandingan tersebut.
press	Banyaknya pressing yang dilakukan setiap pemain terhadap pemain lawan yang memegang bola pada pertandingan tersebut.
tackle	Banyaknya tackle yang dilakukan setiap pemain terhadap pemain lawan yang memegang bola pada pertandingan tersebut.
intercept	Banyaknya kejadian pemain tersebut memotong operan yang dilakukan oleh musuh pada pertandingan tersebut.
blocks	Banyaknya kejadian pemain tersebut menahan tembakan atau umpan yang dilakukan oleh musuh pada pertandingan tersebut.
shotCreation	Total dua tindakan ofensif/menyerang yang langsung menghasilkan tembakan ke gawang.
goalCreation	Goal yang diciptakan pemain.
passComplete	Operan berhasil yang dilakukan pemain pada pertandingan tersebut.
passAttempt	Percobaan operan yang dilakukan pemain pada pertandingan tersebut.
passProgresive	Operan yang berhasil dikirimkan terhadap rekan tim yang berada di sekitar kotak pertahanan musuh.
carries	Total upaya pemain melakukan dribble bola dalam satuan yard.

Nama Fitur	Keterangan
carries>5yard	Total upaya pemain melakukan dribble bola yang mengancam pertahanan musuh dalam satuan yard.
DribblesSucces	Total upaya pemain melakukan dribble bola yang berhasil pada pertandingan tersebut.
Dribbles.1	Total upaya pemain melakukan dribble bola pada pertandingan tersebut.
Nama	Nama pemain.
goal	Total goal yang dimasukkan tim tersebut dalam pertandingan.
Kemasukan	Total goal yang memasukkan tim tersebut dalam pertandingan.
Minggu	Matchweek atau minggu ke berapa pertandingan tersebut dimainkan.
homeAway	Dimana pertandingan dilaksanakan. angka 1 jika bermain sebagai <i>Home</i> , dan angka 2 untuk bermain sebagai <i>away</i> .
Starting 11	Apakah pemain tersebut menjadi <i>starting line up</i> atau tidak. angka 1 jika pemain tersebut merupakan <i>starting line up</i> dan angka 2 jika tidak.
KlasemenTimMusuh	Berisikan posisi klasemen tim yang dilawan pada kompetisi Liga Inggris di musim lalu atau musim 2021-2022

Setelah data dibersihkan melalui tahapan *pre-processing* sesuai dengan kebutuhan penelitian dan telah siap untuk digunakan, selanjutnya kumpulan data tersebut diolah dengan menggunakan metode regresi polinomial berganda. Metode polinomial berganda yang dilakukan pada penelitian ini yaitu dengan melakukan perkalian terhadap setiap fitur yang ada. Sebelum dilakukan perkalian setiap fitur, dilakukan pengelompokan berdasarkan 11 pemain yang bermain dalam setiap pertandingan. Contoh perkalian antar fitur serta pengelompokan dapat dilihat pada Tabel 4.3

Tabel 4.3. Lima contoh fitur yang telah dikelompokkan dan dikalikan tiap fitur.

Fitur	Keterangan
shotTotal11intercept2	fitur shotTotal pemain ke-11 dikalikan fitur intercept pemain ke-11
shotTotal11tackle2	fitur shotTotal pemain ke-11 dikalikan fitur tackle pemain ke-2
menit11passComplete11	fitur menit pemain ke-11 dikalikan fitur passComplete pemain ke-11
menit1goalCreation10	fitur menit pemain ke-1 dikalikan fitur goalCreation pemain ke-10
passAttempt11goalCreation6	fitur passAttempt pemain ke-11 dikalikan fitur goalCreation pemain ke-6

4.3 Korelasi terhadap fitur goal dan Kemasukan

Penelitian ini melakukan analisis menggunakan 2 kasus yang dipisah dengan menggunakan 2 dataset, yaitu dataset 1 merupakan dataset yang menggunakan fitur goalCreation (gls), dan dataset 2 yaitu dataset yang tidak menggunakan fitur goalCreation (gls).

4.3.1 Korelasi untuk dataset 1 yang memiliki fitur goalCreation

Tabel 4.4 menunjukkan 5 fitur yang memiliki nilai korelasi positif tertinggi dibandingkan fitur yang lain yaitu touches6goalCreation6, passComplete5goalCreation6, touches5goalCreation6, passAttempt5goalCreation6 dan goalCreation6passAttempt6. Sedangkan Tabel 4.5 menunjukkan 5 fitur yang memiliki korelasi negatif tertinggi terhadap fitur goal yaitu KlasemenTeam, Kemasukan, press2blocks2, press2press4 dan blocks2press4.

Tabel 4.4. Korelasi positif tertinggi terhadap fitur goal dengan menggunakan dataset 1

Fitur	Hasil korelasi positif terhadap goal
touches6goalCreation6	0.445
passComplete5goalCreation6	0.444
touches5goalCreation6	0.442
passAttempt5goalCreation6	0.441
goalCreation6passAttempt6	0.439

Tabel 4.5. Korelasi negatif tertinggi terhadap fitur goal dengan menggunakan dataset 1

Fitur	Hasil korelasi negatif terhadap goal
KlasemenTeam	-0.232
Kemasukan	-0.197
press2blocks2	-0.176
press2press4	-0.167
blocks2press4	-0.164

Nilai korelasi *Pearson* untuk seluruh fitur terhadap goal dengan menggunakan dataset 1 berkisar

antara -0.23 sampai dengan -0.000001 untuk korelasi negatif dan berkisar antara 0.0000006 sampai dengan 0.445 untuk korelasi positif.

Tabel 4.6 menunjukkan fitur yang memiliki korelasi positif tertinggi terhadap fitur Kemasukan yaitu tackle1homeAway1, blocks2yellowCard8, press2blocks9, tackle1homeAway9 dan tackle1homeAway8. Sedangkan Tabel 4.7 menunjukkan 5 fitur yang memiliki korelasi negatif tertinggi terhadap fitur Kemasukan yaitu touches5menit11, passAttempt5menit11, passComplete5menit11, touches5menit10 dan touches5touches11.

Tabel 4.6. Korelasi positif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 1

Fitur	Hasil korelasi positif terhadap Kemasukan
tackle1homeAway1	0.172
blocks2yellowCard8	0.171
press2blocks9	0.167
tackle1homeAway9	0.166
tackle1homeAway8	0.165

Tabel 4.7. Korelasi negatif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 1

Fitur	Hasil korelasi negatif terhadap Kemasukan
touches5menit11	-0.319
passAttempt5menit11	-0.314
passComplete5menit11	-0.312
touches5menit10	-0.301
touches5touches11	-0.3

Nilai korelasi *Pearson* untuk seluruh fitur terhadap kemasukan dengan menggunakan dataset 1 berkisar antara -0.319 sampai dengan -0.0001 untuk korelasi negatif dan berkisar antara 0.00008 sampai dengan 0.172 untuk korelasi positif.

4.3.2 Korelasi untuk dataset 2 tanpa fitur goalCreation

Berdasarkan Tabel 4.4 dapat dilihat bahwa fitur goalCreation sangat mendominasi dan pada penelitian ini tidak mengharapkan salah satu fitur terlalu mendominasi. Penelitian ini mencoba menggunakan dataset lain dengan menghapus fitur goalCreation.

Tabel 4.8 menunjukkan 5 fitur yang memiliki nilai korelasi positif tertinggi terhadap fitur goal yaitu fitur passComplete5shotCreation6, passAttempt5shotCreation6, touches5shotCreation6, passComplete5menit9 dan carries5shotCreation6. Sedangkan Tabel 4.9 menunjukkan 5 fitur yang memiliki korelasi negatif tertinggi terhadap fitur goal yaitu KlasemenTeam, Kemasukan, press2blocks2, press2press4 dan blocks2press4.

Nilai korelasi *Pearson* untuk seluruh fitur terhadap goal dengan menggunakan dataset 2 berkisar antara -0.23 sampai dengan -0.000001 untuk korelasi negatif dan berkisar antara 0.00001 sampai dengan 0.334 untuk korelasi positif.

Tabel 4.10 menunjukkan fitur yang memiliki nilai korelasi positif tertinggi terhadap fitur Kemasukan yaitu tackle1homeAway1, blocks2yellowCard8, press2blocks9, tackle1homeAway9 dan tackle1homeAway8.

Tabel 4.8. Korelasi positif tertinggi terhadap fitur goal dengan menggunakan dataset 2

Fitur	Hasil korelasi terhadap goal
passComplete5shotCreation6	0.33
passAttempt5shotCreation6	0.325
touches5shotCreation6	0.319
passComplete5menit9	0.318
carries5shotCreation6	0.315

Tabel 4.9. Korelasi negatif tertinggi terhadap fitur goal dengan menggunakan dataset 2

Fitur	Hasil korelasi terhadap goal
KlasemenTeam	-0.232
Kemasukan	-0.197
press2blocks2	-0.176
press2press4	-0.167
blocks2press4	-0.164

Sedangkan Tabel 4.11 menunjukkan 5 fitur yang memiliki korelasi negatif tertinggi terhadap fitur kemasukan yaitu touches5menit11, passAttempt5menit11, passComplete5menit11, touches5menit10 dan touches5touches11.

Tabel 4.10. Korelasi positif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 2

Fitur	Hasil korelasi terhadap Kemasukan
tackle1homeAway1	0.172
blocks2yellowCard8	0.171
press2blocks9	0.167
tackle1homeAway9	0.166
tackle1homeAway8	0.165

Tabel 4.11. Korelasi negatif tertinggi terhadap fitur Kemasukan dengan menggunakan dataset 2

Fitur	Hasil korelasi terhadap Kemasukan
touches5menit11	-0.319
passAttempt5menit11	-0.314
passComplete5menit11	-0.312
touches5menit10	-0.301
touches5touches11	-0.3

Nilai korelasi *Pearson* untuk seluruh fitur terhadap Kemasukan dengan menggunakan dataset 2 berkisar antara -0.00005 sampai dengan -0.319 untuk korelasi negatif dan berkisar antara 0.00006 sampai dengan 0.172 untuk korelasi positif.

4.4 Hasil Uji Model

4.4.1 Dataset 1

Tabel 4.12 menunjukkan hasil evaluasi model dataset 1 dengan menggunakan R2 memiliki hasil sebesar -7.14 untuk kemasukan dan -1.72 untuk mencetak goal. Lalu terdapat nilai *Mean Absolute Error* sebesar 2.98 untuk kemasukan dan 1.53 untuk mencetak goal. Dalam membangun model, data set dipisah dengan perbandingan 80 : 20 di mana 80% data digunakan untuk training dan 20% digunakan untuk testing. Hasil testing ini menunjukkan bahwa nilai ketelitian model dibandingkan dengan kondisi kemenangan tim sebenarnya adalah 0.57. Model berikutnya dibangun dengan menggunakan 100% data set untuk training. Lalu model ini di-testing menggunakan 100% data yang sama. Nilai ketelitian model dibandingkan dengan kondisi kemenangan tim sebenarnya adalah 0.85. Dengan metode yang sama, hasil regresi linier memiliki nilai ketelitian 0.61 dan 0.94 seperti yang dapat dilihat pada Tabel 4.13

Tabel 4.12. R2 Score dan Mean Absolute Error Model 1

Model 1	Hasil
R2 score Test Kemasukan	-7.148214634223612
Mean Absolute Error Kemasukan	2.9837405204645706
R2 score test Goal	-1.721272419183216
Mean Absolute Error	1.5372002758989907

Tabel 4.13. Hasil performa model menggunakan kolom goal creation

Nama	Hasil
Hasil Prediksi dengan GoalCreation 20%	0.5714285714285714
Hasil Prediksi dengan GoalCreation	0.8597285067873304
Hasil Prediksi Linear dengan GoalCreation 20%	0.6165413533834586
Hasil Prediksi Linear dengan GoalCreation	0.9411764705882353

4.4.2 Dataset 2

Tabel 4.14 menunjukkan hasil evaluasi model dataset 2 dengan menggunakan R2 memiliki hasil sebesar -7.36 untuk kemasukan dan -2.88 untuk mencetak goal. Lalu terdapat nilai *Mean Absolute Error* sebesar 3.02 untuk kemasukan dan 2.03 untuk mencetak goal. Dalam membangun model, data set dipisah dengan perbandingan 80 : 20 di mana 80% data digunakan untuk training dan 20% digunakan untuk testing. Hasil testing ini menunjukkan bahwa nilai ketelitian model dibandingkan dengan kondisi kemenangan tim sebenarnya adalah 0.51. Model berikutnya dibangun dengan menggunakan 100% data set untuk training. Lalu model ini di-testing menggunakan 100% data yang sama. Nilai ketelitian model dibandingkan dengan kondisi kemenangan tim sebenarnya adalah 0.89. Dengan metode yang sama, hasil regresi linier memiliki nilai ketelitian 0.57 dan 0.76 seperti yang dapat dilihat pada Tabel 4.15

4.5 Klasemen Berdasarkan Performa Model

Gambar 4.1a merupakan hasil klasemen akhir Liga Inggris musim 2021-2022 berdasarkan hasil uji coba model dataset 1 sedangkan Gambar 4.1b merupakan klasemen akhir yang sebenarnya.

Tabel 4.14. R2 Score dan Mean Absolute Error Model 2

Nama	Hasil
R2 score Test Kemasukan	-7.36058733050341
Mean Absolute Error Kemasukan	3.0239641896865685
R2 score test Goal	-2.8811805298024726
Mean Absolute Error	2.0340968040283234

Tabel 4.15. Hasil Performa model tanpa kolom goal creation

Model	Hasil
Hasil Prediksi Tanpa GoalCreation 20%	0.5112781954887218
Hasil Prediksi Tanpa GoalCreation	0.8914027149321267
Hasil Prediksi Linear Tanpa GoalCreation 20%	0.5789473684210527
Hasil Prediksi Linear Tanpa GoalCreation	0.7616892911010558

Prediksi Klasemen(Model 1)				
Standings		Team	goal_difference	Points
0	1	Manchester City	7.000000	102
1	2	Liverpool	6.000000	99
2	3	Chelsea	7.000000	84
3	4	Arsenal	5.000000	75
4	5	Tottenham	5.000000	75
6	6	Manchester Utd	4.000000	66
6	7	West Ham	4.000000	63
7	8	Southampton	3.000000	63
8	9	Leicester City	4.000000	60
9	10	Brentford	3.000000	60
10	11	Wolves	4.000000	57
11	12	Newcastle Utd	3.000000	57
12	13	Brighton	4.000000	54
13	14	Crystal Palace	3.000000	54
14	15	Aston Villa	4.000000	48
15	16	Leeds United	3.000000	42
16	17	Everton	3.000000	39
17	18	Burnley	2.000000	6

Rk	Squad	MP	W	D	L	GF	GA	GD	Pts
1	Manchester City	38	29	6	3	99	26	+73	93
2	Liverpool	38	28	8	2	94	26	+68	92
3	Chelsea	38	21	11	6	76	33	+43	74
4	Tottenham	38	22	5	11	69	40	+29	71
5	Arsenal	38	22	3	13	61	48	+13	69
6	Manchester Utd	38	16	10	12	57	57	0	58
7	West Ham	38	16	8	14	60	51	+9	56
8	Leicester City	38	14	10	14	62	59	+3	52
9	Brighton	38	12	15	11	42	44	-2	51
10	Wolves	38	15	6	17	38	43	-5	51
11	Newcastle Utd	38	13	10	15	44	62	-18	49
12	Crystal Palace	38	11	15	12	50	46	+4	48
13	Brentford	38	13	7	18	48	56	-8	46
14	Aston Villa	38	13	6	19	52	54	-2	45
15	Southampton	38	9	13	16	43	67	-24	40
16	Everton	38	11	6	21	43	66	-23	39
17	Leeds United	38	9	11	18	42	79	-37	38
18	Burnley	38	7	14	17	34	53	-19	35
19	Watford	38	6	5	27	34	77	-43	23
20	Norwich City	38	5	7	26	23	84	-61	22

(a) Klasemen Berdasarkan Performa Model 1

(b) Klasemen asli Sumber: FBREF
(<https://fbref.com/en/comps/9/2021-2022/2021-2022-Premier-League-Stats>)

Gambar 4.1. Klasemen Berdasarkan uji Performa Model 1 dan Klasemen Asli Liga Inggris 2021-2022

4.6 Prediksi Tahun 2022-2023

Penelitian ini melakukan prediksi hasil pertandingan (Goal dan Kemasukan) untuk setiap tim di setiap pertandingan pada Liga Inggris musim 2022-2023. Tabel 4.16 menunjukkan 10 data pertama dari hasil prediksi yang telah dilakukan. Total terdapat 272 baris di dataframe hasil prediksi tersebut.

Tabel 4.17 merupakan hasil prediksi klasemen Liga Inggris musim 2022-2023 dengan menggunakan data statistik pemain di Liga Inggris musim 2021-2022. Prediksi dilakukan menggunakan model yang dibangun berdasarkan data tanpa goalCreation. Prediksi ini memiliki ketelitian sebesar

Tabel 4.16. Hasil prediksi goal dan kemasukan di Liga Inggris tahun 2022-2023

KlasemenTeam	goal	Kemasukan	KlasemenTimMusuh
1	2.355977	1.157862	2
1	2.379541	1.113089	3
1	2.403106	1.068316	4
1	2.426670	1.023543	5
1	2.450235	0.978770	6
1	2.473799	0.933997	7
1	2.497363	0.889224	8
1	2.520928	0.844451	9
1	2.544492	0.799678	10
1	2.568056	0.754905	11

0.4140625. Hal ini menunjukkan bahwa model yang dibuat tidak cukup baik untuk melakukan prediksi hasil pertandingan. Klasemen Liga Inggris dibuat berdasarkan total point yang didapatkan tim di setiap pertandingan. Tim yang memenangi pertandingan akan mendapatkan 3 point, tim yang seri akan mendapatkan 1 point, sedangkan tim yang kalah akan mendapatkan 0 point. Jika terdapat kondisi di mana terdapat dua tim atau lebih memiliki point yang sama, maka klasemen dilihat berdasarkan *goal difference* atau goal dikurangi kemasukan yang dialami tim selama satu musim.

Tabel 4.17. Hasil prediksi Klasemen 2023

Standings	Team	Goal Difference	Points
1	Manchester City	2.223176	48
2	Liverpool	2.098790	48
3	Chelsea	1.444492	48
4	Tottenham	1.086861	45
5	Arsenal	0.669405	27
6	West Ham	0.556949	27
7	Crystal Palace	0.454143	18
8	Leicester City	0.395457	18
9	Manchester Utd	0.305720	15
10	Aston Villa	0.285458	15
11	Brighton	0.303442	12
12	Wolves	0.213704	12
13	Newcastle Utd	-0.131998	0
14	Everton	-0.270770	0
15	Southampton	-0.279102	0
16	Brentford	-0.293313	0
17	Leeds United	-0.642788	0





BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini mengusulkan penggunaan model regresi polinomial berganda kuadratik untuk memprediksi hasil pertandingan sepakbola liga inggris musim 2022-2023. Data yang digunakan diambil dari data terbuka yang tersedia dari situs FBREF. Pada tahap pengolahan data didapatkan bahwa data yang digunakan berdimensi tinggi, sehingga penelitian ini melakukan pengurangan fitur dengan hanya menggunakan fitur yang sangat berpengaruh terhadap target. Pemilihan fitur yang sesuai dilakukan dengan menghitung nilai korelasi dengan menggunakan metode Korelasi *Pearson*. Hasil tabel korelasi tersebut menunjukkan bahwa fitur-fitur yang sangat berpengaruh terhadap banyaknya goal yang dicetak di setiap pertandingan yaitu fitur banyaknya operan berhasil yang dilakukan oleh pemain ke-5 (passComplete5) dikalikan fitur upaya melakukan serangan yang menghasilkan tendangan mengarah ke gawang yang dilakukan oleh pemain ke-6 (shotCreation6), upaya percobaan melakukan operan yang dilakukan oleh pemain ke-5 (passAttempt5) dikalikan fitur shotCreation6, fitur banyaknya bola yang didapatkan pemain ke-5 (touches5) dikalikan fitur shotCreation6, fitur operan berhasil yang dilakukan oleh pemain ke-5 (passComplete5) dikalikan fitur menit bermain dari pemain ke-9 (menit9) dan fitur pemain ke-5 melakukan *dribble* bola (carries5) dikalikan fitur shotCreation6. Untuk memprediksi kemasukan goal tim tersebut di pertandingan, fitur yang baik digunakan ialah fitur upaya tackle yang dilakukan oleh pemain ke-1 (tackle1) dikalikan fitur dimana pertandingan dilaksanakan (HomeAway1), fitur upaya pemain ke-2 menahan tendangan dari musuh (blocks2) dikalikan fitur kartu kuning yang didapatkan pemain ke-8 (yellowCard8), fitur upaya pemain ke-2 melakukan pressing (press2) dikalikan fitur upaya pemain ke-9 menahan tendangan dari musuh (blocks9), fitur tackle1 dikalikan fitur HomeAway9, Dan fitur tackle1 dikalikan fitur HomeAway8. Evaluasi model dilakukan dengan menggunakan metrik R2, MAE serta dilakukan uji model dengan melihat keakuratan dalam memprediksi hasil pertandingan sepak bola. Hasil evaluasi juga dibandingkan dengan model regresi linear sebagai *baseline*. Hasil evaluasi performa model menunjukkan bahwa performa model regresi polinomial berganda kuadratik memiliki akurasi sedikit lebih tinggi sebesar 0.13 dibandingkan dengan model regresi linear untuk dataset yang tidak menggunakan fitur goalCreation. Namun, performa model regresi linear sedikit lebih tinggi sebesar 0,09 untuk pemakaian di dataset yang menggunakan fitur goalCreation. Prediksi hasil pertandingan liga inggris musim 2022-2023 dengan menggunakan data performa rata-rata dari pemain di liga inggris musim 2021-2022 memiliki ketelitian yang rendah yaitu 0.4140625.

5.2 Saran

Hasil dari penelitian ini berupa fitur apa saja yang baik untuk digunakan memprediksi hasil pertandingan sepak bola, serta melihat performa dari penggunaan regresi polinomial berganda untuk melakukan prediksi hasil pertandingan sepak bola. Hasil prediksi menggunakan model pada penelitian ini akan lebih baik jika data yang digunakan untuk melakukan prediksi adalah data rata-rata setiap pemain yang akan memainkan pertandingan tersebut.

1. Dilarang mengutip karya tulis ini, kecuali:
 - a. menyebutkan sumber sesuai kaidah kecependekiaan;
 - b. pengutipan hanya untuk keperluan pendidikan, penulisan karya ilmiah atau penelitian;
 - c. pengutipan tidak merugikan Universitas Pertamina.
2. Dilarang mempublikasikan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa adanya izin dari Universitas Pertamina.



DAFTAR PUSTAKA

- Ahmad, S., Wasim, S., Irfan, S., Gogoi, S., Srivastava, A., and Farheen, Z. (2019). Qualitative v/s. quantitative research-a summarized review. *Journal of Evidence-Based Medicine and Healthcare*, 6(43):2828–2832.
- Chakraborty and Niladri (2022). How football clubs make money & earn profits? know how football clubs meet their budget & earn profit. kreedon.com/how-football-clubs-make-money-and-earn-profit/?amp. Diakses pada 19-11-2021.
- Cho, I. (2018). Moneyball for business: Becoming an analytics-driven business. <https://medium.com/nexus-analytics/moneyball-for-business-becoming-an-analytically-driven-business-60093799fc03>. Diakses pada 19-11-2021.
- Detle, H. and Studden, W. J. (1997). *The theory of canonical moments with applications in statistics, probability, and analysis*, volume 338. John Wiley & Sons.
- Gough, C. (2022). Total sports market revenue worldwide 2021-2026. <https://www.statista.com/statistics/370560/worldwide-sports-market-revenue/>. Diakses pada 18-11-2022.
- Hasibuan, P. D. and Zainal, A. (2007). Metodologi penelitian pada bidang ilmu komputer dan teknologi informasi; konsep, teknik, dan aplikasi.
- Hiran, K. K., Jain, R. K., Lakhwani, K., and Doshi, R. (2021). *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition)*. United kingdom: BPB Publications.
- Imran, H., Al-Abdaly, N. M., Shamsa, M. H., Shatnawi, A., Ibrahim, M., and Ostrowski, K. A. (2022). Development of prediction model to predict the compressive strength of eco-friendly concrete using multivariate polynomial regression combined with stepwise method. *Materials*, 15(1):317.
- Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4):531–538.
- Kaplan, D. (2022). Multivariate polynomial regression python (full code). <https://enjoymachinelearning.com/blog/multivariate-polynomial-regression-python/>. Diakses pada 21-11-2021.
- Kumar, S. and Chong, I. (2018). Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. *International journal of environmental research and public health*, 15(12):2907.
- Kurniawan, D. (2022). *Pengenalan Machine Learning dengan Python*. Jakarta: Elex Media Komputindo.
- Nourse and Garret (2021). A machine learning algorithm for predicting outcomes of mlb games. <https://towardsdatascience.com/a-machine-learning-algorithm-for-predicting-outcomes-of-mlb-games-fa17710f3c04>. Diakses pada 18-11-2021.
- Prabowo, D. B. (2020). Prediksi hasil pertandingan sepakbola english premier league dengan menggunakan algoritma k-nearest neighbors dan naïve bayes classifier.
- Rebala, G., Ravi, A., and Churiwala, S. (2019). *An introduction to machine learning*. USA: Springer.

Smith, C. (2022). Premier league winners: 1992-2022. <https://soccerprime.com/premier-league-winners/>. Diakses pada 19-11-2021.

Spacey, A. (2022). 10 reasons why the EPL is the world's best. <https://howtheyplay.com/team-sports/10-Reasons-Why-All-The-World-Loves-The-EPL>. Diakses pada 06-03-2023.

Veroutsos, E. (2022). The most popular sports in the world. <https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html>. Diakses pada 18-11-2021.

Young, B. and Starkes, J. (2005). Career-span analyses of track performance: Longitudinal data present a more optimistic view of age-related performance decline. *Experimental aging research*, 31:69–90.

1. Dilarang mengutip karya tulis ini, kecuali:

- a. menyebutkan sumber sesuai kaidah kecependekiaan;
- b. pengutipan hanya untuk keperluan pendidikan, penulisan karya ilmiah atau penelitian;
- c. pengutipan tidak merugikan Universitas Pertamina.

2. Dilarang mempublikasikan atau memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa adanya izin dari Universitas Pertamina.



Universitas
Pertamina

LAMPIRAN A

Pengumpulan Data

Pengumpulan dataset 1

```
1  #@title
2  import requests
3  df = pd.DataFrame()
4  dictionary_pemain = {}
5  list_nama = []
6
7  url = 'https://fbref.com/en/players/8f62b6ee/matchlogs/' + \
8        '2021-2022/summary/Hugo-Lloris-Match-Logs'
9  html = requests.get(url).content
10 df_list = pd.read_html(html)
11 df = df_list[-1]
12 #print(df)
13 df.to_csv('my data.csv')
14 nama = url[66:-11]
15 dictionary_pemain[nama] = df
16 list_nama.append(nama)
17 print(nama)
18 dictionary_pemain[nama]['Nama']=nama
19 df2=dictionary_pemain[nama]
```

Pengumpulan dataset 2

```
1  #@title
2  url = 'https://fbref.com/en/players/32f60ed8/matchlogs/' + \
3        '2021-2022/summary/Morgan-Gibbs-White-Match-Logs'
4  html = requests.get(url).content
5  df_list = pd.read_html(html)
6  df = df_list[-1]
7  #print(df)
8  df.to_csv('my data.csv')
9  nama = url[66:-11]
10 dictionary_pemain[nama] = df
11 list_nama.append(nama)
12 print(nama)
13 dictionary_pemain[nama]['Nama']=nama
14 df2=pd.concat([df2,dictionary_pemain[nama]])
15 dictionary_pemain[nama]
16 df2
```



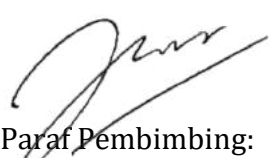

Form TA-2 Bimbingan Tugas Akhir
FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER

Nama Mahasiswa : Rifqi Julian Hasyari

NIM : 105218037

Nama Pembimbing : Rangga Ganzar Noegraha, Ph. D

NIP : 119005

No. 1	Hari/Tanggal: Kamis , 15 september 2022
<p>Hal yang menjadi perhatian:</p> <ul style="list-style-type: none">– Penjelasan mengenai topik penelitian– Dataframe yang digunakan <p style="text-align: right;"> Paraf Pembimbing:</p>	
No. 2	Hari/Tanggal: Senin, 26 september 2022
<p>Hal yang menjadi perhatian:</p> <ul style="list-style-type: none">– Pengumpulan Dataframe <p style="text-align: right;"> Paraf Pembimbing:</p>	



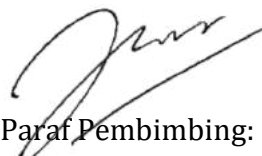
Form TA-2 Bimbingan Tugas Akhir
FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER

No. 3

Hari/Tanggal: Rabu, 5 Oktober 2022

Hal yang menjadi perhatian:

- Dataset yang telah dikumpulkan, perlu di olah Kembali
- Buat kolom baru yang berisikan goal dan kemasukan di tiap pertandingan



Paraf Pembimbing:

No. 4

Hari/Tanggal: Senin, 10 Oktober 2022

Hal yang menjadi perhatian:

- Melakukan Pre-Processing data
- Dicek lagi data-data yang double input
- Hapus data yang double input



Paraf Pembimbing:



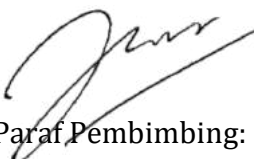
Form TA-2 Bimbingan Tugas Akhir
FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER

No. 5

Hari/Tanggal: Senin, 17 Oktober 2022

Hal yang menjadi perhatian:

- Melakukan Melakukan Pre-Processing data
- Cari semua data pemain yang double secara manual
- Kirimkan Draft bab 1



Paraf Pembimbing:

No. 6

Hari/Tanggal : Jumat, 21 Oktober 2022

Hal yang menjadi perhatian:

- Data pemain yang double telah dihapus.
- Penjelasan mengenai Regresi polynomial berganda untuk melakukan prediksi performa sepak bola
- Revisi laporan



Paraf Pembimbing:



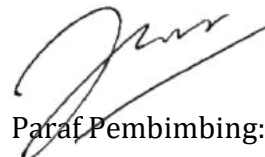
Form TA-2 Bimbingan Tugas Akhir
FAKULTAS SAINS DAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER

No. 7

Hari/Tanggal: Rabu, 2 November 2022

Hal yang menjadi perhatian:

- Melakukan pre-processing data menjadi sebuah numerik



Paraf Pembimbing:

No. 8

Hari/Tanggal: Jumat, 10 November 2022

Hal yang menjadi perhatian:

- Penulisan laporan untuk seminar kemajuan



Paraf Pembimbing:



Form TA-2 Bimbingan Tugas Akhir

FAKULTAS SAINS DAN ILMU KOMPUTER



PROGRAM STUDI ILMU KOMPUTER

Nama Mahasiswa : Rifqi Julian Hasyari

NIM :105218037

Nama Pembimbing : Ade Irawan, Ph.D

NIP : 116130

No.1	Hari/Tanggal: 1 december 2022
<p>Hal yang menjadi perhatian:</p> <ul style="list-style-type: none">– Terdapat beberapa hal yang perlu direvisi dalam laporan untuk seminar kemajuan.– Terdapat beberapa hal tentang topik yang perlu dipahami lagi. <p style="text-align: right;"> Paraf Pembimbing:</p>	
No.2	Hari/Tanggal: 5 december 2022
<p>Hal yang menjadi perhatian:</p> <ul style="list-style-type: none">– Tambahkan tabel 3c2s.– Revisi laporan untuk seminar kemajuan <p style="text-align: right;"> Paraf Pembimbing:</p>	



Form TA-2 Bimbingan Tugas Akhir

FAKULTAS SAINS DAN ILMU KOMPUTER

PROGRAM STUDI ILMU KOMPUTER

No.3

Hari/Tanggal: 12 december 2022

Hal yang menjadi perhatian:

- Finalisasi laporan untuk seminar kemajuan.

Paraf Pembimbing:

No.4

Hari/Tanggal: 1 februari 2023

Hal yang menjadi perhatian:

- Progres dari penelitian yang telah dilakukan.
- Perlu ditambahkan model pembanding dari metode yang digunakan.

Paraf Pembimbing:



Form TA-2 Bimbingan Tugas Akhir

FAKULTAS SAINS DAN ILMU KOMPUTER

PROGRAM STUDI ILMU KOMPUTER

No.5

Hari/Tanggal: 2 februari 2023

Hal yang menjadi perhatian:

- Cara pembuatan tabel yang baik di latex overleaf.
- Penulisan referensi perlu dicek Kembali.

Paraf Pembimbing:

No.6

Hari/Tanggal: 3 februari

Hal yang menjadi perhatian:

- Penulisan lampiran kode python dalam laporan dengan menggunakan latex.

Paraf Pembimbing:



Form TA-2 Bimbingan Tugas Akhir

FAKULTAS SAINS DAN ILMU KOMPUTER

PROGRAM STUDI ILMU KOMPUTER

No.7

Hari/Tanggal: 6 februari 2023

Hal yang menjadi perhatian:

- Revisi bab 2 pada laporan tugas akhir.
- Tambahkan penjelasan tentang korelasi untuk pemilihan fitur.

Paraf Pembimbing:

No. 8

Hari/Tanggal: Kamis , 9 Februari 2023

Hal yang menjadi perhatian:

- Penulisan bab 4 hasil dan pembahasan untuk laporan tugas akhir

Paraf Pembimbing:



Form TA-2 Bimbingan Tugas Akhir

FAKULTAS SAINS DAN ILMU KOMPUTER

PROGRAM STUDI ILMU KOMPUTER

No. 9

Hari/Tanggal: 10 februari 2023

Hal yang menjadi perhatian:

- Pengiriman draft laporan
- Pembuatan abstract

Paraf Pembimbing:

SURAT KETERANGAN PERBAIKAN

Yang bertanda tangan dibawah ini menerangkan bahwa :

NIM : 105218037

Nama : Rifqi Julian Hasyari

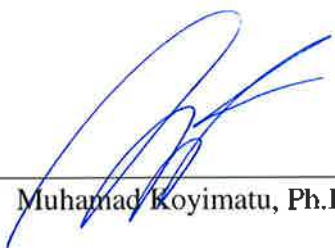

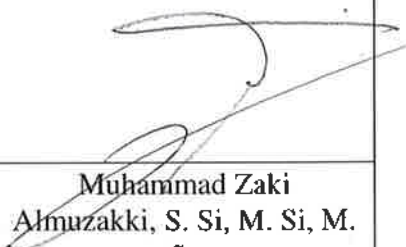
Fakultas/Prodi : Sains dan Ilmu Komputer / Ilmu Komputer

Judul Skripsi : APAKAH REGRESI POLINOMIAL BERGANDA KUADRATIK TEPAT
UNTUK ANALISIS PERFORMA TIM SEPAK BOLA? STUDI KASUS LIGA INGGRI
2022-2023.


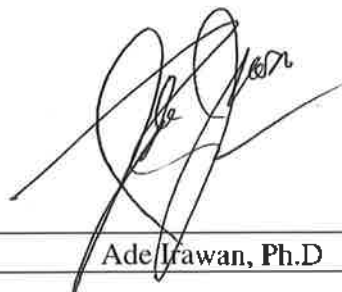
Mahasiswa tersebut diatas telah menyelesaikan perbaikan naskah Tugas Akhir sesuai petunjuk perbaikan berita acara dan arahan dosen penguji dan pembimbing.

Demikian surat keterangan ini dibuat untuk diperhatikan dan dilaksanakan.

Jakarta, 13 Maret 2023

Penguji 1	Penguji 2	Penguji 3
		
Muhammad Koyimatu, Ph.D	Dr. Ariana Yunita, M.I.T., MBA	Muhammad Zaki Almuzakki, S. Si, M. Si, M. Sc

Mengelahui

Pembimbing 1	Pembimbing 2
	
Rangga Ganzar Noegraha, Ph. D	Ade Irawan, Ph.D