

**PREDIKSI KECEPATAN RATA-RATA BERSEPEDA  
BERDASARKAN KONDISI TOPOGRAFI DAN FAKTOR  
CUACA MENGGUNAKAN *XGBOOST* DARI DATA STRAVA**

**LAPORAN TUGAS AKHIR**

Laporan ini disusun untuk memenuhi salah satu syarat memperoleh  
Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika  
Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



**Disusun Oleh :**  
**Rifqy Ramdhani Hakim**  
**NIM 32602100111**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG  
2025**

***FINAL PROJECT***  
***PREDICTION OF AVERAGE CYCLING SPEED BASED***  
***ON TOPOGRAPHIC CONDITIONS AND WEATHER***  
***FACTORS USING XGBOOST FROM STRAVA DATA***

*Proposed to complete the requirement to obtain a bachelor's degree (SI)*  
*at Informatic Engineering Department of Industrial Technology Faculty*  
*Sultan Agung Islamic University*



Arranged By :  
**Rifqy Ramdhani Hakim**  
**NIM 32602100111**

**MAJORING OF INFORMATIC ENGINEERING**  
**FACULTY OF INDUSTRIAL TECHNOLOGY**  
**SULTAN AGUNG ISLAMIC UNIVERSITY**  
**SEMARANG**  
**2025**

**LEMBAR PENGESAHAN  
TUGAS AKHIR**

**PREDIKSI KECEPATAN RATA-RATA BERSEPEDA BERDASARKAN  
KONDISI TOPOGRAFI DAN FAKTOR CUACA MENGGUNAKAN  
XGBOOST DARI DATA STRAVA**

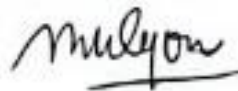
**Rifqy Ramdhani Hakim  
NIM 32602100111**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir  
Program Studi Teknik Informatika  
Universitas Islam Sultan Agung  
Pada tanggal : 2 September 2025

**TIM PENGUJI UJIAN SARJANA :**

**Ir. Sri Mulyono, M.Eng**  
NIK.  
(210616049)

Tanda Tangan



Tanggal

2/9  
2025

**Moch Taufik MIT**  
NIK.  
(210604034)

Tanda Tangan



Tanggal

2/9  
2025

**SAM FARISA CHAERUL**  
**HAVIANA, ST., M.KOM**  
NIK. 210615046  
(Pembimbing)

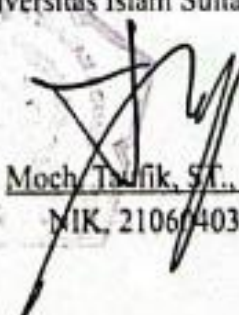
Tanda Tangan



Tanggal

2/9  
2025

Semarang, 2 September 2025  
Mengetahui,  
Kaprodi Teknik Informatika  
Universitas Islam Sultan Agung

  
**Moch Taufik, ST., MIT**  
NIK. 210604034

## **SURAT PERNYATAAN KEASLIAN TUGAS AKHIR**

Yang bertanda tangan dibawah ini :

Nama : Rifqy Ramdhani Hakim

NIM : 32602100111

Judul Tugas Akhir : PREDIKSI KECEPATAN RATA-RATA BERSEPEDA  
BERDASARKAN KONDISI TOPOGRAFI DAN FAKTOR CUACA  
MENGUNAKAN *XGBOOST* DARI DATA STRAVA

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 2 September 2025

Yang Menyatakan,



Rifqy Ramdhani Hakim

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Rifqy Ramdhani Hakim

NIM : 32602100111

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : Semarang

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : Prediksi Kecepatan Rata-rata Bersepeda Berdasarkan Kondisi Topografi dan Faktor Cuaca Menggunakan *XGBoost* Dari Data Starava. Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan di internet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung.

Semarang, 2 September 2025

Yang menyatakan,



Rifqy Ramdhani Hakim



## KATA PENGANTAR

Puji syukur penulis panjatkan kepada ALLAH SWT, yang telah memberikan rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Prediksi Kecepatan Rata-rata Bersepeda Berdasarkan Kondisi Topografi dan Faktor Cuaca Menggunakan *XGBoost* Dari Data Starava” ini dengan baik. Dengan penuh rasa hormat, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T., M.T
3. Dosen pembimbing penulis Bapak Sam Farisa Chaerul Haviana, ST., M.kom yang telah memberikan arahan, bimbingan, dan saran yang berarti dalam penyelesaian tugas akhir ini.
4. Seluruh dosen Program Studi Teknik Informatika, Fakultas Teknologi Industri UNISSULA yang telah memberikan ilmunya kepada penulis.
5. Orang tua penulis, Bapak Lukman Hakim dan Ibu Muflikha yang selalu memberikan segala doa, dukungan, dan motivasi dengan penuh limpahan kasih sayang sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.
6. Untuk Diri Sendiri yang tidak pernah menyerah dan berhenti mencoba sesulit apapun rintangan kuliah dan selama proses penyusunan laporan ini.
7. Dan kepada seluruh pihak yang tidak dapat saya sebutkan satu persatu.

Semarang, 2 September 2025

Rifqy Ramdhani Hakim

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN TUGAS AKHIR.....</b>	<b>iii</b>
<b>SURAT PERNYATAAN KEASLIAN TUGAS AKHIR .....</b>	<b>iv</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....</b>	<b>v</b>
<b>KATA PENGANTAR .....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR GAMBAR .....</b>	<b>x</b>
<b>DAFTAR TABEL.....</b>	<b>xi</b>
<b>ABSTRAK.....</b>	<b>xii</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1    Latar Belakang.....	1
1.2    Perumusan Masalah.....	2
1.3    Pembatasan Masalah .....	2
1.4    Tujuan.....	3
1.5    Manfaat.....	3
1.6    Sistematika Penulisan.....	3
<b>BAB II    TINJAUAN PUSTAKA DAN DASAR TEORI .....</b>	<b>5</b>
2.1    Tinjauan Pustaka .....	5
2.2    Dasar Teori.....	14
2.2.1    Sepeda .....	14
2.2.2    Perfoma Kecepatan .....	14
2.2.3    Topografi dan Cuaca .....	15
2.2.4    Xtreme Gradient Boosting ( <i>XGBoost</i> ) .....	16
2.2.5    Strava.....	17
2.2.6    Waktu Tidur .....	18
<b>BAB III METODE PENELITIAN .....</b>	<b>19</b>
3.1    Metode Penelitian.....	19
3.1.1    Studi Literatur .....	19
3.1.2    Pengumpulan Dataset.....	20

3.1.3	Pre Processing .....	21
3.1.4	Data Cleaning.....	21
3.1.5	Normalisasi Data.....	22
3.1.6	Split Data.....	23
3.1.7	Pembuatan <i>XGBoost</i> .....	23
3.1.8	<i>Evaluasi Model</i> .....	24
3.1.9	Halaman Awal Sistem Web .....	27
3.1.10	Visualisasi Hasil.....	29
3.1.11	Software software yang digunakan .....	30
3.1.12	Library yang digunakan .....	31
<b>BAB IV HASIL DAN ANALISIS PENELITIAN .....</b>		<b>34</b>
4.1	Hasil Penelitian.....	34
4.1.1	Gambaran Umum Dataset.....	34
4.1.2	<i>Data Selection</i> .....	35
4.1.3	<i>Missing Value</i> .....	35
4.1.4	Mengecek <i>duplicate data</i> .....	36
4.2	Preprocessing Data.....	36
4.2.1.	<i>Data cleaning</i> .....	36
4.2.2.	Normalisasi Data.....	36
4.2.3.	Pembagian Data ( <i>Data Splitting</i> ).....	37
4.3	Pembuatan Model.....	37
4.4.1.	Pembuatan Model <i>XGBoost</i> .....	37
4.4.2.	Evaluasi model.....	40
4.4.1.	<i>Wireframe</i> Halaman Awal.....	42
4.4.2.	<i>Wireframe</i> Halaman Hasil Prediksi.....	43
4.4.3.	Hasil dan Pembahasan.....	45
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>48</b>
5.1	Kesimpulan.....	48
5.2	Saran.....	49
<b>DAFTAR PUSTAKA .....</b>		<b>50</b>
<b>HALAMAN LAMPIRAN.....</b>		<b>53</b>





## DAFTAR GAMBAR

Gambar 2. 1 Arsitektur <i>XGBoost</i> .....	16
Gambar 3. 1 Alur Penelitian.....	19
Gambar 3. 2 Alur Sistem.....	26
Gambar 3. 3 Awal Sistem .....	28
Gambar 3. 4 Visualisasi Hasil .....	29
Gambar 4. 1 Alur Model <i>XGBoost</i> .....	38
Gambar 4. 2 Perancangan Sistem.....	41
Gambar 4. 3 Halaman Awal.....	42
Gambar 4. 4 Halaman Hasil Prediksi .....	43
Gambar 4. 5 Penjelasan Parameter.....	44



## DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka .....	7
Tabel 3. 1 Dataset .....	20
Tabel 4. 1 Dataset .....	34
Tabel 4. 2 Splitting Data .....	37
Tabel 4. 3 Hyperparameter .....	39
Tabel 4. 4 Evaluasi Model .....	41
Tabel 4. 5 Hasil Prediksi .....	45



## ABSTRAK

Kegiatan bersepeda merupakan kegiatan atau aktivitas yang cukup banyak diminati masyarakat sebagai pilihan olahraga. Seiring meningkatnya minat terhadap gaya hidup sehat dan berbasis teknologi, aplikasi seperti Strava semakin banyak digunakan oleh pesepeda. Strava memungkinkan penggunaannya untuk merekam data aktivitas bersepeda secara rinci, seperti jarak tempuh, kecepatan, elevasi, serta waktu pelaksanaan. Salah satu algoritma yang dapat digunakan untuk memahami dan memprediksi hubungan antara kondisi lingkungan dengan performa bersepeda adalah *XGBoost* (*Extreme Gradient Boosting*). Dengan menggunakan algoritma ini, penelitian bertujuan untuk membangun model prediksi kecepatan rata-rata bersepeda berdasarkan data historis aktivitas dari Strava dan variabel lingkungan seperti topografi dan cuaca. Metode yang digunakan pada penelitian ini yaitu pengumpulan data, pengolahan data, data selection, pembuatan model dan pengujian. Sumber data diperoleh menggunakan data aktivitas bersepeda pribadi yang diperoleh melalui aplikasi Strava yang kemudian dikumpulkan dan di input kedalam excel. Hasil penelitian dan penerapan sistem prediksi kecepatan rata-rata saat bersepeda menggunakan algoritma *XGBoost* menunjukkan bahwa sistem ini mampu memperkirakan kecepatan rata-rata bersepeda dengan tingkat akurasi yang cukup baik. Kombinasi antara kondisi topografi dan faktor cuaca terbukti memberikan pengaruh signifikan terhadap performa pesepeda.

Kata Kunci: Kecepatan Bersepeda, Strava, Topografi, Cuaca, *XGBoost*

## ABSTRACT

Cycling is a popular activity for many people as a sport. With the growing interest in a healthy, technology-based lifestyle, applications like Strava are becoming increasingly popular among cyclists. Strava allows users to record detailed cycling activity data, such as distance traveled, speed, elevation, and time. One algorithm that can be used to understand and predict the relationship between environmental conditions and cycling performance is *XGBoost* (*Extreme Gradient Boosting*). Using this algorithm, this study aims to develop a prediction model for average cycling speed based on historical activity data from Strava and environmental variables such as topography and weather. The methods used in this study included data collection, data processing, data selection, model creation, and testing. The data source was personal cycling activity data obtained through the Strava application, which was then compiled and input into Excel. The results of the research and implementation of the average cycling speed prediction system using the *XGBoost* algorithm indicate that this system is capable of estimating average cycling speed with a fairly good level of accuracy. The combination of topography and weather factors has been shown to significantly influence cyclist performance.

Keywords: *Comfort Pace, Strava, Topography, Weather, XGBoost*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kegiatan bersepeda merupakan kegiatan atau aktifitas yang cukup banyak diminati masyarakat sebagai pilihan olahraga. Sepeda adalah alat transportasi yang sangat umum dan luas penggunaannya di dunia yang digunakan oleh semua orang dari berbagai kalangan usia. Tidak hanya sebagai alat transportasi sepeda juga menjadi salah satu hobi yang banyak diminati oleh masyarakat dan sudah merupakan gaya hidup bagi sebagian masyarakat kota (Utomo 2016). Saat ini olahraga sepeda menjadi sebuah kebutuhan utama dari setiap orang, bukan hanya untuk prestasi, namun juga untuk kepentingan kesehatan, untuk kepentingan hiburan atau rekreasi, untuk kepentingan mencari relasi dan berbagai kepentingan lainnya. (Surojo, dkk. 2022).

Seiring meningkatnya minat terhadap gaya hidup sehat dan berbasis teknologi, aplikasi seperti Strava semakin banyak digunakan oleh pesepeda. Strava memungkinkan penggunanya untuk merekam data aktivitas bersepeda secara rinci, seperti jarak tempuh, kecepatan, elevasi, serta waktu pelaksanaan. Data ini membuka peluang besar untuk dianalisis secara ilmiah guna memahami faktor-faktor yang memengaruhi performa bersepeda (Rupaka, dkk. 2021).

Salah satu indikator performa pesepeda yang umum digunakan adalah kecepatan rata-rata. Nilai ini dapat dipengaruhi oleh berbagai faktor, di antaranya adalah kondisi topografi lintasan seperti elevasi dan kemiringan, serta cuaca, seperti suhu dan kelembapan udara. Di samping itu, terdapat pula faktor-faktor non-teknis seperti kualitas tidur, kondisi tubuh, dan permukaan jalan yang basah setelah hujan. Meskipun faktor non-teknis ini tidak tercatat secara eksplisit dalam data, pemahaman terhadapnya penting untuk memberi konteks tambahan terhadap hasil analisis performa (Widodo dkk. 2023).

Untuk memahami dan memprediksi hubungan antara kondisi lingkungan dengan performa bersepeda, diperlukan pendekatan komputasional berbasis pembelajaran mesin (machine learning). Salah satu algoritma yang dapat digunakan adalah *XGBoost* (Extreme Gradient Boosting), yaitu metode prediksi berbasis

pohon keputusan yang terkenal efektif dalam memodelkan hubungan non-linear serta menghasilkan akurasi tinggi. Dengan menggunakan algoritma ini, penelitian bertujuan untuk membangun model prediksi kecepatan rata-rata bersepeda berdasarkan data historis aktivitas dari Strava dan variabel lingkungan seperti topografi dan cuaca

Urgensi penelitian ini terletak pada pentingnya memahami bagaimana faktor-faktor seperti elevasi, kondisi cuaca, dan kualitas waktu tidur dapat memengaruhi kecepatan rata-rata bersepeda. Elevasi dan kemiringan jalur berhubungan langsung dengan tingkat kesulitan lintasan yang dapat memperlambat atau mempercepat kayuhan. Faktor cuaca, seperti hujan atau suhu udara, dapat memengaruhi kenyamanan, keamanan, serta efisiensi tenaga pesepeda. Sementara itu, kualitas waktu tidur yang baik berkontribusi terhadap kesiapan fisik dan daya tahan tubuh, yang pada akhirnya memengaruhi performa saat bersepeda. Oleh karena itu, analisis mendalam terhadap variabel-variabel tersebut tidak hanya penting untuk menghasilkan prediksi yang akurat, tetapi juga bermanfaat bagi pesepeda dalam merencanakan strategi latihan, mengatur pola tidur, serta menyesuaikan aktivitas dengan kondisi lingkungan untuk mencapai performa optimal.

## **1.2 Perumusan Masalah**

Penggunaan data pribadi dari aktivitas bersepeda melalui Strava menimbulkan keterbatasan dalam memperoleh gambaran umum mengenai pengaruh topografi, cuaca, dan waktu tidur. faktor individu terhadap kecepatan bersepeda. Kondisi ini menimbulkan permasalahan mengenai bagaimana merancang sistem prediksi kecepatan rata-rata bersepeda yang mampu mengintegrasikan faktor-faktor tersebut menggunakan algoritma *XGBoost* guna memberikan estimasi yang lebih akurat?"

## **1.3 Pembatasan Masalah**

Untuk menjaga fokus dan ruang lingkup penelitian, terdapat beberapa Batasan yang ditetapkan, yaitu :

1. Data yang digunakan bersumber dari aktivitas bersepeda pribadi yang direkam menggunakan aplikasi Strava



2. Variabel yang digunakan terbatas pada data yang tersedia yaitu, Jam mulai bersepeda, total durasi, elevasi, jarak tempuh, kecepatan maksimum
3. Penelitian ini hanya berfokus pada prediksi kecepatan rata-rata, bukan kecepatan instan atau performa (perkilometer, per elevasi)

#### **1.4 Tujuan**

Tujuan penelitian ini adalah untuk mengimplementasikan model *XGBoost* dalam memprediksi Kecepatan Rata-rata Bersepeda Berdasarkan Kondisi Topografi, Faktor Cuaca dan Waktu Tidur

#### **1.5 Manfaat**

Penelitian ini memberikan manfaat berupa gambaran estimasi kecepatan rata-rata bersepeda berdasarkan kondisi topografi, cuaca, dan kualitas tidur sehingga dapat membantu pesepeda dalam merencanakan strategi bersepeda yang lebih efisien dan aman. Dari sisi akademik, penelitian ini berkontribusi pada penerapan algoritma *XGBoost* dalam bidang olahraga dan kesehatan, sekaligus membuka peluang untuk diintegrasikan dengan aplikasi seperti Strava guna menghadirkan fitur prediksi performa. Selain itu, penelitian ini juga mendukung inovasi teknologi berbasis machine learning dalam pengembangan olahraga baik untuk tujuan rekreasi maupun profesional.

#### **1.6 Sistematika Penulisan**

##### **BAB I : PENDAHULUAN**

Pada BAB I menjelaskan tentang latar belakang, pemilihan judul, rumusan masalah, Batasan masalah, tujuan penelitian, metodologi penelitian dan sistematika penelitian.

##### **BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI**

Pada BAB II memuat tentang penelitian terdahulu dan landasan teori yang berkaitan untuk membantu memahami konsep algoritma *XGBoost* dalam memprediksi kecepatan rata-rata bersepeda.

##### **BAB III : METODE PENELITIAN**

Pada BAB III menjelaskan proses penelitian yang dimulai dari pengumpulan data hingga evaluasi hasil prediksi.

##### **BAB IV : HASIL DAN ANALISIS PENELITIAN**

Pada BAB IV berisi tentang pemaparan hasil penelitian yang dimulai dari pembuatan sistem sampai dengan proses deployment

## BAB V : KESIMPULAN DAN SARAN

Pada BAB V merangkum keseluruhan proses penelitian dari awal sampai akhir



## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Pada Penelitian Prediksi Angka Harapan Hidup Penduduk Menggunakan Model *XGBoost* menunjukkan performa prediksi yang sangat baik dengan tingkat akurasi mencapai 96,8%. Nilai *Mean Absolute Error (MAE)* sebesar 0,97 mengindikasikan bahwa rata-rata selisih antara hasil prediksi dan data aktual relatif kecil. Hasil ini menegaskan bahwa *XGBoost* Regressor dapat digunakan secara efektif untuk memperkirakan angka harapan hidup di negara-negara kawasan Asia. Temuan tersebut memiliki implikasi penting, khususnya dalam mendukung pengambilan keputusan berbasis data serta perencanaan kebijakan di bidang kesejahteraan masyarakat. (Kurniawan dkk. 2024).

Penelitian Pemodelan ritme pembakaran kalori selama aktivitas bersepeda menggunakan *Feedforward Neural Network (FFNN)* menghasilkan konfigurasi optimal dengan 59 neuron pada lapisan tersembunyi. Model ini mampu memprediksi jumlah kalori yang terbakar (*Calt*) dengan tingkat error sekitar 7%. Hasil pengujian juga menunjukkan bahwa suhu memiliki pengaruh positif, di mana semakin tinggi suhu maka kalori yang terbakar juga meningkat. Secara keseluruhan, rancangan FFNN dalam penelitian ini dapat dikatakan cukup efektif dengan tingkat akurasi mencapai 93% dalam memprediksi kalori terbakar per detik. (Ichwan dkk. 2024).

Penelitian Sebelumnya Penerapan Algoritma *Extreme Gradient Boosting (XGBoost)* untuk Analisis Risiko Kredit Hasil analisis menggunakan algoritma *XGBoost* menunjukkan bahwa metode ini cukup efektif dalam melakukan klasifikasi pada data kredit. Pada tahap data mining, dilakukan pembangunan 10 model untuk memperoleh performa terbaik. Sebelum proses pemodelan, data terlebih dahulu diseimbangkan dengan *SMOTE* untuk mengatasi masalah ketidakseimbangan kelas. Dari hasil evaluasi diperoleh bahwa algoritma *XGBoost* dapat dijadikan pendekatan untuk membantu proses pengambilan keputusan dalam menerima atau menolak pengajuan kredit. Evaluasi juga memperlihatkan bahwa kinerja model meningkat setelah penerapan *SMOTE*, yang ditunjukkan dengan

perbaikan nilai akurasi maupun *AUC*. Model terbaik diperoleh pada skenario pertama (90% data latih dan 10% data uji) dengan capaian akurasi sebesar 0,83 dan *AUC* 0,918.(Saputra dkk. 2024).

Berdasarkan hasil penelitian, dapat disimpulkan bahwa minat masyarakat terhadap olahraga bersepeda tergolong sangat tinggi. Hasil observasi menunjukkan bahwa bersepeda bukan hanya menjadi tren, tetapi juga dipandang sebagai pilihan olahraga yang tepat dan sesuai untuk berbagai kalangan. Dengan demikian, bersepeda dapat dikategorikan sebagai aktivitas olahraga yang populer serta layak direkomendasikan bagi masyarakat secara luas. (Irawati, dkk. 2024).

Penelitian ini memanfaatkan *XGBoost* untuk prediksi penyakit stroke dan menghasilkan performa yang baik (akurasi 95,4%, precision 94,3%, recall 96,6%, F1-score 95,4%, *AUC* 95,4%). Meski hasilnya tinggi, performanya masih di bawah model Stacking dan Random Forest dari penelitian sebelumnya. Teknik *SMOTE* digunakan untuk mengatasi data imbalance, dan Bayesian Optimization untuk tuning hyperparameter. Peneliti menyarankan optimasi lebih lanjut karena ruang pencarian hyperparameter masih luas, membuka peluang peningkatan di penelitian selanjutnya. (Murdiansyah 2024).

Penelitian ini menghasilkan model prediksi diabetes berbasis algoritma *XGBoost* dengan performa yang sangat baik, ditunjukkan oleh akurasi sebesar 96,33% dan nilai recall 97,16% pada dataset *Pima Indians*. Berdasarkan analisis confusion matrix, model hanya menghasilkan 6 kasus false negative (1,4%) dan 16 kasus false positive (3,8%). Selain itu, nilai ROC-*AUC* sebesar 0,99 membuktikan kemampuan diskriminasi model yang mendekati sempurna. Hasil penelitian juga memperlihatkan bahwa *XGBoost* memiliki kinerja lebih unggul dibanding algoritma lain setelah dilakukan optimasi *hyperparameter*. Meski demikian, terdapat beberapa aspek yang perlu diperhatikan untuk pengembangan lebih lanjut, di antaranya adalah validasi eksternal pada populasi yang lebih beragam, pengurangan jumlah *false positive*, serta integrasi model ke dalam antarmuka klinis yang ramah pengguna. Dengan potensi yang dimiliki, model ini berpeluang besar menjadi alat bantu dalam pengambilan keputusan medis, namun implementasi

klinis tetap membutuhkan pendekatan bertahap dengan memastikan faktor keamanan pasien dan validasi menyeluruh.(Susanto dkk 2025)

Tabel 2. 1 Tinjauan Pustaka

Judul	Metode	Hasil Penelitian
Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode <i>XGBoost</i>	<i>XGBoost</i> (eXtreme Gradient Boosting)	Penelitian ini berhasil mengembangkan model prediksi angka harapan hidup dengan tingkat akurasi sebesar 96,8%. Nilai <i>Mean Absolute Error (MAE)</i> sebesar 0,97 menunjukkan rata-rata selisih prediksi dengan data aktual relatif kecil. Hasil tersebut menegaskan bahwa algoritma <i>XGBoost</i> Regressor memiliki kemampuan yang cukup andal dalam mengestimasi angka harapan hidup di negara-negara Asia. Temuan ini juga memberikan kontribusi penting bagi proses perumusan kebijakan berbasis data, khususnya dalam bidang kesehatan dan kesejahteraan masyarakat (Kurniawan dkk. 2024)
Pemodelan Ritme Kalori Terbakar Setiap Waktu Selama Bersepeda dengan Feedforward Neural Network	Feedforward Neural Network	Pada penelitian ini, model Feedforward Neural Network (FFNN) yang dirancang menghasilkan 59 neuron pada lapisan tersembunyi. Output utama jaringan ini adalah prediksi jumlah kalori yang terbakar (Calt) selama aktivitas bersepeda. Hasil evaluasi menunjukkan bahwa metode prediksi FFNN memiliki tingkat kesalahan sebesar 7%, dengan akurasi yang tergolong baik. Selain itu, ditemukan hubungan bahwa semakin tinggi suhu lingkungan, maka

		<p>kalori yang terbakar juga cenderung meningkat. Berdasarkan perancangan dan pengujian yang dilakukan, dapat disimpulkan bahwa FFNN cukup optimal digunakan untuk memprediksi kalori terbakar per detik dengan akurasi yang memadai. yang digunakan pada penelitian ini cukup optimal untuk memprediksi kalori yang terbakar per detik dengan akurasi yang cukup baik yaitu sebesar 93%. (Ichwan. 2024)</p>
<p>Model Prediksi Kepadatan Lalu Lintas: Perbandingan Antara Algoritma Random Forest dan <i>XGBoost</i></p>	<p>Random Forest dan <i>XGBoost</i></p>	<p>Kemampuan Hasil penelitian menunjukkan bahwa algoritma Random Forest dan <i>XGBoost</i> tidak memiliki perbedaan signifikan dalam hal akurasi prediksi volume lalu lintas. Namun, nilai <math>R^2</math> pada <i>XGBoost</i> lebih tinggi, sehingga kelayakan model ini dapat dikatakan lebih baik dibanding Random Forest. Perbedaan mencolok justru terlihat pada waktu pemrosesan, di mana <i>XGBoost</i> terbukti jauh lebih efisien, yaitu sekitar 532% lebih cepat daripada Random Forest. Berdasarkan hasil perbandingan tersebut, dapat disimpulkan bahwa <i>XGBoost</i> merupakan algoritma yang lebih tepat digunakan untuk membangun model prediksi kepadatan lalu lintas karena menawarkan akurasi tinggi sekaligus efisiensi waktu pemrosesan. (Lisanthoni dkk. 2023)</p>



<p>The Level Of Motivation Parents To Involve Their Children In Cycling Activities</p>	<p>Kuantitatif</p>	<p>Berdasarkan hasil penelitian, dapat disimpulkan bahwa motivasi utama orang tua dalam mengikutsertakan anak pada aktivitas bersepeda untuk mengisi waktu luang tergolong penting, yakni sebesar 66,7%. Sejalan dengan temuan pada bagian hasil dan pembahasan, mayoritas orang tua memandang bersepeda sebagai kegiatan bermanfaat untuk mengisi waktu senggang, sementara 29,2% menekankan pada aspek perkembangan motorik anak, dan 12,5% memfokuskan pada peningkatan imunitas tubuh. Keputusan orang tua dalam mengizinkan anak bersepeda perlu dikaji lebih mendalam, apakah sekadar mengikuti tren atau memang didasarkan pada tujuan tertentu. Selain itu, penelitian ini juga berpotensi diperluas dengan membandingkannya pada konteks masyarakat perkotaan, sehingga dapat memberikan gambaran yang lebih menyeluruh mengenai fenomena ini. (Pendidikan dkk . 2021)</p>
<p>Penerapan Algoritma <i>Extreme Gradient Boosting</i> (<i>XGBoost</i>) Untuk Analisis Risiko Kredit</p>	<p><i>XGBoost</i> (eXtreme Gradient Boosting)</p>	<p>Hasil analisis dengan menggunakan algoritma <i>XGBoost</i> menunjukkan bahwa metode ini memiliki kinerja yang sangat baik dalam mengolah dan mengevaluasi data kredit. Pada tahap data mining, dilakukan pembangunan sebanyak 10 model untuk menentukan model yang paling optimal. Sebelum proses pemodelan</p>

		<p>terhadap <i>German Credit Data</i>, dilakukan penyeimbangan kelas data dengan teknik <i>SMOTE</i> agar distribusi kelas lebih proporsional. Berdasarkan hasil evaluasi, penerapan <i>XGBoost</i> terbukti efektif dalam menganalisis data kredit untuk mendukung keputusan menerima atau menolak pengajuan kredit. Selain itu, performa model semakin meningkat setelah penerapan <i>SMOTE</i>, ditunjukkan dengan kenaikan nilai <i>accuracy</i> dan <i>AUC</i>. Model terbaik diperoleh pada skenario 1 (dengan pembagian data latih 90% dan data uji 10%) yang menghasilkan nilai <i>accuracy</i> sebesar 0,83 dan <i>AUC</i> sebesar 0,918. (Saputra dkk. 2024)</p>
Kepopuleran Olahraga Sepeda sebagai Bagian dari Pola Hidup Sehat	Kuantitatif	<p>Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa tingkat minat masyarakat terhadap aktivitas olahraga bersepeda tergolong sangat tinggi. Hasil observasi menunjukkan bahwa bersepeda saat ini menjadi salah satu pilihan olahraga yang tepat dan populer di kalangan masyarakat. Dari deskripsi serta analisis data yang diperoleh, dapat diketahui bahwa olahraga bersepeda tidak hanya diminati secara luas, tetapi juga sesuai untuk dilakukan oleh berbagai kalangan tanpa batasan usia. (Irawati, Irwin, and Anshar 2024).</p>
Upaya bersepeda sebagai moda transportasi serta gaya hidup baru menjaga kebugaran jasmani	Kuantitatif	<p>Kebugaran jasmani penting untuk mendukung produktivitas, dan dapat dicapai melalui pola hidup sehat seperti</p>

		<p>makan bergizi, istirahat cukup, serta olahraga teratur. Salah satu olahraga aerobik yang dianjurkan adalah bersepeda (gowes), karena memiliki banyak manfaat, seperti meningkatkan kebugaran, memperbaiki sirkulasi darah, metabolisme, serta imunitas tubuh. Aktivitas ini memerlukan persiapan alat, pemanasan, latihan sesuai program, dan diakhiri dengan pendinginan. (Utomo 2016)</p>
<p>Prediksi Jumlah Pasien Medical Check Up Berdasarkan Time Series Forecasting Menggunakan Algoritma <i>XGBoost</i></p>	<p><i>XGBoost</i> (eXtreme Gradient Boosting)</p>	<p>Penelitian ini memprediksi kunjungan pasien medical checkup menggunakan <i>XGBoost</i>. Data mengalami kekosongan dan outlier akibat pandemi, yang ditangani melalui pembersihan dan manipulasi data. Feature engineering dilakukan dengan menambahkan moving average, yang terbukti berkorelasi positif terhadap kunjungan pasien. Delapan model dibangun dari empat versi dataset, dengan Model 7 sebagai yang terbaik (MAE: 8,607), menggunakan manipulasi outlier dan fitur moving average. Penanganan outlier dan penambahan fitur terbukti meningkatkan akurasi prediksi. (Abdullah 2025)</p>
<p>Prediksi Stroke Menggunakan Extreme Gradient Boosting</p>	<p><i>XGBoost</i> (eXtreme Gradient Boosting)</p>	<p>Penelitian ini memanfaatkan <i>XGBoost</i> untuk prediksi penyakit stroke dan menghasilkan performa yang baik (akurasi 95,4%, precision 94,3%, recall 96,6%, F1-score 95,4%, <i>AUC</i> 95,4%). Meski hasilnya tinggi, performanya masih di bawah model Stacking dan Random Forest dari penelitian sebelumnya. Teknik</p>

		<p><i>SMOTE</i> digunakan untuk mengatasi data imbalance, dan Bayesian Optimization untuk tuning hyperparameter. Peneliti menyarankan optimasi lebih lanjut karena ruang pencarian hyperparameter masih luas, membuka peluang peningkatan di penelitian selanjutnya. (Murdiansyah 2024)</p>
<p>Survei Antusiasme Dan Ketertarikan Masyarakat Dalam Bersepeda Di Kota Semarang</p>	Kualitatif	<p>Antusiasme serta minat masyarakat Kota Semarang terhadap aktivitas bersepeda tergolong dalam kategori sedang. Hal ini dipengaruhi oleh beberapa faktor, terutama karena budaya hidup aktif melalui olahraga bersepeda di Kota Semarang belum sepenuhnya ditunjang oleh ketersediaan fasilitas umum yang memadai. Selain itu, tingkat kepadatan lalu lintas serta kondisi kontur jalan yang relatif kurang aman bagi pesepeda turut menjadi kendala dalam meningkatkan minat masyarakat. (Surojo dkk, 2022)</p>
<p>Komunitas Gowes Sepeda Menjadi Gaya Hidup Sehat (Pada Komunitas Sepeda Gowes Rumah Sakit Soeradji Tirtonegoro Klaten Atau SCCK</p>	Kuantitatif	<p>Komunitas gowes berpengaruh positif terhadap gaya hidup sehat anggotanya. Anggota mampu mengontrol pola makan karena pengalaman dalam komunitas mengenai makanan bergizi dan yang harus dihindari. Motivasi bergabung beragam, mulai dari rasa jenuh, dorongan diri sendiri, hingga pengaruh lingkungan, dengan tujuan utama menjaga kesehatan dan menghilangkan penat. Bersepeda berkelompok memberi manfaat biologis dan psikologis, memacu semangat, serta membantu menjaga konsistensi. Selain itu, anggota</p>

		menjadi lebih sadar untuk menjauhi hal-hal yang merusak kesehatan dan terdorong untuk membentuk perilaku hidup sehat melalui pengalaman bersama komunitas. (Al Anshori Rosid Fachrudin 2013)
Penerapan Algoritma <i>XGBoost</i> untuk Prediksi Diabetes: Analisis Confusion Matrix dan ROC Curve	<i>XGBoost</i>	Model prediksi diabetes berbasis <i>XGBoost</i> yang dikembangkan pada penelitian ini menunjukkan kinerja yang sangat baik dengan tingkat akurasi sebesar 96,33% dan recall mencapai 97,16% pada dataset Pima Indians. Berdasarkan analisis confusion matrix, hanya terdapat 6 kasus false negative (1,4%) serta 16 kasus false positive (3,8%). Nilai ROC-AUC sebesar 0,99 menegaskan bahwa model memiliki kemampuan diskriminasi yang hampir sempurna. Setelah melalui prosese estimasi hyperparameter, <i>XGBoost</i> terbukti lebih unggul dibandingkan dengan algoritma pembandingan lainnya. Walaupun hasil yang diperoleh cukup menjanjikan, penelitian ini tetap merekomendasikan beberapa langkah lanjutan, di antaranya: melakukan validasi eksternal pada populasi yang lebih beragam, mengurangi jumlah false positive secara lebih optimal, serta mengembangkan antarmuka klinis yang ramah pengguna. Dengan demikian, meskipun model ini memiliki potensi besar sebagai alat bantu pengambilan keputusan medis, penerapannya di lingkungan klinis tetap harus dilakukan



		secara bertahap dengan memperhatikan aspek keselamatan pasien serta validasi yang komprehensif. (Susanto dkk 2025)
--	--	--

## 2.2 Dasar Teori

### 2.2.1 Sepeda

Sepeda pertama kali ditemukan oleh Baron Karl Drais von Sauerbronn, yang lebih dikenal dengan nama Karl Drais. Ia lahir pada 29 April 1785 di kota Karlsruhe, Jerman. (Arsyad and Wahyuni 2021). Bersepeda merupakan aktivitas fisik yang melibatkan pengendaraan sepeda dengan tenaga otot dan kini berkembang menjadi salah satu cabang olahraga yang populer di Indonesia. Dalam dunia olahraga, terdapat beberapa jenis sepeda berdasarkan jenis lintasannya, antara lain: sepeda gunung (MTB/*mountain bike*), BMX (*bicycle moto cross*), sepeda jalan raya (*road bike*), dan sepeda trek (*track*).

Olahraga sepeda dapat dilakukan di berbagai lintasan seperti jalan raya, gunung, bukit, hingga velodrome atau lintasan buatan. Di Indonesia, olahraga balap sepeda dinaungi oleh Pengurus Besar Ikatan Sport Sepeda Indonesia (PB ISSI), sedangkan secara internasional, organisasi tertinggi adalah Union Cycliste Internationale. (Hermawan, dkk 2024)

### 2.2.2 Perfoma Kecepatan

Kecepatan adalah kemampuan seseorang dalam melakukan suatu tindakan atau mencapai jarak tertentu dalam waktu yang singkat (Rambe dkk. 2023). Dalam konteks olahraga bersepeda, kecepatan sangat berperan penting untuk mencapai hasil terbaik, terutama dalam konteks lomba atau kompetisi, karena berkaitan langsung dengan efisiensi waktu tempuh. Kecepatan tidak dapat dipisahkan dari beberapa faktor utama yang menunjang performa pesepeda, antara lain: teknik, daya tahan, kekuatan, strategi, mental, dan kondisi fisik. Teknik adalah metode gerakan yang dikembangkan secara efisien untuk menghasilkan kecepatan yang optimal.

Sementara itu, strategi dan mental akan membantu pesepeda mengambil keputusan yang tepat selama aktivitas berlangsung (Samodra dkk. 2022). Dari



keseluruhan faktor tersebut, kondisi fisik menjadi elemen paling dominan, kondisi fisik adalah prasyarat utama dalam mengembangkan prestasi olahraga secara optimal, dan harus disesuaikan dengan karakteristik masing-masing cabang olahraga. Tanpa kondisi fisik yang prima, faktor-faktors lainnya tidak dapat berjalan maksimal. Hal ini sejalan dengan pendapat yang menyatakan bahwa kondisi fisik merupakan dasar utama untuk mendukung penampilan olahraga yang efektif dan maksimal. (Herlambang and Wahyuri 2023)

### 2.2.3 Topografi dan Cuaca

Indikator evaluasi sangat penting untuk menilai kinerja model yang dilatih dengan data beranotasi terbatas, yang ditugaskan untuk mengklasifikasikan kategori baru secara akurat hanya dengan beberapa sampel berlabel. Biasanya, evaluasi ini dilakukan dalam pengaturan N-way K-shot, di mana setiap tugas terdiri dari support set dan verification set yang mencakup N kategori (way) dan K sampel dalam support set untuk setiap kategori. Akurasi algoritma ditentukan dengan mengujinya pada verification set, mengulangi evaluasi beberapa kali, dan menghitung rata-rata akurasi. Secara lebih spesifik, akurasi klasifikasi pada metode few-shot learning dihitung berdasarkan proporsi instance yang diklasifikasikan dengan benar dibandingkan dengan ukuran total dataset.

Topografi adalah kondisi permukaan bumi yang menggambarkan kemiringan atau kontur lahan. Semakin besar kontur atau elevasi suatu lahan, maka kemiringan lerengnya juga semakin tinggi, yang berdampak pada tingkat kesulitan dalam bersepeda (Yasmin 2022). Dalam analisis aktivitas bersepeda, topografi umumnya dinyatakan dalam dua indikator utama :

- Elevasi total (*elevation gain*) : jumlah total kenaikan ketinggian yang dicapai selama aktivitas bersepeda.
- Kemiringan (*gradient/slope*) : rasio antara perubahan elevasi vertikal terhadap jarak horizontal

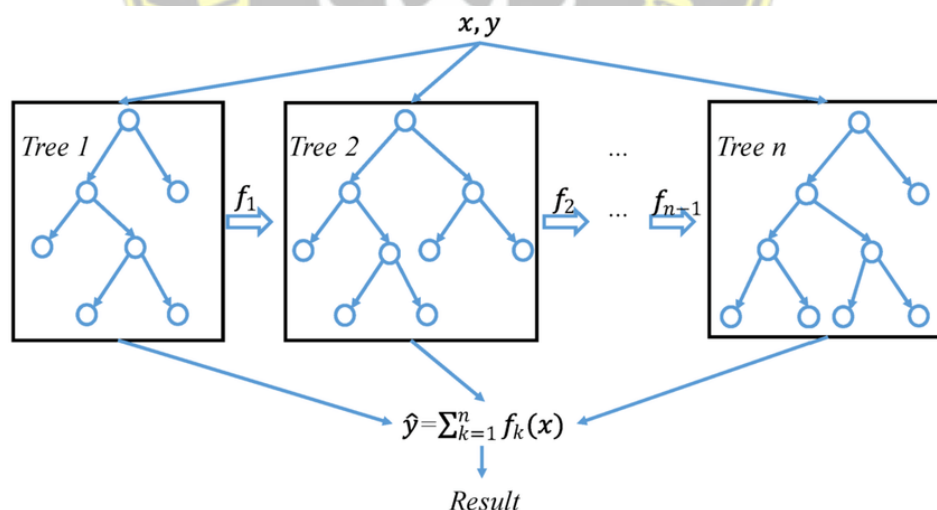
Topografi yang curam atau berbukit akan meningkatkan beban kerja pesepeda dan secara umum akan menurunkan kecepatan rata-rata

Sementara itu, cuaca juga merupakan faktor penting yang memengaruhi kenyamanan, keamanan, dan performa dalam bersepeda. Beberapa kondisi cuaca

yang signifikan dalam memengaruhi kecepatan antara lain, suhu tinggi, angin kencang, hujan dan kondisi jalan basah sehingga cenderung memperlambat laju untuk menjaga keselamatan. Selain itu, beberapa studi menyebutkan bahwa faktor-faktor seperti cuaca buruk, keselamatan lalu lintas, dan kondisi jalan menjadi alasan utama seseorang memilih untuk tidak bersepeda. Bahkan untuk pesepeda aktif sekalipun, keputusan untuk tidak bersepeda dapat dipengaruhi oleh jarak yang terlalu jauh, kurangnya fasilitas, atau kebutuhan menggunakan kendaraan bermotor untuk aktivitas tertentu. (Kurniasih dkk, Bappeda Kota Yogyakarta 2023)

#### 2.2.4 Xtreme Gradient Boosting (*XGBoost*)

*XGBoost* (Xtreme Gradient Boosting) adalah penyempurnaan dari algoritma Gradient Boosting yang dikenal memiliki kecepatan eksekusi yang lebih tinggi serta tingkat akurasi yang lebih baik. *XGBoost* termasuk dalam metode ensemble learning yang bekerja dengan menggabungkan sejumlah model pohon keputusan (decision tree), lalu secara bertahap memperbaiki kesalahan prediksi dengan menambahkan pohon baru pada setiap iterasi. *XGBoost* dikenal memiliki performa tinggi dalam berbagai tugas prediksi regresi maupun klasifikasi (Astutiningsih, dkk. 2023)



Gambar 2.1 Arsitektur *XGBoost*

Pada gambar 2.1 Algoritma *XGBoost* bekerja dengan cara membangun pohon keputusan secara bertahap (boosting). Pada setiap iterasi, algoritma akan

meng-update bobot dari pohon (tree) sebelumnya berdasarkan error yang masih tersisa, kemudian membangun pohon baru untuk memperbaiki prediksi.

Setelah seluruh pohon dibangun, prediksi akhir diperoleh dengan menjumlahkan bobot kontribusi dari setiap pohon dan kemudian memasukkannya ke dalam fungsi objektif. Secara umum, fungsi prediksi *XGBoost* dapat ditulis sebagai:

$$\hat{y} = \sum_{k=1}^K f_k(x_i) \dots \dots \dots (01)$$

dengan:

$\hat{y}$  adalah output atau hasil prediksi dari model untuk data ke- $i$ ,

$K$  adalah jumlah total pohon (round boosting),

$f_k$  merupakan fungsi dari pohon keputusan ke- $k$  yang memetakan input  $x_i$  ke nilai prediksi

$x_i$  adalah vektor fitur dari data ke- $i$ .

Setiap fungsi  $f_k$  termasuk dalam ruang fungsi  $F$ , yang didefinisikan sebagai sekumpulan pohon keputusan dengan struktur tertentu. Model *XGBoost* dibangun secara bertahap dengan meminimalkan fungsi loss yang dilengkapi dengan regularisasi untuk menghindari overfitting.

### 2.2.5 Strava

Strava adalah aplikasi olahraga yang pertama kali diluncurkan pada tanggal 18 Agustus 2009. Strava dikembangkan oleh dikembangkan oleh Michael Horvath dan Mark Gainey, Aplikasi Strava ialah aplikasi yang sangat populer di kalangan penggemar olahraga, lari, bersepeda, dan hiking. Pelari dan pengendara sepeda biasanya menggunakan aplikasi ini untuk merekam kegiatan olahraga mereka, seperti melacak jarak dan waktu yang ditempuh saat berlari atau bersepeda (Su 2018).

Keunggulan Strava juga terletak pada kemampuan analisis data personal dan komunitas, yang membuatnya banyak digunakan dalam studi akademik, riset perilaku olahraga, hingga pemetaan infrastruktur sepeda. Strava memiliki sejumlah fitur, antara lain :

- Perekaman rute berbasis GPS
- Pencatatan jarak tempuh, waktu tempuh, elevasi dan kecepatan

- Perhitungan kecepatan rata-rata dan kecepatan maksimum

Cuaca, kelembaban dan angin

#### 2.2.6 Waktu Tidur

Tidur merupakan salah satu kebutuhan fisiologis yang sangat penting bagi tubuh manusia. Kualitas dan kuantitas tidur berpengaruh langsung terhadap fungsi kognitif, fisik, dan psikologis seseorang. Menurut National Sleep Foundation (NSF) dan World Health Organization (WHO), (Surojo, dkk. 2022), durasi tidur yang ideal untuk orang dewasa berada pada kisaran 7–9 jam per malam untuk mendukung pemulihan tubuh secara optimal, dalam konteks performa fisik seperti bersepeda, tidak semua orang mendapatkan waktu tidur yang cukup. Penelitian yang dilakukan oleh Watson et al. (2015) menunjukkan bahwa tidur di bawah 5 jam per malam tergolong tidur kurang (sleep deprivation) dan dapat menyebabkan:

- Penurunan daya tahan tubuh.
- Penurunan kemampuan koordinasi motorik.
- Peningkatan rasa lelah saat beraktivitas.
- Penurunan kemampuan fokus dan reaksi.

Sebaliknya, durasi tidur di atas 5 jam sudah dapat memberikan waktu pemulihan minimum bagi tubuh untuk mempertahankan fungsi dasar, walaupun masih belum optimal jika dibandingkan dengan durasi tidur 7–9 jam yang direkomendasikan. Bagi pesepeda, kurang tidur dapat berdampak pada penurunan performa bersepeda, khususnya pada aspek kekuatan, kecepatan, dan daya tahan. Oleh karena itu, dalam penelitian ini digunakan batasan bahwa:

- Tidur  $\geq 5$  jam dianggap cukup untuk mempertahankan performa minimal.
- Tidur  $< 5$  jam dianggap kurang, berpotensi menurunkan performa bersepeda.

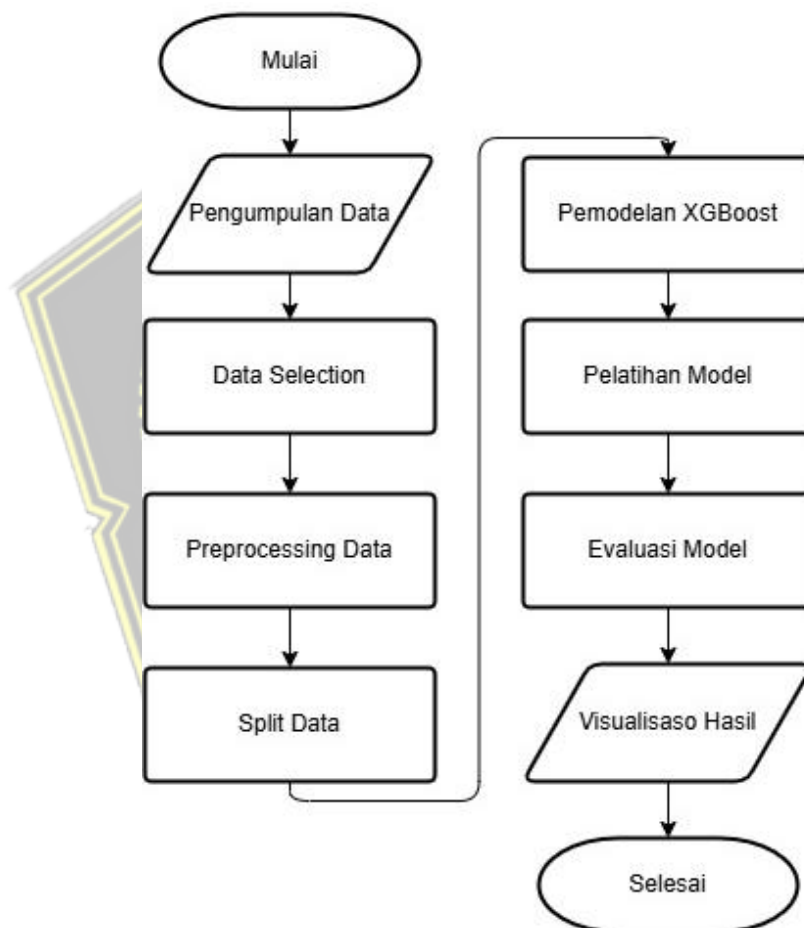
Batasan ini sejalan dengan kajian literatur yang menyatakan bahwa durasi tidur di bawah 5 jam berhubungan dengan peningkatan risiko kelelahan dan penurunan performa fisik. (Kim dkk. 2022).

## BAB III

### METODE PENELITIAN

#### 3.1 Metode Penelitian

Bagian ini akan membahas mengenai langkah-langkah yang akan digunakan dalam penelitian ini. Langkah pertama yaitu pengumpulan dan pengolahan data, di tahap kedua adanya data selection dan dilanjutkan pemodelan atau pembuatan model, tahap ke empat yaitu pengujian dan diakhiri dengan penarikan Kesimpulan sebagai hasil akhir.



Gambar 3. 1 Alur Penelitian

##### 3.1.1 Studi Literatur

Pada proses ini adalah pengumpulan data yang diperlukan oleh peneliti. Sumber data pada penelitian ini menggunakan data aktivitas bersepeda pribadi yang diperoleh melalui aplikasi Strava, dengan rentang waktu mulai dari 21 Oktober 2024 hingga 1 Juli 2025. Data tersebut dikumpulkan oleh peneliti secara langsung

dari Strava langsung diinput ke dalam Excel. Dataset yang digunakan berisi informasi aktivitas bersepeda seperti waktu mulai, durasi aktivitas, jarak tempuh, total elevasi, kecepatan rata-rata, dan kecepatan maksimum. Data tersebut diekspor dalam format .csv, yang kemudian diolah dan dianalisis.

### 3.1.2 Pengumpulan Dataset

Pada proses ini adalah pengumpulan data yang diperlukan oleh peneliti. Sumber data pada penelitian ini menggunakan data aktivitas bersepeda pribadi yang diperoleh melalui aplikasi Strava, dengan rentang waktu mulai dari 2 Juni 2024 hingga 9 Agustus 2025. Data tersebut dikumpulkan oleh peneliti secara langsung dari Strava langsung diinput ke dalam Excel. Dataset yang digunakan berisi informasi aktivitas bersepeda seperti waktu mulai, durasi aktivitas, jarak tempuh, total elevasi, kecepatan rata-rata, dan kecepatan maksimum. Data tersebut diekspor dalam format .csv, yang kemudian diolah dan dianalisis

Tabel 3. 1 dataset

Tanggal	Jam Mulai	Waktu Total	Elevasi	Jarak Tempuh	Kecepatan Rata-rata	Kecepatan Max	Cuaca	Waktu Tidur
2 Juni 24	15.42	40.13.23	80 m	15.67 km	25,5 km/jam	45.3 km/jam	0	7
9 Juni 24	5.41	2.45.12	723 m	29.12 km	21,2 km/jam	52,6 km/jam	0	6
15 Juni 24	5.33	1.24.35	275 m	33,72 km	26,3 km/jam	49,5 km/jam	1,2	4
23 Juni 24	15.53	1.14.45	57 m	25,3 km	18,7 km/jam	37,2 km/jam	3,5	7
1 Juli 24	5.45	1.40.50	427 m	36,6 km	22,4 km/jam	32,7 km/jam	1,7	7

Data yang digunakan dalam penelitian ini diperoleh dari riwayat aktivitas bersepeda yang direkam menggunakan aplikasi Strava. Seluruh data diunduh melalui fitur ekspor Strava dalam format Excel sehingga memudahkan proses pengolahan lebih lanjut. Dataset berisi catatan aktivitas mulai tanggal 2 Juni 2024 hingga akhir periode pengumpulan data. Setiap baris pada dataset mewakili satu sesi bersepeda, sedangkan setiap kolom berisi atribut (fitur) yang relevan untuk penelitian ini. Terdapat sembilan kolom (fitur) utama, yaitu:



- Tanggal – tanggal pelaksanaan aktivitas bersepeda.
- Jam Mulai – waktu mulai aktivitas bersepeda.
- Waktu Total – durasi total aktivitas bersepeda dari awal hingga selesai.
- Elevasi – total kenaikan ketinggian yang ditempuh (dalam meter).
- Jarak Tempuh – panjang rute bersepeda (dalam kilometer).
- Kecepatan Rata-rata – kecepatan rata-rata selama bersepeda (km/jam).
- Kecepatan Maksimum – kecepatan tertinggi yang dicapai dalam satu sesi bersepeda (km/jam).
- Cuaca – kondisi cuaca saat bersepeda (cerah, mendung, hujan).
- Waktu Tidur – lama waktu istirahat sebelum bersepeda (dalam jam).

Data ini kemudian digunakan sebagai bahan utama dalam tahap pembersihan data (*data cleaning*), *feature engineering*, dan pembuatan model *XGBoost* untuk memprediksi kecepatan rata-rata bersepeda

### 3.1.3 Pre Processing

Tahapan ini adalah proses mengolah data mentah menjadi data yang siap digunakan untuk pemodelan dengan tujuan untuk menghasilkan data yang lebih akurat. Data preprocessing meliputi tahapan sebagai berikut :

### 3.1.4 Data Cleaning

*Data cleaning* adalah proses yang dilakukan untuk memastikan data mentah memiliki kualitas yang layak digunakan dalam analisis. Proses ini mencakup penanganan terhadap data yang hilang, penghapusan data yang duplikat, identifikasi dan penanganan outlier, serta eliminasi data yang tidak relevan

- Penanganan Nilai Yang Hilang (*Missing Value*)

Penanganan nilai yang hilang adalah langkah pertama dan salah satu yang paling fundamental dalam tahap *data cleaning*. Keberadaan data yang kosong atau tidak lengkap (*NaN/null*) dapat menyebabkan kesalahan komputasi selama pelatihan model dan menghasilkan analisis yang bias serta prediksi yang tidak akurat. Oleh karena itu, identifikasi dan penanganan nilai yang hilang menjadi wajib dilakukan. Untuk mengidentifikasi keberadaan nilai yang hilang secara sistematis, penelitian ini menggunakan pustaka *pandas* pada *Python*. Prosesnya adalah dengan menerapkan fungsi `.isnull().sum()`. Fungsi ini akan memindai seluruh

DataFrame dan menghasilkan sebuah objek baru dengan ukuran yang sama, di mana setiap sel yang berisi nilai hilang akan ditandai sebagai `True`, dan yang berisi data akan ditandai sebagai `False`. Hasilnya adalah sebuah daftar yang menunjukkan nama setiap kolom beserta jumlah total nilai yang hilang di dalamnya.

#### - Penanganan data duplikat

Tujuan dari tahap penanganan data duplikat ini adalah untuk memastikan bahwa setiap baris data dalam dataset bersifat unik. Keberadaan data duplikat dapat menyebabkan bias pada model *machine learning*, karena model akan memberikan bobot yang tidak semestinya pada data yang berulang. Hal ini juga dapat mengarah pada hasil evaluasi yang terlalu optimis dan tidak mencerminkan kinerja model di dunia nyata. Implementasi untuk menangani data duplikat dilakukan menggunakan pustaka `pandas` dengan fungsi `.duplicated().sum()`. Fungsi ini mengidentifikasi apakah terdapat baris data yang identik. Dalam konteks data deret waktu, ini biasanya berarti memeriksa apakah ada catatan dengan stempel waktu (*timestamp*) dan nilai-nilai fitur yang sama persis. Jika hasil dari langkah pertama menunjukkan adanya data duplikat (lebih dari 0), maka tindakan penghapusan akan dilakukan.

### 3.1.5 Normalisasi Data

Normalisasi data merupakan teknik krusial dalam tahap pra-pemrosesan, terutama karena setiap variabel dalam dataset umumnya memiliki skala nilai yang berbeda. Dalam penelitian ini, peneliti menerapkan metode min-max scaler untuk melakukan normalisasi, yaitu dengan mengubah nilai-nilai data ke dalam rentang tertentu umumnya antara 0 hingga 1 agar setiap variabel berada pada skala yang sebanding dan dapat dibandingkan secara seimbang (Salsabil et al., 2024)

$$x' = \frac{(x - \min_x)}{(\max_x - \min_x)} \dots\dots\dots (02)$$

Dimana :

$X_{scaled}$  adalah nilai data setelah dinormalisasi.

$X$  adalah nilai data asli.

$X_{min}$  adalah nilai minimum dari seluruh data pada fitur tersebut.

$X_{max}$  adalah nilai maksimum dari seluruh data pada fitur tersebut

### 3.1.6 Split Data

Setelah data dibersihkan dan dinormalisasi, Langkah berikutnya adalah melakukan pemisahan dataset ke dalam dua kelompok, yaitu data latih (*training data*) dan data uji (*testing data*). Proses ini bertujuan untuk memberikan data pembelajaran kepada model sekaligus menyediakan data terpisah guna menguji kinerja model setelah pelatihan sekaligus dievaluasi secara objektif. merupakan prosedur fundamental dalam pengembangan model *machine learning* untuk mengevaluasi kinerja model secara objektif. Tujuan utama dari pembagian data adalah untuk melatih model pada satu set data (data latih) dan kemudian menguji seberapa baik model tersebut dapat melakukan generalisasi pada data baru yang belum pernah dilihat sebelumnya (data uji). Pembagian data tidak boleh dilakukan secara acak. Pembagian harus dilakukan secara kronologis untuk menjaga dependensi temporal data. Hal ini mensimulasikan skenario dunia nyata di mana kita menggunakan data masa lalu untuk memprediksi data di masa depan. Dalam penelitian ini, rasio pembagian yang digunakan adalah 80:20, dengan rincian yaitu 80% sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*).

### 3.1.7 Pembuatan XGBoost

Tahap pembuatan model merupakan inti dari penelitian ini karena pada bagian ini dilakukan implementasi dan pendekatan pemodelan yaitu model *XGBoost* (Extreme Gradient Boosting) yang merupakan algoritma machine learning berbasis pohon keputusan. Tujuan dari tahap ini adalah membangun kedua model dengan arsitektur dan parameter yang sesuai.

Langkah pembuatan model *XGBoost* :

1. Persiapan Data untuk *XGBoost*: Berbeda dengan GRU yang memproses data sekuensial, *XGBoost* memerlukan data dalam format tabular (baris dan kolom). Oleh karena itu, data deret waktu perlu diubah menjadi format *supervised learning*. Proses ini biasanya melibatkan pembuatan fitur (*feature engineering*) dari data masa lalu, seperti menggunakan harga penutupan hari sebelumnya ( $t-1$ ,  $t-2$ , dst.) dan nilai indikator teknikal sebagai fitur input untuk memprediksi harga hari ini ( $t$ ).

2. Inisialisasi Model *XGBoost*: Model *XGBoost* tidak dibangun lapis demi lapis seperti jaringan saraf. Sebagai gantinya, sebuah objek model diinisialisasi, yaitu *XGBRegressor*, karena tugasnya adalah prediksi regresi (memprediksi nilai kontinu).
3. Menentukan *Hyperparameter*: Sejumlah *hyperparameter* utama ditentukan untuk mengontrol kinerja dan kompleksitas model. Beberapa di antaranya adalah:
  - *n\_estimators*: Jumlah total pohon keputusan (*decision tree*) yang akan dibangun.
  - *learning\_rate*: Mengontrol laju belajar model untuk mencegah *overfitting*.
  - *max\_depth*: Kedalaman maksimum dari setiap pohon untuk membatasi kompleksitas.
  - *objective*: Fungsi tujuan yang dioptimalkan, misalnya 'reg:squarederror' untuk masalah regresi.

Melatih model: Model dilatih menggunakan fungsi `.fit()` pada data latih. Seringkali, teknik Early Stopping digunakan selama pelatihan, di mana performa model dipantau pada data validasi. Pelatihan akan berhenti jika tidak ada peningkatan performa setelah beberapa iterasi tertentu untuk mendapatkan model yang optimal.

### 3.1.8 Evaluasi Model

Setelah model *XGBoost* berhasil dilatih menggunakan data latih (training data), tahap selanjutnya adalah melakukan evaluasi untuk mengukur kinerja dan akurasi prediksi kecepatan rata-rata bersepeda. Evaluasi ini bertujuan untuk mengetahui sejauh mana model mampu memprediksi dengan benar pada data yang belum pernah digunakan selama proses pelatihan. Data yang digunakan untuk evaluasi adalah data uji (test data), sehingga hasilnya dapat mencerminkan kemampuan generalisasi model terhadap data baru.

Dalam penelitian ini, metrik evaluasi yang digunakan adalah Root Mean Squared Error (RMSE) dan R-squared ( $R^2$ ).

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots\dots\dots (03)$$

Dimana :

$y_i$ : nilai aktual

$\hat{y}_i$ : nilai prediksi

$\frac{1}{n}$ : menghitung rata-rata dari kuadrat eror

$\sum$  : menjumlahkan semua eror dari semua data

Dari perhitungan tersebut memberikan pengertian bahwa semakin kecil RMSE yang dihasilkan maka akurasi model yang digunakan semakin tinggi dan sebaliknya apabila hasil RMSE semakin besar maka nilai akurasi pada model semakin kecil. RMSE ini dipilih karena keunggulan yang dimiliki oleh RMSE dibandingkan dengan perhitungan akurasi ya

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots\dots\dots (04)$$

Dimana :

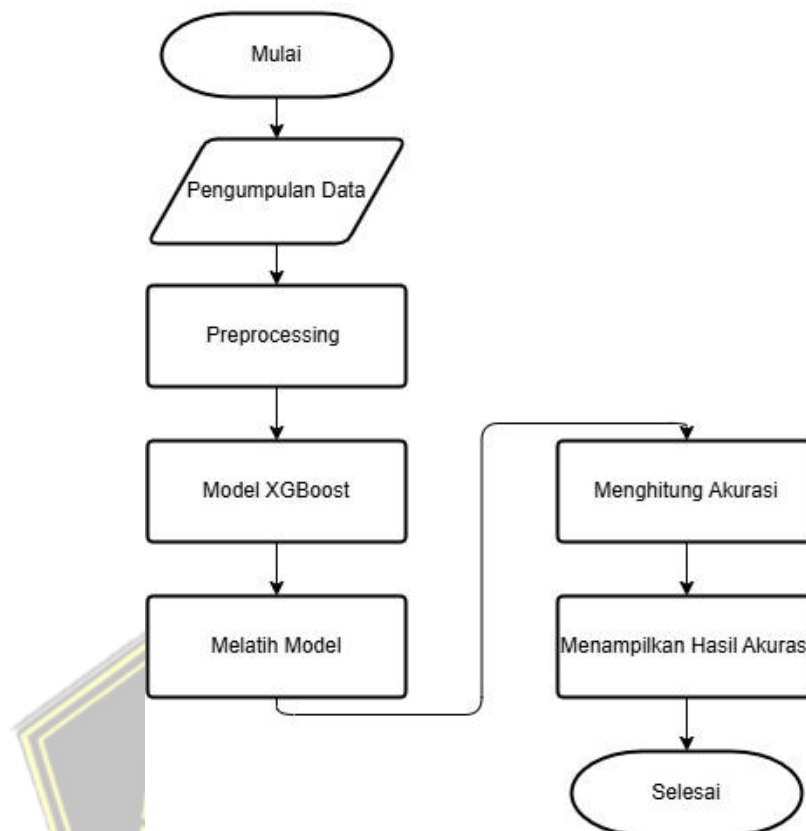
$n$  jumlah total data

$y_i$  adalah nilai harga aktual pada data ke-I

$\hat{y}_i$  adalah nilai harga prediksi yang dihasilkan oleh model pada data ke-i

$\bar{y}$  adalah nilai rata-rata dari seluruh harga aktual

Selain RMSE, penelitian ini juga menggunakan metrik *R-squared* ( $R^2$ ) atau Koefisien Determinasi untuk mengukur kecocokan model (*goodness-of-fit*). Metrik ini berfungsi untuk mengukur seberapa besar proporsi variasi yang berhasil dijelaskan oleh model prediksi. Nilai  $R^2$  berkisar antara 0 hingga 1 (atau 0% hingga 100%), di mana skor yang lebih tinggi menunjukkan performa yang lebih baik. Secara matematis,  $R^2$  dihitung dengan membandingkan jumlah kuadrat kesalahan prediksi model terhadap jumlah kuadrat total dari data, sehingga memberikan gambaran komprehensif tentang kemampuan prediktif model.



Gambar 3. 2 Alur Sistem

Gambar 3.2 menunjukkan alur sistem prediksi kecepatan rata-rata bersepeda menggunakan model *XGBoost*. Proses dimulai dari tahapan awal hingga sistem menghasilkan keluaran berupa nilai akurasi dan hasil prediksi. Alurnya dapat dijelaskan sebagai berikut:

1. Mulai

Sistem dimulai dengan inisialisasi proses penelitian untuk membangun model prediksi kecepatan rata-rata bersepeda.

2. Pengumpulan Data

Data dikumpulkan dari aktivitas bersepeda pribadi menggunakan aplikasi Strava. Data mencakup berbagai faktor seperti jarak, elevasi, durasi, kondisi cuaca, dan waktu tidur.

3. Preprocessing Data

Data mentah yang diperoleh kemudian melalui proses preprocessing, Pembersihan data (menghapus data kosong/NaN), transformasi data



(konversi satuan, pengolahan tanggal/jam), feature engineering (membuat variabel turunan seperti *SleepPenalty* atau *Cuaca\_Efek*).

#### 4. Model *XGBoost*

Setelah data siap, algoritma *XGBoost* dipilih karena memiliki performa tinggi pada data tabular, mampu menangani hubungan non-linear, serta efisien dalam komputasi.

#### 5. Melatih Model

Dataset dibagi menjadi data latih (training) dan data uji (testing). Data latih digunakan untuk membangun model *XGBoost* agar dapat mempelajari pola antara input (topografi, cuaca, tidur) dengan output (kecepatan rata-rata).

#### 6. Menghitung Akurasi

Setelah model selesai dilatih, dilakukan evaluasi dengan data uji menggunakan metrik akurasi, seperti  $R^2$  (koefisien determinasi) dan RMSE (Root Mean Square Error) untuk menilai sejauh mana model mampu memprediksi dengan tepat.

#### 7. Menampilkan Hasil Akurasi

Hasil evaluasi ditampilkan berupa nilai akurasi prediksi. Hal ini menjadi dasar untuk menilai kualitas model dalam memprediksi kecepatan rata-rata bersepeda.

#### 8. Selesai

Proses berakhir setelah hasil akurasi ditampilkan. Sistem dapat digunakan untuk melakukan prediksi baru berdasarkan input parameter dari pengguna.

### 3.1.9 Halaman Awal Sistem Web

Gambar berikut merupakan rancangan awal alur kerja sistem prediksi kecepatan rata-rata bersepeda berbasis *XGBoost*. Desain ini menggambarkan tahapan utama yang akan dijalankan oleh sistem, mulai dari proses pengumpulan data aktivitas bersepeda dari Strava, pemilihan dan pembersihan data, hingga tahap pemodelan dan pelatihan menggunakan algoritma *XGBoost*. Rancangan ini juga

menampilkan proses evaluasi kinerja model serta visualisasi hasil prediksi, yang bertujuan untuk memberikan gambaran yang jelas mengenai urutan proses dari awal hingga akhir sebelum sistem diimplementasikan secara penuh.

## Cycling Speed Predictor

Prediksi Kecepatan Rata-rata Bersepeda Berdasarkan Kondisi Topografi & Cuaca

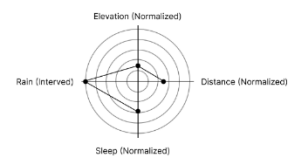
Kembali ke Home atau kunjungi Profile untuk informasi developer

Input Parameters

Jarak (KM)	25.00	Curah Hujan (mm)	0.00
5.00	50.00	0.00	100.000
Elevasi (m)	200	Jam Tidur (jam)	7
50	700	1	12
Jam Mulai	06.00		
Predict Speed			

Input Visualization

Input Parameters Overview



Kondisi: Cerah

Gambar 3. 3 Halaman Sistem

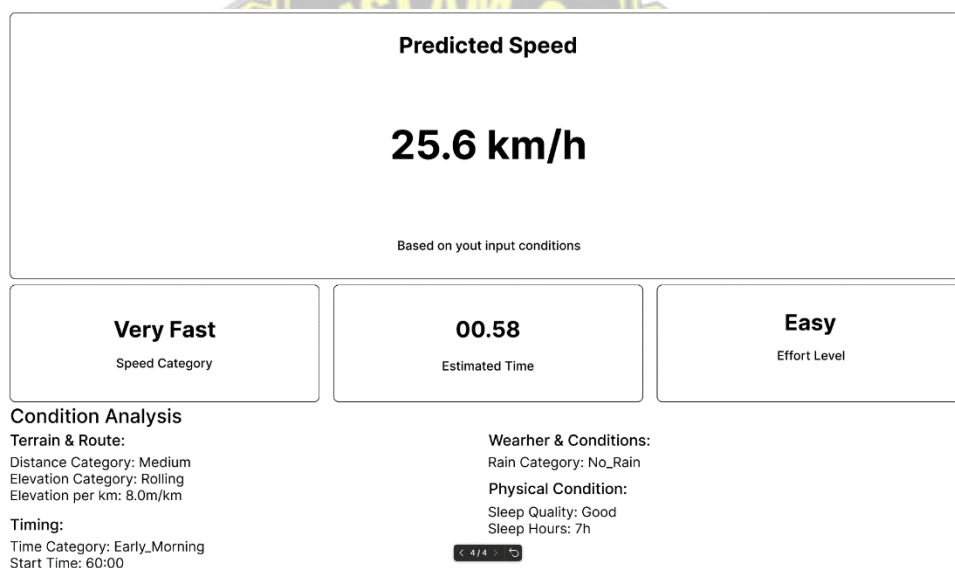
Pada Gambar 3.3 menampilkan rancangan antarmuka pengguna (*user interface*) pada halaman awal sistem prediksi kecepatan rata-rata bersepeda. Halaman ini merupakan titik awal interaksi pengguna dengan sistem, di mana pengguna dapat langsung memasukkan parameter yang diperlukan untuk melakukan prediksi.

Desain antarmuka dibuat sederhana dan intuitif agar mudah digunakan oleh siapapun. Pada bagian utama halaman, terdapat komponen input berupa *slider* untuk mengatur nilai Elevasi (meter), Jarak Tempuh (km), dan Durasi (menit), serta *radio button* untuk memilih kondisi Waktu Tidur dan *dropdown menu* untuk memilih Kondisi Cuaca. Setelah pengguna mengisi semua parameter, mereka dapat menekan tombol Prediksi. Hasil prediksi kecepatan rata-rata bersepeda akan ditampilkan secara langsung dalam bentuk angka yang mudah dibaca, sehingga pengguna dapat segera mengetahui estimasi kecepatan yang akan dicapai berdasarkan kondisi yang diinputkan.

### 3.1.10 Visualisasi Hasil

Gambar berikut merupakan *wireframe* atau rancangan awal antarmuka pengguna (*user interface*) dari sistem prediksi kecepatan rata-rata bersepeda. Desain ini menggambarkan tampilan halaman hasil prediksi yang dirancang untuk menampilkan estimasi kecepatan secara jelas setelah pengguna memasukkan seluruh parameter seperti Elevasi, Jarak Tempuh, Durasi, Waktu Tidur, dan Kondisi Cuaca.

Pada bagian utama halaman, hasil prediksi disajikan dalam bentuk angka kecepatan yang dilengkapi ikon atau simbol pendukung agar lebih mudah dipahami. Desain dibuat sederhana dan informatif sehingga pengguna dapat langsung memahami output yang diberikan sistem tanpa memerlukan penjelasan tambahan.



Gambar 3. 4 Halaman Hasil

Pada gambar 3.4 diatas merupakan tampilan sistem prediksi kecepatan rata-rata bersepeda setelah pengguna memasukkan seluruh parameter yang diperlukan. Antarmuka ini menampilkan hasil akhir dari proses perhitungan menggunakan model prediksi berbasis *machine learning*.

Hasil prediksi ditampilkan dalam bentuk teks di bagian bawah, dengan label “Prediksi Kecepatan: XX km/jam” yang ditandai dengan warna latar mencolok agar mudah terlihat. Label ini menunjukkan estimasi kecepatan rata-rata yang dapat dicapai berdasarkan input Elevasi, Jarak Tempuh, Durasi, Waktu Tidur, dan

Kondisi Cuaca yang telah dimasukkan oleh pengguna. Antarmuka ini dirancang dengan pendekatan minimalis dan fokus pada fungsionalitas, sehingga memudahkan pengguna dalam memahami hasil prediksi.

### 3.1.11 Software software yang digunakan

Dalam penelitian ini menggunakan beberapa *software* yang akan digunakan untuk mengolah data, pembuatan model serta evaluasi hasil. Berikut daftar *software* yang akan digunakan :

#### 1. *Google Colaboratory*

*Google Colab* salah satu platform Google ini dipilih karena menyediakan lingkungan *notebook Jupyter* yang dapat diakses langsung melalui peramban web, sehingga menghilangkan kebutuhan instalasi lokal dan yang terpenting, memberikan akses gratis ke sumber daya komputasi berkinerja tinggi. Secara spesifik, ketersediaan GPU (*Graphics Processing Unit*) di Google Colab menjadi fungsi krusial yang dimanfaatkan untuk secara signifikan mempercepat waktu pelatihan model *deep learning* yang bersifat komputasional intensif. Selain itu, Colab juga mempermudah manajemen pustaka *python* serta memungkinkan penyimpanan dan akses kerja yang fleksibel melalui integrasinya dengan Google Drive, menjadikannya platform yang sangat efektif dan efisien untuk melaksanakan penelitian ini.

#### 2. *Visual Studio Code*

*Visual Studio Code* merupakan sebuah *software* kode editor modern yang dikembangkan oleh Microsoft. Perangkat lunak ini bersifat gratis (*open-source*), dan dapat dijalankan di berbagai sistem operasi, termasuk Windows, macOS, dan Linux. Dalam penelitian ini, VS Code digunakan sebagai lingkungan pengembangan utama untuk proses pengkodean. Penggunaan VS Code dipilih karena fleksibilitasnya, dukungan ekstensi yang lengkap untuk pengembangan Python, dan integrasi terminal yang mempercepat alur kerja dari penulisan kode hingga proses *deployment* aplikasi.

#### 3. *Streamlit*

Streamlit adalah *framework python open-source* yang dalam penelitian ini berfungsi sebagai alat utama untuk tahap *deployment*, yaitu proses mengubah model

prediksi yang telah dilatih menjadi sebuah aplikasi web yang fungsional dan interaktif. Penggunaannya meliputi perancangan dan pembangunan seluruh antarmuka pengguna (UI). Secara fungsional, streamlit menjembatani interaksi antara pengguna dan model yang berjalan di latar belakang, skripnya bertanggung jawab untuk menerima input pilihan model (*XGBoost*), memuat file model yang sesuai, memicu proses prediksi, dan kemudian menyajikan outputnya secara komprehensif. Hasil yang disajikan mencakup nilai prediksi numerik dan juga visualisasi grafik yang dinamis, sehingga secara efektif mengubah hasil penelitian dari sekadar kode menjadi sebuah produk aplikasi yang dapat didemonstrasikan.

### 3.1.12 Library yang digunakan

Dalam penelitian ini menggunakan beberapa *library python* yang akan digunakan. Berikut daftar *library python* yang akan digunakan :

#### 1. Strava

Strava adalah aplikasi olahraga yang pertama kali diluncurkan pada tanggal 18 Agustus 2009. Strava dikembangkan oleh Michael Horvath dan Mark Gainey, Aplikasi Strava ialah aplikasi yang sangat populer di kalangan penggemar olahraga, lari, bersepeda, dan hiking. Pelari dan pengendara sepeda biasanya menggunakan aplikasi ini untuk merekam kegiatan olahraga mereka, seperti melacak jarak dan waktu yang ditempuh saat berlari atau bersepeda (Su 2018).

Keunggulan Strava juga terletak pada kemampuan analisis data personal dan komunitas, yang membuatnya banyak digunakan dalam studi akademik, riset perilaku olahraga, hingga pemetaan infrastruktur sepeda

#### 2. Pandas

Dalam *library Python*, Pandas merupakan pustaka esensial yang dirancang untuk keperluan analisis dan manipulasi data, dengan *DataFrame* sebagai struktur datanya yang paling utama. *DataFrame* sendiri adalah sebuah objek tabular dua dimensi berkinerja tinggi yang terdiri dari sumbu baris dan kolom yang dapat diberi label.

#### 3. NumPy

*NumPy (Numerical Python)* adalah pustaka fundamental untuk komputasi numerik dalam *python*, yang menyediakan objek utama berupa array multidimensi berkinerja tinggi beserta serangkaian fungsi untuk operasi matematis. Dalam penelitian ini, peran utamanya adalah pada tahap persiapan data untuk pemodelan, di mana *NumPy* digunakan untuk mentransformasikan data dari struktur *pandas DataFrame* menjadi format array numerik murni. Transformasi ini merupakan langkah krusial, terutama untuk model, karena model *deep learning* memerlukan input dalam bentuk array dengan dimensi yang spesifik. *NumPy* digunakan untuk melakukan manipulasi bentuk (*reshaping*) pada array tersebut untuk menciptakan sekuens data yang sesuai dengan arsitektur input model. Selain penggunaan langsung, *NumPy* juga berfungsi sebagai fondasi bagi pustaka lain seperti *pandas* dan *TensorFlow*, sehingga perannya sangat esensial dalam keseluruhan alur kerja komputasi.

#### 4. *Matplotlib/Seaborn*

*Matplotlib/seaborn* adalah pustaka visualisasi data yang fundamental dalam ekosistem *Python*, yang dalam penelitian ini memegang peranan krusial pada tahap evaluasi model dan *deployment*. Secara spesifik, pustaka ini digunakan untuk membuat grafik garis (*line chart*) yang membandingkan secara visual antara data deret waktu dengan data harga prediksi yang dihasilkan oleh model *XGBoost*. Visualisasi ini berfungsi sebagai alat analisis kualitatif yang penting untuk mengamati seberapa baik setiap model mampu menangkap tren dan volatilitas pasar. Grafik yang telah dihasilkan ini kemudian diintegrasikan dan ditampilkan di dalam antarmuka aplikasi *streamlit* untuk menyajikan hasil prediksi kepada pengguna akhir. Dengan demikian, *matplotlib* berfungsi sebagai jembatan untuk mengubah hasil numerik yang kompleks menjadi representasi visual yang intuitif dan mudah diinterpretasikan.

#### 5. *Scikit-learn*

*Scikit-learn* adalah pustaka machine learning yang sangat fundamental dan populer dalam ekosistem *Python*, yang menyediakan serangkaian alat komprehensif untuk tugas-tugas analisis data prediktif. Meskipun tidak digunakan secara langsung untuk membangun model *XGBoost*, pustaka ini memainkan peran



pendukung yang sangat penting dalam dua tahapan krusial penelitian ini. Pertama, pada tahap pra-pemrosesan data, kelas *MinMaxScaler* dari modul *sklearn.preprocessing* dimanfaatkan untuk melakukan normalisasi, yaitu mengubah skala seluruh fitur ke dalam rentang seragam. Kedua, pada tahap evaluasi model, fungsi-fungsi dari modul *sklearn.metrics*, seperti *mean\_squared\_error* digunakan untuk menghitung skor kuantitatif RMSE secara akurat. Dengan demikian, *Scikit-learn* berfungsi sebagai perangkat esensial yang menyediakan fungsi-fungsi standar industri untuk persiapan data dan validasi model, memastikan alur kerja penelitian ini dapat diandalkan.

#### 6. Joblib

Joblib adalah sebuah pustaka Python yang dirancang untuk menyediakan pipelining yang ringan, dengan salah satu fungsi utamanya adalah untuk menyimpan dan memuat objek Python secara efisien. Dalam alur kerja machine learning, sangat penting untuk menyimpan model yang telah dilatih agar dapat digunakan kembali tanpa perlu mengulang proses pelatihan yang memakan waktu. Dalam penelitian ini, joblib secara spesifik digunakan untuk proses serialisasi atau penyimpanan model *XGBoost* yang telah final ke dalam sebuah file. Pustaka ini dipilih karena sangat efisien dalam menangani objek yang berisi array NumPy berukuran besar, yang merupakan komponen inti dari model seperti *XGBoost*. Setelah pelatihan selesai, fungsi `joblib.dump()` digunakan untuk menyimpan model.

## BAB IV

### HASIL DAN ANALISIS PENELITIAN

#### 4.1 Hasil Penelitian

##### 4.1.1 Gambaran Umum Dataset

Tahap awal dalam penelitian ini adalah pengumpulan data historis yang akan digunakan untuk melatih dan menguji model prediksi. Data yang relevan dan berkualitas merupakan pondasi utama untuk membangun model *machine learning* yang akurat. Sistem yang dibangun pada penelitian ini merupakan sebuah aplikasi prediksi kecepatan rata-rata bersepeda yang memanfaatkan algoritma *Extreme Gradient Boosting (XGBoost)* sebagai model pembelajaran mesin. Data yang digunakan berasal dari catatan aktivitas bersepeda pribadi yang diunduh dari aplikasi Strava, kemudian dipadukan dengan data Elevasi, Cuaca dan Kondisi Tidur pesepeda

Data diperoleh dari *export* aktivitas Strava dalam bentuk Excel/CSV, yang berisi informasi seperti jarak tempuh, durasi, kecepatan rata-rata, kecepatan maks, elevasi, dan kecepatan rata-rata. Data cuaca diambil berdasarkan tanggal dan waktu aktivitas menggunakan sumber *historical weather*. Data kondisi tidur diinput secara manual sesuai catatan distarava pesepeda.

Pengumpulan data dilakukan mulai 3 Juni 2024 hingga 9 Agustus 2025. Selama periode ini yg nantinya akan digunakan sebagai data training dan testing, setiap aktivitas bersepeda dicatat secara konsisten menggunakan perangkat GPS dan data cuaca diperoleh baik dari catatan manual maupun integrasi informasi cuaca pada aplikasi Strava.

Tabel 4. 1 Dataset

Tanggal	Jam Mulai	Waktu Total	Elevasi	Jarak Tempuh	Kecepatan Rata-rata	Kecepatan Max	Cuaca	Waktu Tidur
2 Juni 24	15.42	40.13.23	80 m	15.67 km	25,5 km/jam	45.3 km/jam	0	7
10 Juni 2025	05.20	2.00.12	723 m	29.12 km	21,2 km/jam	52,6 km/jam	0	6
3 Juli 2025	16.17	1.11.44	112 m	24,5 km	20,2 km/jam	49,5 km/jam	1,2	6
9 Agustus 2025	05.38	1.44.07	217 m	25,3 km	23,7 km/jam	97,0 km/jam	3,5	8

1 Juli 24	5.45	1.40.50	427 m	36,6 km	22,4 km/jam	32,7 km/jam	1,7	7
-----------	------	---------	-------	---------	-------------	-------------	-----	---

Fitur- Fitur yang diambil meliputi

- Tanggal : Aktivitas bersepeda dilakukan
- Jam mulai : Waktu pukul aktivitas bersepeda
- Waktu total : Total durasi bersepeda
- Elveasi : Total kenaikan ketinggian
- Jarak tempuh : Panjang total route yang ditempuh
- Kecepatan rata-rata : Nilai kecepatan rata-rata bersepeda
- Kecepatan maks : Kecepatan tertinggi yang dicapai selama aktivitas bersepeda
- Cuaca : Kondisi cuaca saat bersepeda, misalnya cerah, mendung, atau hujan.
- Waktu tidur : Jumlah jam tidur sebelum aktivitas bersepeda

Fitur – fitur diatas digunakan dalam proses analisis dan pemodelan prediksi kecepatan rata-rata bersepeda .

#### 4.1.2 Data Selection

Proses seleksi data dilakukan untuk memilih data yang relevan dengan kebutuhan analisis dan pemodelan. Dari seluruh data aktivitas bersepeda yang dikumpulkan melalui Strava, hanya atribut-atribut yang memiliki pengaruh langsung terhadap kecepatan rata-rata yang dipertahankan, data yang tidak relevan seperti catatan pribadi, data sensor yang tidak konsisten, atau atribut yang memiliki nilai kosong hampir seluruhnya, dihapus untuk meningkatkan kualitas dataset

#### 4.1.3 Missing Value

Penanganan *missing value* atau data yang hilang. Dataset hasil ekspor dari Strava terkadang memiliki nilai kosong, terutama pada kolom Cuaca atau Waktu Tidur yang tidak selalu tercatat, pada tahapan ini setiap data dari Strava

Tahapan ini untuk melihat adanya *missing value* pada data strava yang hasil dari pengecekan adalah tidak adanya *missing value* yang dibuktikan dengan 0 data ditemukan sebagai *missing value* sehingga data tersebut dapat digunakan untuk proses penelitian tanpa perlu mengisi/menggantikan dengan rata-rata.

#### 4.1.4 Mengecek *duplicate data*

Pada tahapan selanjutnya adalah melihat *duplicate data* dari setiap data. Tahap ini menghasilkan tidak adanya data duplikat yang terdapat pada ketiga perusahaan sehingga data ini tidak perlu dirubah lagi untuk digunakan dalam proses selanjutnya

### 4.2 Preprocessing Data

Pada tahapan *preprocessing* data ini akan mengolah data yang sudah dibersihkan dari *missing value* dan duplikasi data yang akan digunakan untuk pemodelan

#### 4.2.1. Data *cleaning*

Dalam data *cleaning* ini dilakukan tahapan untuk mengatasi *outliers* yang ada pada dataset. Penanganan outliers ini menggunakan metode Z-SCORE. Metode ini adalah salah satu metode untuk mengatasi outliers dengan menggunakan nilai mean (rata-rata) yang dimiliki oleh data set. Dalam dataset ini banyak *outliers* ditemukan akan tetapi outliers ini tidak diproses karena beberapa hal. Salah satunya adalah apabila menghilangkan/mengganti *outlier* ini akan memberikan dampak dalam hasil prediksi Kecepatan rata – rata yang dilakukan. Selanjutnya, penulis melakukan normalisasi data yang menjadi solusi untuk data yang melebihi nilai rata-rata data dengan metode *min-max scaler*

#### 4.2.2. Normalisasi Data

Proses ini diambil untuk mengubah nilai dataset ke dalam skala 0 hingga 1 dengan tujuan setelah dilakukan normalisasi data ini dapat meminimalkan error.

Proses normalisasi data yang menggunakan metode *min-max scaler* dengan menggunakan *library MinMaxScaler* untuk merubah dataset ke dalam skala 0-1. Proses normalisasi data ini sangat penting karena membantu percepatan konvergensi dalam metode *XGBoost* sehingga pemrosesan dataset tidak membutuhkan waktu yang lama. Selain itu, normalisasi data ini juga berfungsi untuk stabilitas perhitungan sehingga nilai akurasi yang dihasilkan lebih stabil. Tahapan selanjutnya adalah merubah data menjadi data sequensial karena metode *XGBoost* ini tidak dapat memproses data biasa

Proses ini diharapkan dapat membantu *XGBoost* dalam menemukan pola hubungan antara topografi, cuaca, dan faktor fisik pengendara secara lebih akurat, serta meminimalkan pengaruh nilai ekstrim yang dapat mengganggu proses pembelajaran model

#### 4.2.3. Pembagian Data (*Data Splitting*)

Tahapan ini digunakan untuk memisahkan data *training* dan data *cleaning* dengan tujuan untuk menghindari overfitting yang akan mengganggu pemrosesan data dalam model *XGBoost*. Pembagian data ini juga diperlukan untuk memastikan model yang sudah dibuat dapat mengikuti dengan baik dan memberikan hasil prediksi yang akurat

Tabel 4. 2 Splitting Data

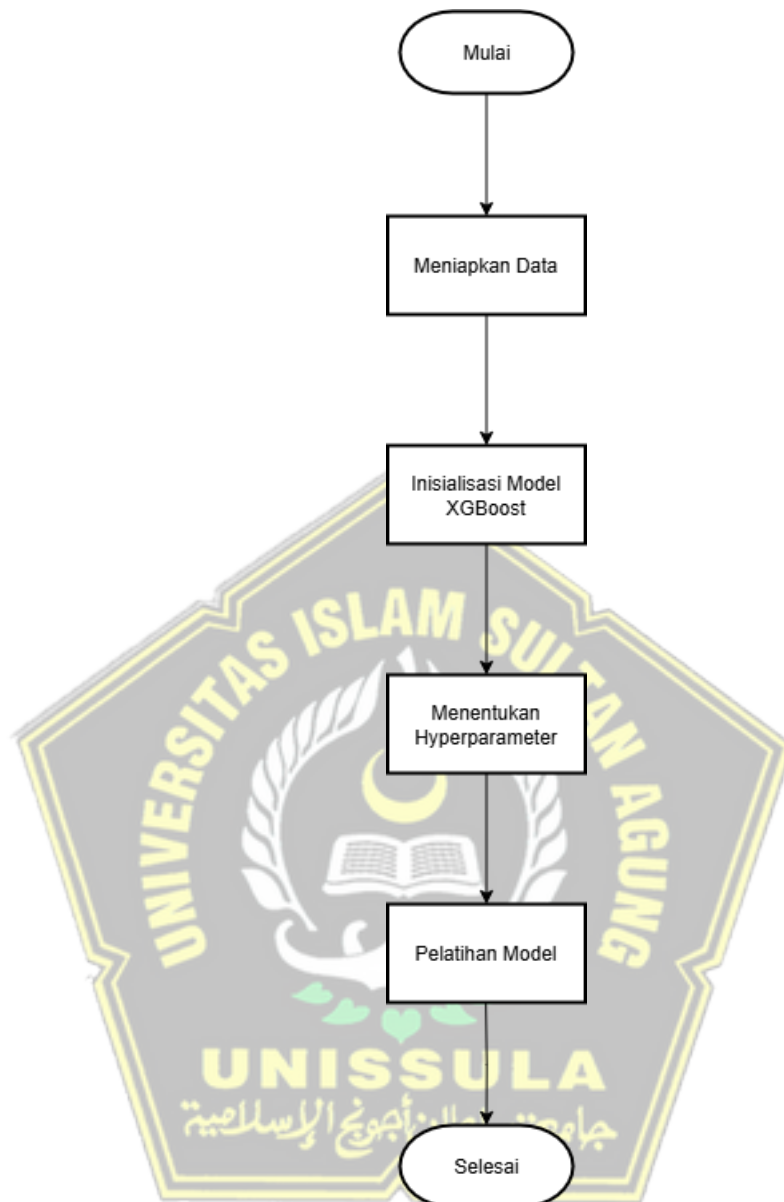
Dat/Kategori	<i>Data Splitting</i>	
	<i>Data Training</i>	<i>Data testing</i>
Elevasi	0.8	0.2
Cuaca	0.8	0.2
Waktu Tidur	0.8	0.2

Pada tabel 4.6 memberikan informasi untuk pembagian data yang akan digunakan dalam pemrosesan data yaitu dengan rasio 80:20 dimana 80% merupakan data *training* dan 20% digunakan untuk data testing.

### 4.3 Pembuatan Model

#### 4.4.1. Pembuatan Model *XGBoost*

Tahapan ini dilakukan proses pembuatan model machine learning menggunakan algoritma *Extreme Gradient Boosting (XGBoost)*. Pemilihan algoritma ini didasarkan pada kemampuannya dalam menangani data berukuran besar, memiliki performa yang baik dalam regresi, serta mampu mengelola hubungan non-linear antar variabel.



Gambar 4. 1 Alur Model *XGBoost*

Langkah pembuatan model *XGBoost* :

1. Persiapan Data untuk *XGBoost*: *XGBoost* memerlukan data dalam format tabular (baris dan kolom). Oleh karena itu, data deret waktu perlu diubah menjadi format *supervised learning*. Proses ini biasanya melibatkan pembuatan fitur (*feature engineering*) dari data masa lalu, seperti menggunakan harga penutupan hari sebelumnya ( $t-1$ ,  $t-2$ , dst.) dan nilai indikator teknikal sebagai fitur input untuk memprediksi harga hari ini ( $t$ ).



2. Inisialisasi Model *XGBoost*: Model *XGBoost* tidak dibangun lapis demi lapis seperti jaringan saraf. Sebagai gantinya, sebuah objek model diinisialisasi, yaitu *XGBRegressor*, karena tugasnya adalah prediksi regresi (memprediksi nilai kontinu).
3. Menentukan *Hyperparameter*: Sejumlah *hyperparameter* utama ditentukan untuk mengontrol kinerja dan kompleksitas model. Beberapa di antaranya adalah:

Tabel 4. 3 Hyperparameter

Model	Hyperparameter				
	n_estimator	learning rate	Max depth	sub sample dan colsample_bytree	Random state
<i>XGBoost</i>	300	0.2	3	0.8/0.8	42

- n-estimator : Jumlah tree (pohon keputusan) yang akan dibangun secara bertahap dalam boosting.
- learning\_rate : Kecepatan model dalam memperbarui bobot setiap kali menambahkan pohon baru.
- Max\_depth : Kedalaman maksimum tiap pohon Keputusan,
- subsample dan colsample\_bytree : Parameter ini digunakan untuk mengontrol random sampling agar model lebih general. Subsample persentase baris data yang diambil untuk membangun tiap pohon sedangkan colsampe\_bytree persentase kolom (fitur) yang diambil untuk mmebangun tiap pohon
- random\_state : seed angka acak yang digunakan agar proses pelatihan reproducible hasilnya sama jika dijalankan ulang.

Model *XGBoost* diinisialisasi sebagai *XGBRegressor* karena tugas yang diemban adalah *regresi* untuk memprediksi nilai kontinu, dengan *objective* diatur ke '*reg:squarederror*' agar model secara eksplisit mengoptimalkan *Mean Squared Error* (RMSE). Untuk membangun model ini, jumlah pohon keputusan (*n\_estimators*) ditetapkan sebanyak 1000, dengan laju belajar (*learning\_rate*) sebesar 0.05 yang memungkinkan model belajar secara perlahan namun lebih

robust. Guna mengontrol kompleksitas dan mencegah *overfitting*, kedalaman maksimum setiap pohon (*max\_depth*) dibatasi hingga 5. Selain itu, teknik regularisasi lebih lanjut diterapkan dengan mengatur parameter subsample dan *colsample\_bytree* ke 0.8. Hal ini berarti setiap pohon hanya menggunakan 80% sampel data dan 80% fitur secara acak, sehingga menambah keragaman dan kemampuan generalisasi model. Terakhir, *random\_state* diatur ke 42 untuk memastikan bahwa hasil pelatihan dapat diulang kembali dengan konsisten.

4. Melatih model: Model dilatih menggunakan fungsi `.fit()` pada data latih. Seringkali, teknik Early Stopping digunakan selama pelatihan, di mana performa model dipantau pada data validasi. Pelatihan akan berhenti jika tidak ada peningkatan performa setelah beberapa iterasi tertentu untuk mendapatkan model yang optimal.

#### 4.4.2. Evaluasi model

Setelah model *XGBoost* berhasil dilatih menggunakan data latih (training data), tahap selanjutnya adalah melakukan evaluasi untuk mengukur kinerja dan akurasi prediksi kecepatan rata-rata bersepeda. Evaluasi ini bertujuan untuk mengetahui sejauh mana model mampu memprediksi dengan benar pada data yang belum pernah digunakan selama proses pelatihan. Data yang digunakan untuk evaluasi adalah data uji (test data), sehingga hasilnya dapat mencerminkan kemampuan generalisasi model terhadap data baru.

Dalam penelitian ini, metrik evaluasi yang digunakan adalah Root Mean Squared Error (RMSE) dan R-Squared ( $R^2$ ). Hasil evaluasi menunjukkan bahwa model menghasilkan nilai RMSE sebesar 1.240. Nilai ini menunjukkan bahwa rata-rata kesalahan prediksi kecepatan model sekitar 1.240 dari nilai aktual. Semakin kecil nilai RMSE yang dihasilkan maka akurasi model semakin tinggi, dan sebaliknya jika RMSE besar maka tingkat kesalahan prediksi juga semakin besar.

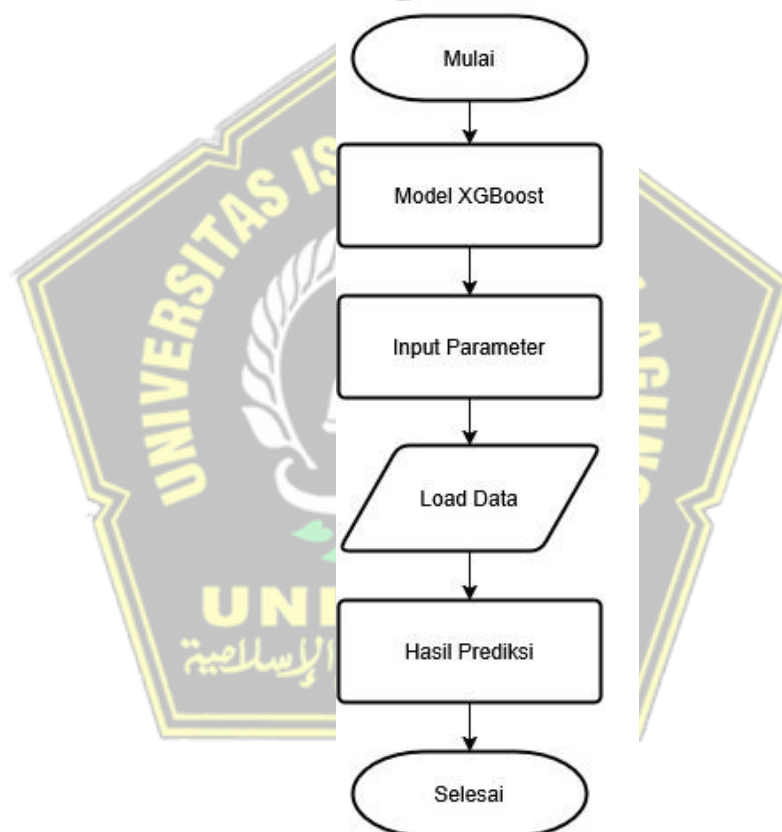
Selain RMSE, penelitian ini juga menggunakan metrik R-squared ( $R^2$ ) atau Koefisien Determinasi untuk mengukur kecocokan model (*goodness-of-fit*). Hasil evaluasi menunjukkan nilai  $R^2$  sebesar 0.800 atau 80%, yang berarti model *XGBoost* mampu menjelaskan sekitar 80% variasi data kecepatan rata-rata bersepeda berdasarkan faktor-faktor masukan seperti topografi, cuaca, dan waktu

tidur. Nilai  $R^2$  yang tinggi ini menandakan bahwa model memiliki performa yang baik dan dapat digunakan untuk melakukan prediksi dengan tingkat keandalan yang cukup tinggi.

Tabel 4. 4 Evaluasi Model

Model	RMSE	$R^2$
<i>XGBoost</i>	1.240	0.800

### 1. Perancangan Sistem



Gambar 4. 2 Perancangan Sistem

Berdasarkan flowchart yang telah dirancang diatas, Berdasarkan flowchart yang telah dirancang di atas, secara visual menggambarkan urutan alur kerja sistem prediksi dari awal hingga akhir. Proses dimulai dari Mulai, kemudian dilanjutkan dengan pemanggilan Model *XGBoost* yang telah dilatih sebelumnya. Selanjutnya, pengguna memasukkan parameter input seperti elevasi, jarak tempuh, durasi, waktu tidur, dan kondisi cuaca. Sistem kemudian melakukan proses load data untuk

mengolah parameter tersebut, dan menghasilkan hasil prediksi kecepatan rata-rata bersepeda. Tahap terakhir adalah Selesai, yang menandakan akhir dari alur kerja sistem.

#### 4.4.1. Wireframe Halaman Awal

##### 1. Hasil implementasi di aplikasi streamlit

Sebelum proses input parameter dilakukan, antarmuka awal dari sistem yang dibangun akan menampilkan halaman utama dengan tampilan sederhana dan informatif. Pada tampilan ini, pengguna diberikan petunjuk untuk memasukkan parameter seperti elevasi, jarak tempuh, curah hujan, jam tidur, jam mulai, dan kemudian melihat hasil prediksi kecepatan rata-rata bersepeda. Hal ini dirancang agar pengguna dapat dengan mudah memahami langkah awal dalam penggunaan sistem. Adapun tampilan antarmuka sebelum input parameter dapat dilihat pada gambar berikut:



Gambar 4. 3 Halaman Awal

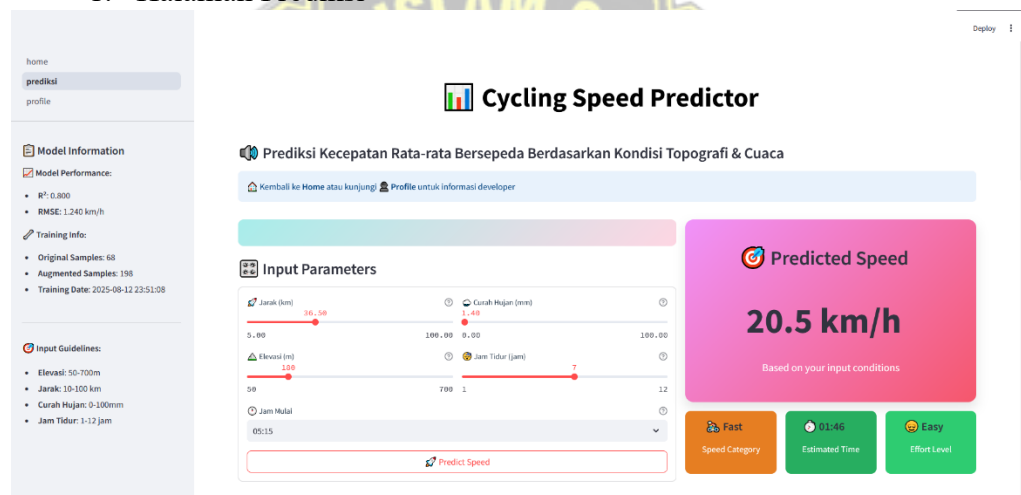
Pada gambar 4.3 menampilkan tampilan Streamlit sebelum pengguna memasukkan parameter pada aplikasi prediksi kecepatan rata-rata bersepeda. Tampilan ini menunjukkan halaman utama aplikasi yang memfasilitasi pengguna untuk mengisi nilai parameter seperti elevasi, jarak tempuh, curah hujan, jam tidur, dan jam mulai. Fitur utama pada halaman ini adalah form input yang dirancang agar pengguna dapat dengan mudah memasukkan data secara manual. Sistem ini

mendukung pengolahan data secara langsung di web tanpa perlu instalasi perangkat lunak tambahan, sehingga memudahkan proses penggunaan.

Secara keseluruhan, tampilan ini tidak hanya mencerminkan aspek fungsional dari sistem, tetapi juga menunjukkan kesiapan aplikasi dalam mendukung prediksi secara real-time berbasis web. Antarmuka yang ramah pengguna ini menjadi bagian penting dari keberhasilan implementasi model prediksi berbasis *XGBoost* yang telah dikembangkan sebelumnya untuk memperkirakan kecepatan rata-rata bersepeda berdasarkan faktor topografi dan cuaca.

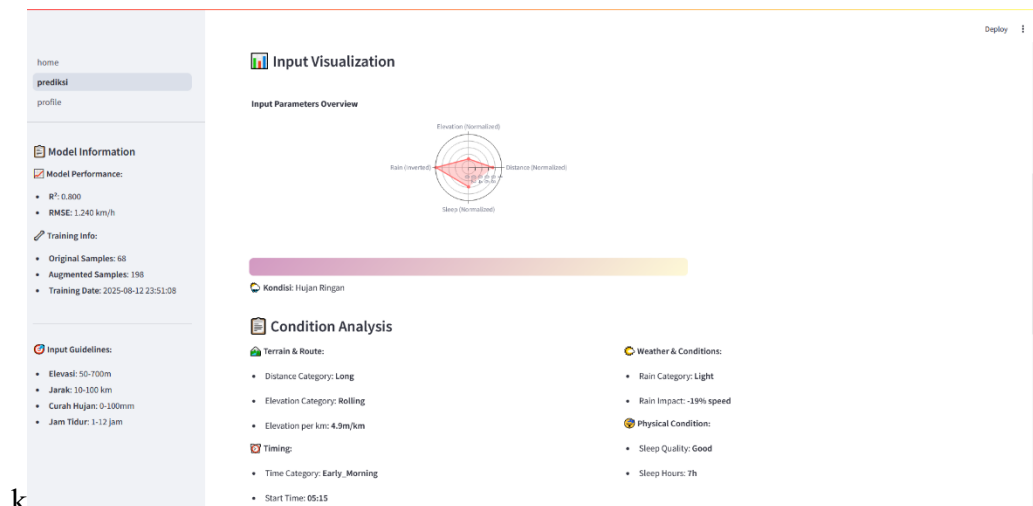
#### 4.4.2. Wireframe Halaman Hasil Prediksi

##### 1. Halaman Prediksi



Gambar 4. 4 Halaman Hasil Prediksi

Berdasarkan input parameter yang dimasukkan ke dalam aplikasi, diperoleh hasil prediksi kecepatan rata-rata bersepeda sebesar 20,5 km/jam. Input yang digunakan antara lain jarak tempuh sejauh 36.50 km, elevasi rute sebesar 100 m, curah hujan 1,40 mm, waktu tidur selama 7 jam, serta jam mulai aktivitas pukul 05:15 pagi. Dari hasil perhitungan model *XGBoost*, sistem mengkategorikan kecepatan yang dihasilkan dalam kategori Fast, dengan estimasi waktu tempuh sekitar 1 jam 46 menit. Tingkat kesulitan yang dihitung oleh sistem berada pada kategori Easy, sehingga kondisi rute dan faktor pendukung lainnya dinilai masih cukup ringan bagi pesepeda.



Gambar 4. 5 Penjelasan Parameter

Gambar tersebut menunjukkan tampilan antarmuka aplikasi prediksi kecepatan rata-rata bersepeda yang dibangun menggunakan framework Streamlit. Aplikasi ini berfungsi untuk memberikan prediksi kecepatan berdasarkan kondisi input pengguna seperti jarak tempuh, elevasi, cuaca, serta kualitas tidur. Pada bagian tengah aplikasi ditampilkan hasil prediksi berupa kecepatan rata-rata, yaitu 20.5 km/h, yang dihitung menggunakan model *XGBoost*. Prediksi ini dilengkapi dengan estimasi waktu tempuh (01:46) serta tingkat kesulitan (Easy) yang divisualisasikan dalam bentuk kotak informasi berwarna sehingga mudah dipahami pengguna.

Berdasarkan pengujian model *XGBoost* terhadap berbagai skenario bersepeda, diperoleh hasil prediksi kecepatan rata-rata serta estimasi waktu tempuh seperti ditunjukkan pada Gambar 4.5. Di bagian kiri layar terdapat informasi kinerja model, meliputi nilai evaluasi  $R^2 = 0.800$  dan  $RMSE = 1.240$  km/h. Selain itu juga ditampilkan detail data pelatihan seperti jumlah sampel asli, sampel hasil augmentasi, serta tanggal model terakhir dilatih. Bagian Condition Analysis menampilkan analisis kondisi yang memengaruhi hasil prediksi, seperti kategori jarak (Long), elevasi (Rolling), waktu mulai (Early Morning, 05:15), kondisi cuaca (No Rain, Cerah), serta kondisi fisik pengendara berdasarkan durasi tidur (7 jam dengan kualitas Good). Informasi ini membantu pengguna memahami faktor-faktor yang memengaruhi kecepatan rata-rata. Secara keseluruhan, aplikasi ini tidak hanya memberikan hasil prediksi kecepatan rata-rata bersepeda, tetapi juga menyajikan



informasi pendukung yang berguna untuk analisis performa dan perencanaan aktivitas bersepeda pengguna.

Hasil ini juga memperlihatkan bahwa aplikasi yang dibangun tidak hanya memberikan angka prediksi, tetapi juga analisis kondisi yang mempermudah interpretasi hasil oleh pengguna.

#### 4.4.3. Hasil dan Pembahasan

Tabel 4. 5 Hasil Prediksi

Elevasi (Meter)	Jarak Tempuh (km)	Curah Hujan	Jam tidur (jam)	Waktu Mulai	Hasil Prediksi Kecepatan (km/h)	Estimasi Waktu Menit
120 m	15.00 km	0	7 jam	06.00	28.0 km/h	31
120 m	15.00 km	10.50	7 jam	06.00	19.5 km/h	53
120m	15.00 km	0	4 jam	06.00	26.00 km/h	38
250 m	25.50 km	0	8 jam	05.15	22.3 km/h	01.08
250 m	25.50 km	6.90	8 jam	05.15	18.5 km/h	01.30
250 m	25.50 km	0	4 jam	05.15	19.3 km/h	01.23
170 m	31.50 km	0	7 jam	15.30	23.1 km/h	01.21
170 m	31.50 km	5.80	7 jam	15.30	20.3 km/h	01.45
170 m	31.50 km	0	4 jam	15.30	20.8 km/h	01.30
370 m	44.50 km	0	5 jam	13.45	15.0 km/h	02.46
370 m	44.50 km	7.60	5 jam	13.45	12.2 km/h	03.27
370 m	44.50 km	0	8 jam	13.45	17.6 km/h	01.36
80 m	37.50 km	0	7 jam	05.45	21.4 km/h	01.32
80 m	37.50 km	14.00	5 jam	05.45	17.5 km/h	02.08
80 m	37.50 km	0	4 jam	0505.45	19.1 km/h	01.51
410 m	48.00 km	0	7 jam	16.00	15.9 km/h	03.00
410 m	48.00 km	4.60	7 jam	16.00	13.7 km/h	03.24
410 m	48.00 km	0	4 jam	15.15	14.9 km/h	03.24

Berdasarkan pengujian model *XGBoost* terhadap berbagai skenario bersepeda, diperoleh hasil prediksi kecepatan rata-rata serta estimasi waktu tempuh seperti ditunjukkan pada Tabel 4.5.

1. Pengaruh Curah Hujan terhadap Kecepatan Dari data terlihat bahwa adanya curah hujan menurunkan kecepatan rata-rata pesepeda secara signifikan.

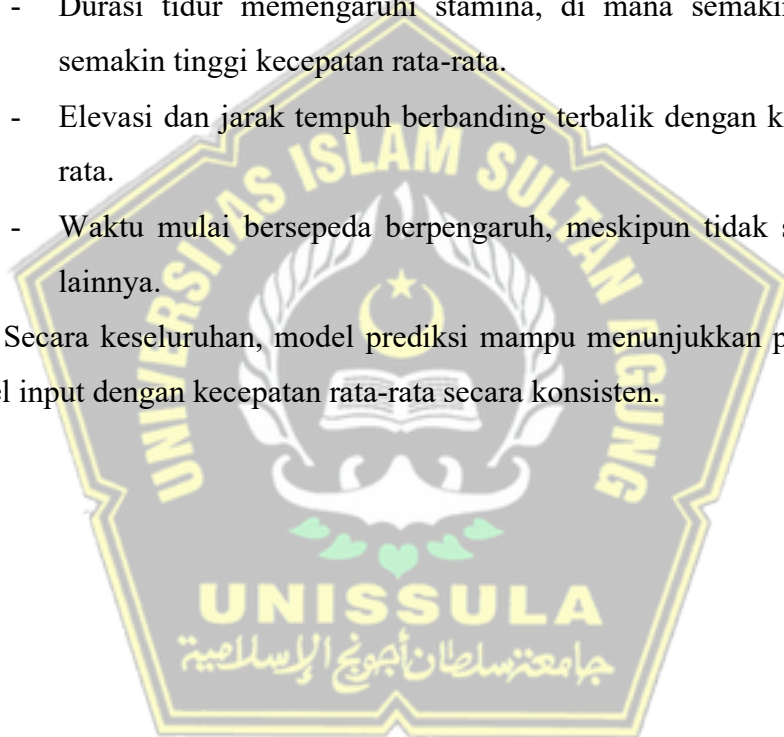
- Pada rute dengan jarak 15 km dan elevasi 120 m, kecepatan turun dari 28,0 km/h (tanpa hujan) menjadi 19,5 km/h (curah hujan 10,5 mm).
  - Pola serupa terlihat pada jarak 44,5 km dengan elevasi 370 m, di mana kecepatan berkurang dari 15,0 km/h menjadi 12,2 km/h saat terdapat curah hujan 7,6 mm. Hal ini menunjukkan bahwa curah hujan merupakan faktor eksternal yang cukup dominan dalam menurunkan kecepatan, karena kondisi jalan yang licin serta meningkatnya resistensi udara.
2. Pengaruh Jam Tidur terhadap Performa Waktu tidur yang cukup memberikan dampak positif pada kecepatan.
- Pada jarak 25,5 km dengan elevasi 250 m, pesepeda dengan tidur 8 jam memiliki kecepatan 22,3 km/h, sedangkan dengan tidur hanya 4 jam kecepatan turun menjadi 19,3 km/h.
  - Kasus lain pada jarak 48 km dengan elevasi 410 m, tidur 7 jam menghasilkan kecepatan 15,9 km/h, sedangkan dengan tidur 4 jam hanya 14,9 km/h. Hasil ini menguatkan bahwa durasi tidur yang lebih panjang meningkatkan stamina dan daya tahan, sehingga kecepatan rata-rata
3. Pengaruh Elevasi dan Jarak Tempuh Secara umum, semakin tinggi elevasi dan semakin jauh jarak tempuh, maka kecepatan rata-rata cenderung menurun.
- Jarak 15 km dengan elevasi 120 m mampu menghasilkan kecepatan hingga 28,0 km/h.
  - Namun pada jarak 48 km dengan elevasi 410 m, kecepatan rata-rata hanya sekitar 15,9 km/h. Hal ini sesuai dengan teori bahwa medan menanjak dan lintasan panjang meningkatkan beban kerja otot, sehingga memperlambat kecepatan rata-rata.
4. Pengaruh Waktu Mulai Aktivitas Waktu mulai bersepeda juga terlihat memengaruhi hasil, meskipun tidak sebesar curah hujan atau jam tidur.
- Pada rute 31,5 km dengan elevasi 170 m, kecepatan saat mulai sore (15.30) adalah 23,1 km/h (7 jam tidur) atau 20,8 km/h (4 jam tidur).

- Sementara itu pada rute pagi hari (05.15) dengan kondisi yang sama (jarak dan elevasi), kecepatan hanya 22,3 km/h (8 jam tidur) atau 19,3 km/h (4 jam tidur). Hal ini menunjukkan bahwa faktor waktu mulai mungkin terkait dengan kondisi cuaca, suhu udara, serta ritme biologis pesepeda.

#### Kesimpulan Pembahasan

- Curah hujan terbukti sebagai faktor paling dominan yang menurunkan kecepatan.
- Durasi tidur memengaruhi stamina, di mana semakin lama tidur, semakin tinggi kecepatan rata-rata.
- Elevasi dan jarak tempuh berbanding terbalik dengan kecepatan rata-rata.
- Waktu mulai bersepeda berpengaruh, meskipun tidak sebesar faktor lainnya.

Secara keseluruhan, model prediksi mampu menunjukkan pola hubungan variabel input dengan kecepatan rata-rata secara konsisten.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian dan implementasi sistem prediksi kecepatan rata-rata bersepeda menggunakan algoritma *XGBoost*, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Model prediksi berbasis *XGBoost* mampu memberikan hasil estimasi kecepatan rata-rata bersepeda dengan tingkat akurasi yang cukup baik. Hal ini ditunjukkan dengan nilai performa model yaitu RMSE sebesar 1,240 km/jam, dan  $R^2$  sebesar 0,800, yang menandakan model cukup representatif dalam memprediksi kecepatan berdasarkan data input.
2. Faktor-faktor yang paling berpengaruh terhadap kecepatan rata-rata bersepeda adalah jarak tempuh, elevasi, durasi aktivitas, curah hujan, dan waktu tidur. Kombinasi antara kondisi topografi dan faktor cuaca terbukti memberikan pengaruh signifikan terhadap performa pesepeda.
3. Aplikasi yang dibangun menggunakan Streamlit berhasil menampilkan hasil prediksi kecepatan secara interaktif, dilengkapi dengan visualisasi input, analisis kondisi rute, dan kategorisasi tingkat kecepatan serta tingkat kesulitan. Hal ini memudahkan pengguna dalam memahami hasil prediksi.
4. Integrasi data Strava dengan faktor cuaca dan kondisi tidur memberikan gambaran yang lebih komprehensif terhadap performa bersepeda, sehingga aplikasi ini berpotensi digunakan sebagai alat bantu perencanaan aktivitas bersepeda bagi pengguna.

## 5.2 Saran

Adapun saran untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Penambahan Variabel Input Perlu ditambahkan faktor lain seperti kecepatan angin, suhu udara, tingkat kelembaban, serta kondisi lalu lintas untuk menghasilkan prediksi yang lebih akurat.
2. Peningkatan Ukuran Dataset Dataset yang digunakan masih relatif terbatas. Oleh karena itu, penelitian selanjutnya sebaiknya menggunakan jumlah data yang lebih besar dan beragam agar model lebih robust dan dapat digeneralisasi.
3. Penggunaan Algoritma Lain untuk Perbandingan Selain *XGBoost*, dapat dilakukan perbandingan dengan algoritma lain seperti Random Forest, LightGBM, atau bahkan metode berbasis deep learning untuk mengetahui model terbaik dalam konteks prediksi ini.
4. Pengembangan Aplikasi Mobile Aplikasi yang sudah berbasis web dapat dikembangkan lebih lanjut ke platform mobile (Android/iOS) agar lebih praktis digunakan saat bersepeda.
5. Integrasi Real-time Data Aplikasi dapat diintegrasikan langsung dengan perangkat GPS atau sensor IoT, sehingga prediksi kecepatan dapat dilakukan secara real-time berdasarkan kondisi aktual pengguna.

## DAFTAR PUSTAKA

- Abdullah, Mohammad Aldinugroho. 2025. "Prediksi Jumlah Pasien Medical Check Up Berdasarkan Time Series Forecasting Menggunakan Algoritma *XGBoost*." *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)* 6(2): 488–97.
- Al Anshori Rosid Fachrudin. 2013. "Komunitas Sepeda Gowes Menjadi Gaya Hidup Sehat." *Journal of Chemical Information and Modeling* 53(9): 1689–99. <http://eprints.ums.ac.id/51455/>.
- Arsyad, Muhammad, and Nur Wahyuni. 2021. "Modifikasi Sepeda Konvensional Menjadi Sepeda Listrik." *Prosiding 5th Seminar Nasional Penelitian & Pengabdian Kepada Masyarakat 2021*: 40–43. <https://www.polygonbikes.com/id/mengenal-sejarah-dan-munculnya-sepeda-di-indonesia/>.
- Astutiningsih, Tiyas, Dewi Retno Sari Saputro, and Sutanto. 2023. "Optimasi Algoritme Xtreme Gradient Boosting (*XGBoost*) Pada Harga Saham PT. United Tractors Tbk." *SPECTA Journal of Technology* 7(3): 632–41.
- Herlambang, Junifer, and Asep Sujana Wahyuri. 2023. "Tinjauan Kecepatan Bersepeda Atlet Pengprov ISSI ( Ikatan Sport Sepeda Indonesia ) Sumatera Barat." 6(11): 62–71.
- Hermawan, Reyhan Dhika, Rony Mohamad Rizal, and M Saleh Paembonan. 2024. "Prosiding Seminar Nasional Pendidikan Olahraga Prodi Pendidikan Jasmani Kesehatan Dan Rekreasi STKIP Pasundan , 26-27 Juni 2024 Faktor Penyebab Menurunnya Performa Atlet Balap Sepeda Road Bike Pada Saat Perlombaan Prosiding Seminar Nasional Pendidikan Ola." : 26–27.
- ICHWAN, MUHAMMAD, and SALMAN FATHUL ALFARISYI. 2024. "Pemodelan Ritme Kalori Terbakar Setiap Waktu Selama Bersepeda Dengan Feedforward Neural Network." *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika* 12(1): 13.
- Irawati, Andi Febi, Irwin, and Andi Muh Anshar. 2024. "Kepopuleran Olahraga Sepeda Sebagai Bagian Dari Pola Hidup Sehat." *Jurnal Ilmiah Global Education* 5(3): 2038–43.
- Kim, Jongsung et al. 2022. "Development of a Deep Learning-Based Prediction Model for Water Consumption at the Household Level." *Water (Switzerland)* 14(9).
- Kurniasih, Diena, and Diena Kurniasih Bappeda Kota Yogyakarta. 2023. "Pemenuhan Jaminan Perlindungan Pesepeda Di Kota Sepeda." *Jurnal Ilmu Hukum dan Sosial* 1(3): 319–38. <https://doi.org/10.51903/hakim.v1i3.1305>.
- Kurniawan, Wildan, and Uce Indahyanti. 2024. "Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode *XGBoost*." *Indonesian Journal of Applied Technology* 1(2): 18.



- Lisanthoni, Angela, Fitri Indah Sari, Ellexia Leonie Gunawan, and Chelsea Ayu Adhigiadany. 2023. "Model Prediksi Kepadatan Lalu Lintas: Perbandingan Algoritma Random Forest Dan *XGBoost*." *Prosiding Seminar Nasional Sains Data* 3(1): 296–303.
- Murdiansyah, Danang Triantoro. 2024. "Prediksi Stroke Menggunakan Extreme Gradient Boosting." *JIKO (Jurnal Informatika dan Komputer)* 8(2): 419.
- Pendidikan, Jurnal et al. 2021. "The Level Of Motivation Parents To Involve Their Children In Cycling Activities Tingkat Motivasi Orang Tua Mengikutsertakan Anaknya Pada Aktivitas Bersepeda Physical Activity ( GAPP 2018-2030 ( More Active People For A Healthier." 5(November): 126–36.
- Rambe, Rizki Ridwan, Asep Sujana Wahyuri, Emral Emral, and Haripah Lwanis. 2023. "Tinjauan Kondisi Fisik Pemain Sepakbola Sekolah Keberbakatan Olahraga U-16 Sumatera Barat." *Jurnal Pendidikan dan Olahraga* 6(7): 111–18.
- Rupaka, Anggun Prima, Aris Budi Sulisty, and Dewa Punia. 2021. "Pemetaan Perilaku Pesepeda Pra Dan Pasca Pandemi Covid-19 Di Provinsi Bali Menggunakan Data Strava Metro." *Jurnal Teknologi Transportasi dan Logistik* 2(2): 119–26.
- Samodra, Touvan Juni et al. 2022. "Profil Hb Dan Vo2 Max Atlet Balap Sepeda Pra Pon." *Journal of Sport Science and Fitness* 8(1): 41–48.
- Saputra, Aldo Andika et al. 2024. "Penerapan Algoritma *Extreme Gradient Boosting (XGBoost)* Untuk Analisis Risiko Kredit." *Jurnal Ilmiah Wahana Pendidikan* 10(7): 27–36. <https://doi.org/10.5281/zenodo.10960080>.
- Su, A. 2018. "Bab II Landasan Teori." *Journal of Chemical Information and Modeling* 53(9): 8–24.
- Surojo, Saka Surya, Fajar Widiyatmoko, and Pandu Kresnapati. 2022. "Survei Antusiasme Dan Ketertarikan Masyarakat Dalam Bersepeda Di Kota Semarang." *Journal of Sport Science and Fitness* 8(1): 63–68.
- Susanto, Erliyan Redy, and Agum Cahyana. 2025. "Penerapan Algoritma *XGBoost* Untuk Prediksi Diabetes: Analisis Confusion Matrix Dan ROC Curve." *Fountain of Informatics Journal* 10(1): 40–50.
- Utomo, Andi Wahyu. 2016. "濟無No Title No Title No Title." *Angewandte Chemie International Edition*, 6(11), 951–952. 1(0): 1–23.
- Widodo, Anindya Pinasih, and Muhammad. 2023. "Profil Kondisi Fisik Atlet Balap Sepeda Jalana Raya Puslatcab Issi Surabaya Dalam Rangka Persiapan Porprov Jawa Timur 2022." *Jurnal Prestasi Olahraga* 6(1): 30–35.
- Yasmin, Nabila Rahyan. 2022. "Gambaran Hasil Skrining Hipotiroid Kongenital (SHK) Berdasarkan Topografi Wilayah Di Kota Bandar Lampung Pada Bulan Mei - Oktober Tahun 2019." *Skripsi* 2019.



**HALAMAN LAMPIRAN**