

**PENERAPAN ALGORITMA *RANDOM FOREST* UNTUK PREDIKSI IPK
MAHASISWA BERDASARKAN POLA GAYA HIDUP HARIAN**

SKRIPSI

Oleh :

WILLYAN IMARTA
NIM. 210605110092



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**PENERAPAN ALGORITMA *RANDOM FOREST* UNTUK PREDIKSI IPK
MAHASISWA BERDASARKAN POLA GAYA HIDUP HARIAN**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
WILLYAN IMARTA
NIM. 210605110092

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN

**PENERAPAN ALGORITMA *RANDOM FOREST* UNTUK PREDIKSI IPK
MAHASISWA BERDASARKAN POLA GAYA HIDUP HARIAN**

SKRIPSI

Oleh :
WILLYAN IMARTA
NIM. 210605110092

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 06 Juni 2025

Pembimbing I,



Syahiduz Zaman, M.Kom
NIP. 19700502 200501 1 005

Pembimbing II,



Dr. Totok Chamidy, M.Kom
NIP. 19691222 200604 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Ir. Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

**PENERAPAN ALGORITMA *RANDOM FOREST* UNTUK PREDIKSI IPK
MAHASISWA BERDASARKAN POLA GAYA HIDUP HARIAN**

SKRIPSI

Oleh :
WILLYAN IMARTA
NIM. 210605110092

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 25 Juni 2025

Susunan Dewan Penguji

Ketua Penguji : Dr. M. Ainul Yaqin, M.Kom
NIP. 19761013 200604 1 004

Anggota Penguji I : Dr. Agung Teguh Wibowo Almais, M.T
NIP. 19860301 202321 1 016

Anggota Penguji II : Syahiduz Zaman, M.Kom
NIP. 19700502 200501 1 005

Anggota Penguji III : Dr. Totok Chamidy, M.Kom
NIP. 19691222 200604 1 001

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Ir. Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Willyan Imarta
NIM : 210605110092
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Penerapan Algoritma *Random Forest* untuk
Prediksi IPK Mahasiswa berdasarkan pola gaya
hidup harian.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 25 Juni 2025
Yang membuat pernyataan,



Willyan Imarta
NIM.210605110092

MOTTO

“Whatever you are, be a good one”

(Abraham Lincoln)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas limpahan karunia, rahmat, dan kasih sayang-Nya yang senantiasa mengiringi setiap langkah penulis hingga terselesaikannya skripsi ini. Karya ini penulis persembahkan kepada:

Ibu tercinta,

Sumber kasih yang tak pernah habis, yang selalu memberikan dukungan, doa, serta semangat tanpa henti. Terima kasih telah menjadi cahaya yang menerangi setiap langkah penulis.

Ayah tercinta,

Teladan dalam keteguhan dan kerja keras, yang selalu memberikan kasih sayang tulus serta menjadi inspirasi dalam setiap perjalanan hidup penulis.

Kakak Perempuan tercinta,

Yang menjadi inspirasi lewat kebijaksanaan dan perhatian, serta memberikan dorongan semangat dan kebaikan yang tak terhingga.

Untuk diri sendiri,

Terima kasih telah terus berjuang, tetap kuat dalam menghadapi setiap rintangan, dan menyelesaikan semua proses ini dengan kesabaran. Semoga karya ini menjadi langkah awal menuju perjalanan yang lebih bermakna.

KATA PENGANTAR

Segala puji syukur senantiasa penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Algoritma *Random Forest* Untuk Prediksi Ipk Mahasiswa Berdasarkan Pola Gaya Hidup Harian” ini dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, suri teladan umat manusia, beserta keluarga, sahabat, dan pengikutnya hingga akhir zaman.

Ucapan terima kasih sebanyak-banyaknya saya ucapkan kepada:

1. Prof. Dr. H. M. Zainuddin, MA selaku Rektor UIN Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT, IPU, selaku Kepala Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
4. Ajib Hanani, M.T, selaku dosen wali yang telah mendampingi saya secara akademik selama perkuliahan, serta memberikan bimbingan dan dukungan yang sangat berarti dalam proses studi saya.
5. Syahiduz Zaman, M. Kom. selaku dosen pembimbing pertama skripsi yang telah membimbing dan memberikan arahan dari awal perkuliahan hingga proses penyusunan skripsi ini.
6. Dr. Totok Chamidy, M.Kom, selaku dosen pembimbing kedua yang dengan penuh perhatian memberikan masukan, motivasi, dan kritik membangun demi penyempurnaan karya ini.
7. Dr. M. Ainul Yaqin, M.Kom, selaku Ketua Dewan Penguji, atas saran dan koreksi yang sangat berharga dalam proses ujian skripsi.

8. Dr. Agung Teguh Wibowo Almais, M.T, selaku penguji kedua, atas masukan dan evaluasi yang telah meningkatkan kualitas dari penelitian ini.
9. Kedua orang tua tercinta, Ibu dan Ayah, serta kakak Perempuan saya, yang selalu menjadi sumber doa, kekuatan, dan inspirasi yang tak ternilai.
10. Seluruh dosen Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, atas ilmu dan dedikasi yang telah diberikan selama masa perkuliahan.
11. Dua sahabat terdekat saya, Adam dan Ammil, yang selalu memberikan semangat, bantuan, serta menjadi tempat berbagi dalam setiap tantangan selama proses penyusunan skripsi ini.
12. Teman-teman terdekat penulis, Gilang, Bambang, Faiz, dan Hasan, kehadiran kalian telah memberikan warna tersendiri dalam perjalanan ini. Bersama kalian, penulis tidak pernah merasa sendiri, bahkan di tengah kesulitan sekalipun.
13. Rekan-rekan seperjuangan Jurusan Teknik Informatika angkatan 2021, serta teman-teman satu bimbingan skripsi, atas dukungan, kerja sama, dan semangat yang saling menguatkan.
14. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan dukungan secara langsung maupun tidak langsung.

Penulis telah berupaya semaksimal mungkin dalam menyusun skripsi ini. Namun, penulis menyadari bahwa karya ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang.

Malang, 10 Juni 2025

Penulis

DAFTAR ISI

HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
ABSTRAK	xiv
ABSTRACT	xv
المخلص	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	6
BAB II STUDI PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Indeks Prestasi Akademik (IPK).....	10
2.3 Gaya Hidup	11
2.4 Normalisasi <i>Z-score Scaling</i>	12
2.5 <i>Random Forest</i>	13
BAB III DESAIN DAN IMPLEMENTASI	15
3.1 Alur Penelitian	15
3.2 Pengumpulan Data	15
3.3 Desain Sistem.....	17
3.4 <i>Preprocessing</i>	18
3.4.1 Pembersihan Data	18
3.4.2 Transformasi Data.....	19
3.4.3 Split Data	19
3.5 Implementasi <i>Random Forest</i>	20
3.6 Evaluasi Model	23
3.7 Skenario Pengujian	25
BAB IV HASIL DAN PEMBAHASAN	27
4.1 Dataset.....	27
4.2 Data <i>Preprocessing</i>	27
4.2.1 <i>Data Cleaning</i>	28
4.2.2 Data Normalisasi (<i>Z-Score Scaling</i>)	28
4.2.3 Split Data	29
4.3 Metode <i>Random Forest</i>	30
4.3.1 <i>Bootstrap Sampling</i>	30
4.3.2 <i>Feature Randomness</i>	31
4.3.3 Split Terbaik	31

4.3.4	Membangun <i>Decision Tree</i>	32
4.3.5	Agregasi Prediksi.....	33
4.3.6	Evaluasi Model.....	34
4.4	Hasil Pengujian <i>Hyperparameter</i>	35
4.5	Hasil Pengujian Kombinasi Fitur.....	38
4.5.1	Hasil Kombinasi Model A.....	39
4.5.2	Hasil Kombinasi Model B.....	40
4.5.3	Hasil Kombinasi Model C.....	41
4.5.4	Hasil Kombinasi Model D.....	43
4.5.5	Hasil Kombinasi Model E.....	44
4.5.6	Hasil Kombinasi Model F.....	46
4.5.7	Hasil Kombinasi Model G.....	47
4.5.8	Hasil Kombinasi Model H.....	49
4.5.9	Hasil Kombinasi Model I.....	50
4.5.10	Hasil Kombinasi Model J.....	52
4.5.11	Hasil Kombinasi Model K.....	53
4.6	Hasil Pengujian <i>Feature Importance</i>	56
4.7	Pembahasan.....	57
4.8	Integrasi Penelitian.....	60
	BAB V KESIMPULAN DAN SARAN.....	63
5.1	Kesimpulan.....	63
5.2	Saran.....	64
	DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 3. 1 Alur penelitian.....	15
Gambar 3. 2 Desain sistem	17
Gambar 4. 1 Hasil salah satu decision tree	32
Gambar 4. 2 Hasil visual agregrasi prediksi	33
Gambar 4. 3 Grafik hasil prediksi parameter terbaik.....	37
Gambar 4. 4 Grafik perbandingan evaluasi model	38
Gambar 4. 5 Hasil visualisasi model A.....	39
Gambar 4. 6 Hasil visualisasi model B.....	41
Gambar 4. 7 Hasil visualisasi model C	42
Gambar 4. 8 Hasil visualisasi model D.....	44
Gambar 4. 9 Hasil visualisasi model E	45
Gambar 4. 10 Hasil visualisasi model F	47
Gambar 4. 11 Hasil visualisasi model G.....	48
Gambar 4. 12 Hasil visualisasi model H.....	50
Gambar 4. 13 Hasil visualisasi model I	51
Gambar 4. 14 Hasil visualisasi model J	53
Gambar 4. 15 Hasil visualisasi model K.....	54
Gambar 4. 16 Hasil Visual Feature Importance.....	57
Gambar 4. 17 Hasil evaluasi MAE setiap model kombinasi	58
Gambar 4. 19 Feature Importance.....	60

DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu.....	9
Tabel 3. 1 Atribut Dataset	16
Tabel 3. 2 Contoh hasil Bootstrap sampling	20
Tabel 3. 3 Contoh pemilihan fitur secara acak.....	21
Tabel 3. 4 Contoh tabel split di 7.85	22
Tabel 3. 5 Parameter yang digunakan	25
Tabel 3. 6 Fitur yang digunakan untuk pengujian.....	26
Tabel 4. 1 Dataset.....	27
Tabel 4. 2 Hasil data cleaning	28
Tabel 4. 3 Hasil normalisasi Z-score scaling	29
Tabel 4. 4 Hasil salah satu Bootstrap sampling	31
Tabel 4. 5 Hasil salah satu split terbaik.....	31
Tabel 4. 6 Hasil agregasi prediksi	33
Tabel 4. 7 Hasil evaluasi model	34
Tabel 4. 8 Hasil kombinasi parameter terbaik	35
Tabel 4. 9 Hasil prediksi parameter terbaik	36
Tabel 4. 10 Hasil evaluasi parameter terbaik dan parameter default	37
Tabel 4. 11 Hasil prediksi model A	39
Tabel 4. 12 Hasil evaluasi model A	40
Tabel 4. 13 Hasil Prediksi Model B	40
Tabel 4. 14 Hasil evaluasi model B	41
Tabel 4. 15 Hasil prediksi model C.....	42
Tabel 4. 16 Hasil evaluasi model C	42
Tabel 4. 17 Hasil prediksi model D	43
Tabel 4. 18 Hasil evaluasi model D	44
Tabel 4. 19 Hasil prediksi model E.....	45
Tabel 4. 20 Hasil evaluasi model E.....	45
Tabel 4. 21 Hasil prediksi model F	46
Tabel 4. 22 Hasil evaluasi model F	47
Tabel 4. 23 Hasil prediksi model G	47
Tabel 4. 24 Hasil evaluasi model G	48
Tabel 4. 25 Hasil prediksi model H	49
Tabel 4. 26 Hasil evaluasi model H	50
Tabel 4. 27 Hasil prediksi model I	51
Tabel 4. 28 Hasil evaluasi model I.....	52
Tabel 4. 29 Hasil prediksi model J.....	52
Tabel 4. 30 Hasil evaluasi model J.....	53
Tabel 4. 31 Hasil prediksi model I	54
Tabel 4. 32 Hasil evaluasi model K	55
Tabel 4. 33 Perbandingan hasil evaluasi setiap model.....	55
Tabel 4. 34 Hasil Feature Importance	56

ABSTRAK

Imarta, Willyan. 2025. **Penerapan Algoritma *Random Forest* untuk Prediksi IPK Mahasiswa Berdasarkan Pola Gaya Hidup Harian**. Skripsi, Program Studi Teknik Informatika Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Syahiduz Zaman, M.Kom (II) Dr. Totok Chamidy, M.Kom.

Kata Kunci: *Random Forest*, Prediksi IPK, Gaya Hidup Mahasiswa, *Machine Learning*, *Data Mining*.

Penelitian ini bertujuan untuk memprediksi Indeks Prestasi Kumulatif (IPK) mahasiswa berdasarkan pola gaya hidup harian menggunakan algoritma *Random Forest*. Lima variabel utama yang dianalisis adalah durasi belajar, durasi tidur, aktivitas fisik, interaksi sosial, dan kegiatan ekstrakurikuler. Data diperoleh dari dataset Kaggle yang telah dimodifikasi, dan dilakukan pra-proses berupa pembersihan data, normalisasi *Z-score*, serta pembagian data latih dan uji dengan rasio 80:20. Evaluasi model dilakukan menggunakan metrik R^2 , MAE, dan RMSE. Hasil penelitian menunjukkan bahwa *Random Forest* mampu memprediksi IPK secara cukup akurat dengan nilai R^2 sebesar 0.52, MAE sebesar 0.168, dan RMSE sebesar 0.044. Pengujian *feature importance* menunjukkan bahwa pola belajar memberikan pengaruh paling besar terhadap prediksi IPK. Hasil ini menunjukkan bahwa gaya hidup mahasiswa sangat berperan dalam pencapaian akademik dan pendekatan machine learning dapat dimanfaatkan untuk memberikan rekomendasi gaya hidup yang lebih baik guna meningkatkan prestasi mahasiswa.

ABSTRACT

Imarta, Willyan. 2025. **Application of the Random Forest Algorithm for Predicting Student GPA Based on Daily Lifestyle Patterns**. Thesis, Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Supervisors: (I) Syahiduz Zaman, M.Kom (II) Dr. Totok Chamidy, M.Kom.

Keywords: *Random Forest, GPA Prediction, Student Lifestyle, Machine Learning, Data Mining.*

This study aims to predict university students' Grade Point Average (GPA) based on daily lifestyle patterns using the Random Forest algorithm. The five main variables analyzed include study duration, sleep duration, physical activity, social interaction, and extracurricular activities. The dataset used was sourced and modified from Kaggle, and preprocessing steps included data cleaning, Z-score normalization, and an 80:20 training-testing split. The model evaluation employed R^2 , MAE, and RMSE metrics. The results indicate that Random Forest is able to predict GPA with reasonable accuracy, achieving an R^2 of 0.52, MAE of 0.168, and RMSE of 0.044. The feature importance analysis revealed that study patterns had the most significant influence on GPA prediction. These findings suggest that students' lifestyles play an essential role in academic performance, and machine learning approaches such as Random Forest can be utilized to generate lifestyle recommendations to help students enhance their academic achievements.

الملخص

إمارتا، ويليان. ٢٠٢٥. تطبيق خوارزمية الغابة العشوائية لتنبؤ المعدل التراكمي للطلبة بناءً على أنماط نمط حياتهم اليومية. بحث تخرج، برنامج هندسة المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانغ. المشرفان: (١) شهيد الزمان، ماجستير في علوم الحاسوب، (٢) الدكتور توتوك شاميدي، ماجستير في علوم الحاسوب.

الكلمات المفتاحية: الغابة العشوائية، تنبؤ المعدل التراكمي، نمط حياة الطلاب، تعلم الآلة، التنقيب عن البيانات.

يهدف هذا البحث إلى التنبؤ بالمعدل التراكمي (GPA) للطلاب الجامعيين بناءً على أنماط حياتهم اليومية باستخدام خوارزمية الغابة العشوائية (Random Forest). تم تحليل خمسة متغيرات رئيسية تشمل مدة الدراسة، مدة النوم، النشاط البدني، التفاعل الاجتماعي، والأنشطة اللا منهجية. تم الحصول على مجموعة البيانات من منصة Kaggle بعد تعديلها، وتم تنفيذ معالجة أولية شملت تنظيف البيانات، وتطبيع Z-score، وتقسيم البيانات بنسبة 80:20 للتدريب والاختبار. تم تقييم النموذج باستخدام مقاييس R^2 و MAE و RMSE. أظهرت النتائج أن خوارزمية Random Forest تستطيع التنبؤ بالمعدل التراكمي بدقة جيدة، حيث بلغت قيمة R^2 نحو 0.52، و MAE مقدار 0.168، و RMSE بمقدار 0.044. أظهرت نتائج أهمية السمات أن نمط الدراسة له التأثير الأكبر على التنبؤ بالمعدل. تشير هذه النتائج إلى أن نمط حياة الطالب له دور كبير في التحصيل الأكاديمي، ويمكن استخدام تقنيات التعلم الآلي لتقديم توصيات تساعد الطلاب على تحسين عاداتهم لتحقيق أداء أكاديمي أفضل.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Prestasi akademik mahasiswa merupakan salah satu indikator yang sangat penting dalam menilai keberhasilan suatu proses pendidikan di tingkat perguruan tinggi (Sakız et al., 2021). Indeks Prestasi Kumulatif (IPK) sering dijadikan parameter utama untuk mengukur tingkat pencapaian mahasiswa. IPK tidak hanya menjadi tolak ukur bagi kesuksesan akademik individu, tetapi juga memiliki implikasi yang lebih luas, seperti peluang mendapatkan pekerjaan, melanjutkan pendidikan ke jenjang lebih tinggi, atau mengakses program pengembangan profesional lainnya. Namun demikian, pencapaian IPK tidak sepenuhnya bergantung pada kemampuan akademik semata, melainkan juga dipengaruhi oleh berbagai faktor eksternal. Salah satu faktor penting adalah pola gaya hidup mahasiswa, yang meliputi kebiasaan belajar, durasi tidur, tingkat aktivitas fisik, hingga keterlibatan dalam kegiatan social (Lone, 2021). Gaya hidup mahasiswa dapat memengaruhi konsentrasi, kesehatan mental, dan produktivitas belajar, sehingga memiliki dampak langsung maupun tidak langsung pada hasil akademik mereka (Shafie et al., 2022).

Di era digital seperti saat ini, mahasiswa sering kali menghadapi tantangan untuk menjaga keseimbangan antara kehidupan akademik dan kehidupan pribadi mereka (Adriansyah & Muhliansyah, 2022). Beberapa mahasiswa mungkin memilih untuk menghabiskan lebih banyak waktu belajar, sementara yang lain

lebih memprioritaskan aktivitas sosial atau kegiatan ekstrakurikuler. Pola ini menciptakan dinamika yang unik dan kompleks dalam gaya hidup mahasiswa. Misalnya, kebiasaan begadang, konsumsi kafein berlebih, pola makan yang tidak teratur, dan minimnya aktivitas fisik merupakan bentuk pola hidup tidak sehat yang dapat berdampak negatif terhadap performa akademik (Himanshu J. Dhide Tycs et al., 2025). Penelitian menunjukkan bahwa gaya hidup yang buruk, seperti kurang tidur dan kurang olahraga, secara signifikan berkontribusi terhadap penurunan fungsi kognitif dan tingkat fokus mahasiswa, yang akhirnya berpengaruh pada capaian Indeks Prestasi Kumulatif (IPK) mereka (Seoane et al., 2020). Di sisi lain, aktivitas fisik yang teratur, seperti olahraga, diketahui dapat meningkatkan daya tahan tubuh dan kesehatan mental, yang pada gilirannya berkontribusi terhadap pencapaian akademik (Teferi, 2020). Hubungan antara faktor-faktor ini, meskipun terlihat jelas, sering kali sulit untuk dipahami secara menyeluruh. Sebagian besar penelitian sebelumnya cenderung fokus pada pengaruh satu variabel secara terpisah, tanpa mempertimbangkan interaksi antar variabel.

Dalam penelitian ini, peneliti memilih lima variabel utama yaitu durasi belajar, durasi tidur, aktivitas fisik, keterlibatan sosial, dan keterlibatan dalam kegiatan ekstrakurikuler yang mana dianggap paling signifikan dalam mempengaruhi pencapaian akademik mahasiswa. Durasi belajar berkaitan langsung dengan penguasaan materi, sedangkan durasi tidur berperan penting dalam pemulihan fisik dan kemampuan konsentrasi. Aktivitas fisik diketahui meningkatkan kesehatan dan ketahanan mental, yang mendukung produktivitas akademik. Keterlibatan sosial membantu mengurangi stres dan meningkatkan

kesejahteraan psikologis, sementara kegiatan ekstrakurikuler berkontribusi pada pengembangan keterampilan non-akademik seperti manajemen waktu dan kepemimpinan. Kombinasi kelima variabel ini diharapkan mampu memberikan gambaran yang komprehensif mengenai faktor-faktor gaya hidup yang berpengaruh terhadap IPK.

Dalam Islam, pendidikan merupakan salah satu jalan utama untuk mencapai kehidupan yang lebih baik. Allah SWT berfirman dalam Alqur'an Surah Al-Mujadilah ayat 11:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِذَا قِيلَ لَكُمْ تَفَسَّحُوا فِي الْمَجَالِسِ فَافْسَحُوا يَفْسَحِ اللَّهُ لَكُمْ وَإِذَا قِيلَ انشُرُوا فَانشُرُوا
يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ

“Wahai orang-orang yang beriman, apabila dikatakan kepadamu “Berilah kelapangan di dalam majelis-majelis,” lapangkanlah, niscaya Allah akan memberi kelapangan untukmu. Apabila dikatakan, “Berdirilah,” (kamu) berdirilah. Allah niscaya akan mengangkat orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat. Allah Mahateliti terhadap apa yang kamu kerjakan.” (QS. Al-Mujadilah: 11).

Ayat ini menegaskan pentingnya ilmu pengetahuan sebagai dasar untuk meningkatkan kualitas hidup manusia. Dalam konteks pendidikan tinggi, pencapaian akademik seperti Indeks Prestasi Kumulatif (IPK) menjadi salah satu ukuran yang sering digunakan untuk menilai keberhasilan mahasiswa. Namun, pencapaian ini tidak hanya dipengaruhi oleh usaha akademik, tetapi juga oleh keseimbangan pola hidup yang sesuai dengan ajaran Islam, seperti menjaga waktu tidur, kesehatan, dan aktivitas bermanfaat lainnya.

Teknologi *big data* dan kecerdasan buatan kini membuka peluang baru untuk memahami pola-pola kompleks dalam data, termasuk data gaya hidup mahasiswa

(Ouatic et al., 2022). Salah satu pendekatan yang semakin populer adalah data mining, yang memungkinkan analisis hubungan antara berbagai variabel secara lebih mendalam. Dalam data mining, algoritma *Random Forest* dikenal sebagai metode yang efektif untuk menangani data multivariabel dan hubungan yang kompleks (Salman et al., 2024). Algoritma ini bekerja dengan membangun sejumlah besar *decision tree*, yang kemudian digabungkan untuk menghasilkan prediksi yang akurat. Dalam konteks gaya hidup mahasiswa, *Random Forest* dapat membantu mengeksplorasi bagaimana berbagai faktor, seperti jam belajar, durasi tidur, aktivitas fisik, dan kegiatan sosial, saling memengaruhi pencapaian IPK (Akbar et al., 2023).

Meskipun algoritma *Random Forest* telah banyak digunakan dalam berbagai bidang, penerapannya untuk memprediksi IPK berdasarkan pola gaya hidup mahasiswa masih jarang dieksplorasi (Falát & Piscová, 2022). Penelitian sebelumnya lebih sering menggunakan metode tradisional yang kurang mampu menangkap pola kompleks dalam data multivariable (Evstatiev & Gabrovska-Evstatieva, 2021). Sebagai contoh, penelitian oleh (Brown et al., 2022) yang hanya berfokus pada pengaruh aktifitas fisik atau durasi tidur terhadap IPK, tanpa mempertimbangkan bagaimana faktor-faktor tersebut saling berinteraksi. Padahal, pemahaman ini dapat memberikan wawasan yang lebih mendalam mengenai strategi gaya hidup yang optimal bagi mahasiswa.

Penelitian ini bertujuan untuk mengeksplorasi potensi algoritma *Random Forest* dalam menganalisis dan memprediksi IPK mahasiswa berdasarkan pola gaya hidup harian mereka. Dengan mengintegrasikan berbagai variabel gaya hidup

dalam satu model, diharapkan penelitian ini dapat memberikan gambaran yang lebih komprehensif tentang faktor-faktor yang paling berpengaruh terhadap IPK. Hasil dari penelitian ini diharapkan tidak hanya memberikan kontribusi pada pengembangan ilmu pengetahuan, tetapi juga dapat digunakan sebagai referensi praktis oleh mahasiswa, pendidik, dan institusi pendidikan dalam merancang strategi yang mendukung prestasi akademik.

Dengan memahami hubungan antara gaya hidup dan pencapaian akademik secara lebih mendalam, penelitian ini juga bertujuan untuk memberikan rekomendasi yang dapat membantu mahasiswa dalam mengelola waktu dan aktivitas mereka secara lebih efektif. Pendekatan berbasis data mining ini diharapkan mampu memberikan wawasan baru yang tidak hanya relevan dalam konteks akademik, tetapi juga dalam membantu mahasiswa mencapai keseimbangan antara kehidupan pribadi dan akademik yang lebih baik. Selain itu, hasil penelitian ini diharapkan dapat menjadi acuan bagi institusi pendidikan dalam mendukung pengembangan mahasiswa secara menyeluruh.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah bagaimana mengidentifikasi faktor risiko akademik mahasiswa berdasarkan pola gaya hidup harian?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini adalah “*Student Lifestyle Dataset*” (Sumit Kumar, 2024), yang diperoleh dari *Kaggle* dan telah dimodifikasi untuk keperluan analisis dalam penelitian ini.
2. Variabel yang dianalisis dalam penelitian ini terbatas pada pola gaya hidup mahasiswa, yaitu durasi belajar, durasi tidur, aktivitas fisik, interaksi sosial, dan keterlibatan dalam kegiatan ekstrakurikuler.
3. Model prediksi hanya menggunakan algoritma *Random Forest*, tanpa membandingkan hasilnya dengan metode *machine learning* lainnya.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian ini adalah untuk memprediksi Indeks Prestasi Kumulatif (IPK) mahasiswa berdasarkan pola gaya hidup harian.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Menambah wawasan dalam bidang machine learning dan data mining, khususnya penerapan algoritma *Random Forest* dalam analisis data pendidikan.
2. Membantu mahasiswa memahami bagaimana gaya hidup mereka berdampak pada IPK, sehingga mereka dapat mengelola waktu dan kebiasaan dengan lebih baik.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terdahulu

Pada penelitian yang dilakukan oleh Zhao et al., (2021) meneliti analisis perilaku mahasiswa menggunakan data multisumber seperti *Small Private Online Course* (SPOC), kartu pintar, dan WiFi. Dengan algoritma *Random Forest* dan SVM, mereka mengembangkan model *AugmentED* yang mampu memprediksi performa akademik dengan akurasi tinggi. Model ini juga memberikan umpan balik visual yang membantu mahasiswa meningkatkan hasil akademik. Studi ini menunjukkan efektivitas algoritma *machine learning* dalam menganalisis data perilaku dan pengaruhnya terhadap pencapaian akademik.

Studi yang dilakukan oleh Behr et al., (2020) menerapkan *Random Forest* berbasis *Conditional Inference Trees* untuk memprediksi kemungkinan mahasiswa mengalami dropout dari universitas. Studi ini menganalisis data dari tiga fase pendidikan dan menunjukkan bahwa prediksi dropout paling akurat dapat dicapai menggunakan data dari fase awal pendidikan. Dengan *Area Under Curve* (AUC) sebesar 0.86, faktor utama yang berkontribusi adalah nilai akhir sekolah, tingkat kepuasan mahasiswa terhadap program studi, dan konsep diri akademik.

Penelitian oleh Asselman et al., (2023) membandingkan algoritma *XGBoost* dengan *Performance Factors Analysis* (PFA) dalam prediksi performa akademik siswa. Hasil penelitian menunjukkan bahwa *XGBoost* mencapai akurasi tertinggi dan meningkatkan nilai AUC hingga 5,2% dibandingkan dengan metode PFA.

Studi ini menyoroti bagaimana teknik machine learning dapat digunakan untuk meningkatkan keakuratan model prediksi akademik berdasarkan berbagai faktor yang memengaruhi hasil belajar siswa.

Penelitian oleh Huynh-Cam et al., (2022) menggunakan metode *Educational Data Mining* (EDM) untuk menganalisis faktor-faktor yang memengaruhi kinerja akademik siswa internasional dan siswa berkebutuhan khusus. Algoritma yang digunakan mencakup *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP), *Decision Tree* (DT), dan *Random Forest* (RF). Hasil menunjukkan bahwa SVM unggul untuk prediksi siswa internasional dengan akurasi 100%, sedangkan RF memberikan hasil terbaik untuk siswa lokal dengan akurasi 98,6%.

Penelitian oleh Vives et al., (2024) meneliti prediksi performa akademik mahasiswa dalam mata kuliah pemrograman dasar menggunakan algoritma *machine learning*, termasuk LSTM, DNN, RF, DT, SVC, LR, dan KNN. Penelitian ini menerapkan SMOTE dan GAN untuk menyeimbangkan data. Hasilnya menunjukkan bahwa LSTM dengan GAN mencapai akurasi tertinggi sebesar 98,3% pada minggu ke-8. SMOTE lebih efektif untuk data tradisional, sementara GAN lebih sesuai untuk *deep learning* seperti LSTM.

Penelitian Huang, (2024) mengembangkan algoritma *Random Forest* multinomial berbasis *entropi Grassberger* untuk mengevaluasi kualitas pengajaran di kursus profesional. Algoritma ini menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data kecil, sehingga menghasilkan sistem evaluasi pengajaran yang efektif. Studi ini menyoroti bagaimana optimasi algoritma *machine learning* dapat digunakan untuk meningkatkan evaluasi kualitas pendidikan.

Tabel 2.1 Penelitian terdahulu

No	Sumber	Input	Metode	Output	Hasil
1	Liang Zhao et al. (2021)	Data multisumber termasuk perilaku mahasiswa (belajar online/offline)	LSTM, Random Forest, Support Vector Classifier	Prediksi kinerja akademik	Model AugmentED mencapai akurasi tinggi dalam memprediksi kinerja mahasiswa berdasarkan data multisumber
2	Andreas Behr et al. (2020)	Data karakteristik mahasiswa, nilai akhir sekolah	Random Forest, Conditional Inference Trees	Prediksi dropout mahasiswa	Akurasi AUC 0.86 dengan faktor utama nilai akhir sekolah dan kepuasan mahasiswa
3	Amal Asselman et al. (2021)	Data kinerja mahasiswa, perilaku selama belajar	XGBoost, Random Forest, AdaBoost	Prediksi kinerja akademik	XGBoost mengungguli model lainnya dengan peningkatan AUC sebesar 5,2% dibandingkan model dasar
4	Thao-Trang Huynh-Cam et al. (2022)	Profil penerimaan mahasiswa, GPA semester pertama	SVM, MLP, RF, DT	Prediksi kinerja belajar	SVM, MLP, dan DT unggul untuk mahasiswa internasional, sedangkan RF unggul untuk mahasiswa lokal
5	Luis Vives et al. (2024)	Data akademik mahasiswa, kehadiran, keterlibatan	LSTM, DNN, DT, RF, LR, SVC, KNN	Prediksi performa akademik mahasiswa pada mata kuliah dasar pemrograman	LSTM dengan GAN mencapai akurasi tertinggi (98.3%) pada minggu ke-8. GAN efektif pada deep learning seperti LSTM, sementara SMOTE lebih cocok untuk data tradisional
6	Chuanbo Huang (2024)	Indikator pengajaran, nilai mahasiswa, data evaluasi	Multinomial Random Forest dengan Grassberger Entropy	Prediksi kualitas pengajaran mahasiswa	Model ini meningkatkan akurasi prediksi dengan deviasi rendah, terutama pada dataset sampel kecil

Penelitian ini berbeda dari studi sebelumnya dalam hal fokus, metode, variabel, dan tujuan. Penelitian sebelumnya seperti Behr (2020) dan Asselman (2023) hanya menekankan prediksi akademik dari data seperti nilai ujian dan

kepuasan studi, tanpa mempertimbangkan faktor gaya hidup secara menyeluruh. Padahal, faktor-faktor seperti pola tidur, aktivitas fisik, waktu belajar, hingga keterlibatan sosial dapat memengaruhi kinerja akademik baik secara langsung maupun tidak langsung. Kebaruan (novelty) dari penelitian ini terletak pada integrasi lima faktor gaya hidup mahasiswa dalam satu model prediktif berbasis *Random Forest*, sehingga memungkinkan analisis interaksi variabel secara holistik. Pendekatan ini memberikan perspektif yang lebih menyeluruh dibandingkan penelitian sebelumnya yang umumnya hanya mengevaluasi satu atau dua faktor secara terpisah.

Dari segi variabel, penelitian ini mencakup faktor akademik dan gaya hidup mahasiswa, seperti durasi belajar, pola tidur, aktivitas fisik, keterlibatan dalam kegiatan ekstrakurikuler, dan interaksi sosial, sehingga menawarkan pemahaman yang lebih komprehensif mengenai pengaruh gaya hidup terhadap pencapaian akademik. Tujuan penelitian ini adalah mengintegrasikan berbagai faktor gaya hidup dalam satu model prediktif berbasis *Random Forest*, untuk memberikan wawasan lebih dalam mengenai faktor-faktor yang berpengaruh terhadap akademik mahasiswa, serta memberikan rekomendasi yang berguna bagi mahasiswa dan institusi pendidikan dalam perencanaan akademik dan pembelajaran yang lebih efektif.

2.2 Indeks Prestasi Akademik (IPK)

Indeks Prestasi Akademik (IPK) adalah ukuran yang mencerminkan hasil evaluasi akademik mahasiswa berdasarkan nilai dari seluruh mata kuliah yang diambil selama studi (Papadogiannis et al., 2023). IPK dihitung menggunakan

sistem penilaian numerik dengan skala 0 hingga 4.0, di mana 4.0 menunjukkan prestasi akademik yang sangat baik. Setiap mata kuliah memiliki bobot tertentu berdasarkan jumlah sks, dan nilai yang diperoleh mahasiswa dalam mata kuliah tersebut dikalikan dengan sks untuk menentukan nilai mutu. Jumlah total nilai mutu kemudian dibagi dengan jumlah total sks yang diambil, menghasilkan IPK akhir. Perhitungan IPK didasarkan pada nilai ujian dan tugas yang diberikan oleh pengajar. Setiap nilai huruf, seperti A, B, C, atau D, memiliki nilai numerik tertentu yang berkontribusi terhadap IPK mahasiswa, dengan mata kuliah berbobot sks besar memberikan pengaruh lebih besar terhadap IPK keseluruhan.

IPK sering dijadikan indikator utama dalam menilai kemampuan akademik mahasiswa (Sternier, 2020). Banyak perguruan tinggi, perusahaan, dan lembaga pemerintah menggunakan IPK sebagai syarat seleksi penerimaan, beasiswa, atau rekrutmen tenaga kerja. Mahasiswa dengan IPK tinggi sering dianggap lebih unggul dan memiliki peluang lebih besar dalam mendapatkan beasiswa atau pekerjaan (Nunik, 2021). Namun, IPK bukan satu-satunya faktor penentu keberhasilan seseorang. Keterampilan praktis, pengalaman, dan kemampuan interpersonal juga sangat berperan dalam dunia kerja. Oleh karena itu, selain fokus pada pencapaian akademik, mahasiswa perlu mengembangkan kompetensi lain yang mendukung kesuksesan mereka di masa depan.

2.3 Gaya Hidup

Gaya hidup merupakan cara individu menjalani kehidupan sehari-hari yang mencakup pola kebiasaan, pilihan, dan kebijakan dalam berbagai aspek, seperti pola makan, aktivitas fisik, kebiasaan tidur, pekerjaan, serta interaksi sosial (Elian

et al., 2023). Gaya hidup mencerminkan nilai-nilai pribadi, budaya, status sosial, dan kondisi ekonomi seseorang. Faktor eksternal, seperti lingkungan, pendidikan, dan media sosial, serta faktor internal, seperti kesehatan fisik, psikologis, dan spiritual, turut memengaruhi gaya hidup seseorang (Bakar et al., 2020). Seiring waktu, gaya hidup dapat berubah berdasarkan pengalaman hidup, perkembangan pribadi, serta perubahan dalam lingkungan sosial dan ekonomi.

Dalam konteks kesehatan, gaya hidup yang sehat melibatkan pola makan bergizi, olahraga teratur, cukup tidur, dan manajemen stres yang baik (Brivio et al., 2023). Konsumsi makanan sehat, aktivitas fisik rutin, serta tidur yang cukup membantu menjaga kesehatan fisik dan mencegah penyakit kronis. Selain aspek kesehatan fisik, gaya hidup juga mencakup dimensi sosial dan psikologis, seperti interaksi dengan orang lain, keterlibatan dalam komunitas, serta pengelolaan emosi dan stress (Buijs et al., 2023). Gaya hidup yang seimbang dapat meningkatkan kebahagiaan dan kesejahteraan seseorang, sementara gaya hidup yang tidak seimbang dapat memicu stres, kecemasan, serta masalah kesehatan mental (Sirgy, 2024). Oleh karena itu, penting bagi setiap individu untuk menciptakan pola hidup yang sehat dan harmonis guna mencapai kualitas hidup yang lebih baik.

2.4 Normalisasi *Z-score Scaling*

Normalisasi merupakan salah satu tahap penting dalam proses prapengolahan data (data preprocessing), terutama ketika data yang digunakan memiliki skala yang berbeda antar fitur (Izonin et al., 2022). Salah satu metode normalisasi yang umum digunakan adalah Standard Scaler atau dikenal juga sebagai *Z-score normalization*. Teknik ini bekerja dengan mengubah skala data sehingga setiap fitur memiliki nilai

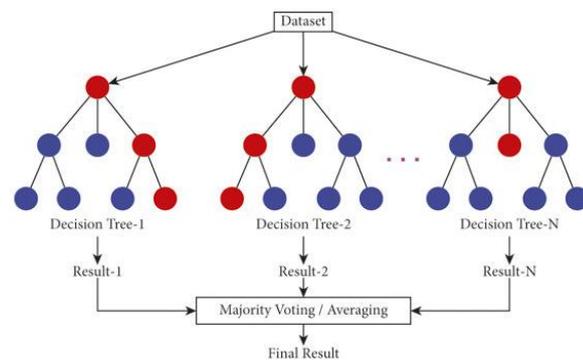
rata-rata nol dan standar deviasi satu. Dengan demikian, skala data menjadi seragam dan lebih sesuai untuk algoritma pembelajaran mesin yang sensitif terhadap perbedaan skala, seperti Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan regresi logistic.

Menurut (Raju et al., 2020), penggunaan Standard Scaler sangat tepat diterapkan pada data yang mendekati distribusi normal karena metode ini mempertahankan bentuk distribusi dan tidak membatasi nilai dalam rentang tertentu, seperti halnya Min-Max Scaling. Normalisasi ini membantu mempercepat proses pelatihan model dan meningkatkan akurasi prediksi karena setiap fitur diperlakukan secara proporsional. Dengan demikian, pemilihan metode normalisasi yang sesuai dengan karakteristik data merupakan langkah krusial dalam membangun model pembelajaran mesin yang optimal dan andal.

2.5 *Random Forest*

Algoritma *Random Forest* adalah metode pembelajaran mesin berbasis *ensemble* yang digunakan untuk tugas klasifikasi dan regresi (Savargiv et al., 2021). Algoritma ini pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 sebagai pengembangan dari *decision tree* yang mengatasi kelemahan seperti *overfitting* dan sensitivitas terhadap data outlier (Schonlau & Zou, 2020). Algoritma ini terdiri dari sekumpulan pohon keputusan (*decision trees*) yang dilatih secara independen menggunakan subset data yang berbeda dengan teknik *bootstrap sampling*. Hasil akhir dari *Random Forest* diperoleh dengan cara mengambil rata-rata prediksi (untuk regresi) atau menggunakan mayoritas suara (voting) dari setiap pohon keputusan (untuk klasifikasi). Keunggulan utama dari *Random Forest* adalah

kemampuannya dalam mengurangi *overfitting* yang sering terjadi pada pohon keputusan tunggal serta meningkatkan akurasi dan kestabilan model dengan mengombinasikan hasil dari banyak pohon (Salman et al., 2024).



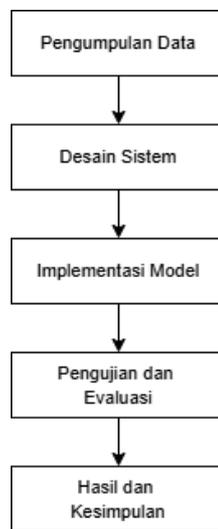
Gambar 2.1 Random Forest

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Alur Penelitian

Tahap desain penelitian merupakan tahap yang digunakan untuk merencanakan alur penelitian yang terstruktur. Alur penelitian tersebut dapat dilihat pada gambar 3.1 dibawah ini.



Gambar 3.1 Alur penelitian

3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari *Student Lifestyle Dataset* yang tersedia di Kaggle dengan tautan sebagai berikut: <https://www.kaggle.com/datasets/steve1215rogg/student-lifestyle-dataset>. Dataset ini berisi 2000 entri dengan 8 fitur, termasuk kebiasaan belajar, pola tidur, aktivitas fisik, dan keterlibatan dalam kegiatan sosial. Dataset ini telah dimodifikasi untuk tujuan penelitian, dengan menambahkan beberapa fitur yang disesuaikan untuk

meningkatkan relevansi dan kualitas data. Rincian lebih lanjut mengenai dataset dapat dilihat pada Tabel 3.1

Tabel 3.1 Atribut Dataset

No	Nama Variabel	Deskripsi	Keterangan
1	<i>Student_ID</i>	Identifikasi unik untuk setiap mahasiswa	Nomor ID unik
2	<i>Study_Hours_Per_Day</i>	Jumlah waktu belajar per hari	Dalam jam
3	<i>Krt_study</i>	Kriteria belajar	Terbalik (1), Acak (2), Teratur (3)
4	<i>Extracurricular_Hours_Per_Day</i>	Waktu yang dihabiskan untuk kegiatan ekstrakurikuler per hari	Dalam jam
5	<i>Sleep_Hours_Per_Day</i>	Durasi tidur mahasiswa setiap hari	Dalam jam
6	<i>Krt_sleep</i>	Kriteria tidur	
7	<i>Social_Hours_Per_Day</i>	Waktu yang dihabiskan untuk interaksi sosial per hari	Dalam jam
8	<i>Physical_Activity_Hours_Per_Day</i>	Waktu yang dihabiskan untuk aktivitas fisik per hari	Dalam jam
9	<i>GPA</i>	Indeks Prestasi Kumulatif mahasiswa	Skala 4.00
10	<i>Stress_Level</i>	Tingkat stres mahasiswa	Dalam skala Rendah, Menengah, Tinggi

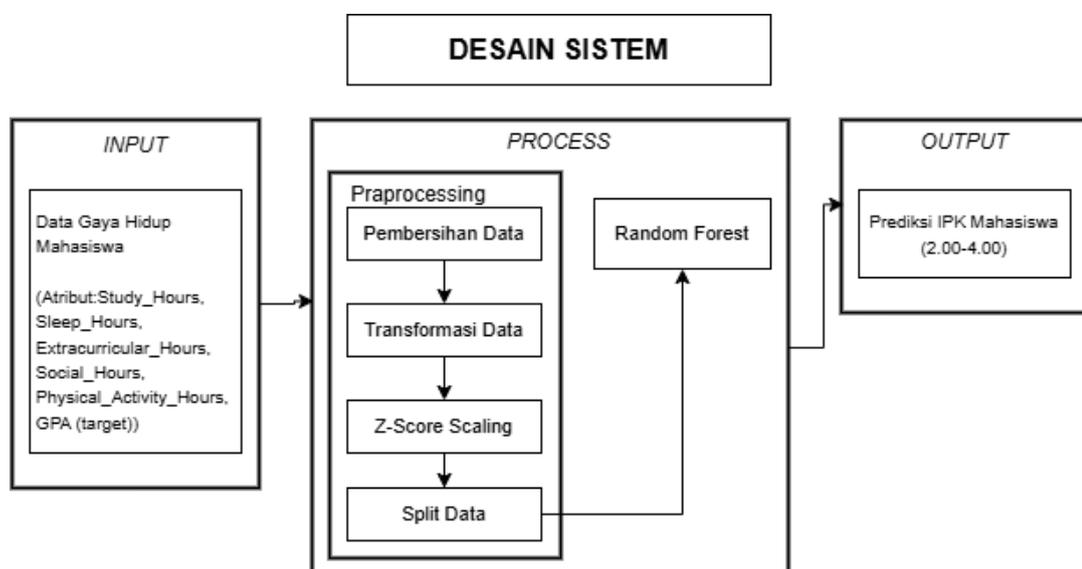
Berdasarkan hasil pengumpulan data yang telah dijelaskan pada Tabel 3.1, berikut adalah distribusi karakteristik variabel tersebut:

1. *Study Hours*: Mayoritas responden (44%) belajar antara 6 hingga 8 jam per hari, diikuti oleh 38% yang belajar lebih dari 8 jam, dan 18% yang belajar kurang dari 6 jam.
2. *Extracurricular Hours*: Sebagian besar responden (53%) menghabiskan 1-3 jam untuk kegiatan ekstrakurikuler, sedangkan 24% berpartisipasi kurang dari 1 jam, dan 23% lebih dari 3 jam.
3. *Sleep Hours*: Sebanyak 39% responden tidur lebih dari 8 jam, 41% tidur antara 6-8 jam, dan 20% tidur kurang dari 6 jam.

4. *Social Hours*: 39% responden bersosialisasi kurang dari 2 jam, 35% antara 2-4 jam, dan 26% lebih dari 4 jam.
5. *Physical Activity*: 55% responden melakukan aktivitas fisik 2-6 jam, 26% lebih dari 6 jam, dan 20% kurang dari 2 jam.
6. *GPA*: Mayoritas responden 53% memiliki GPA antara 3.0-3.5, diikuti oleh 35% dengan GPA 2.5-3.0, 10% GPA diatas 3.5, dan 2% dengan GPA < 2.5.
7. *Stress Level*: Sebagian besar responden 51% berada dalam kategori stres tinggi, diikuti oleh 34% dengan stres moderat, dan 15% dengan stres rendah.

3.3 Desain Sistem

Desain sistem dalam penelitian ini mencakup beberapa tahapan yang telah digambarkan pada Gambar 3.2 dibawah ini.



Gambar 3.2 Desain sistem

Penelitian ini menggunakan bahasa pemrograman Python dengan dataset “*Student Lifestyle*” yang diperoleh dari *Kaggle*, yang memuat informasi gaya hidup harian mahasiswa. Tahapan *preprocessing* diawali dengan pembersihan data untuk menangani nilai hilang serta menghapus data duplikat. Selanjutnya, dilakukan transformasi data melalui teknik *Z-Score Scaling* guna memastikan setiap fitur berada pada skala yang seragam. Data kemudian dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian. Setelah tahap *preprocessing* selesai, algoritma *Random Forest Regression* diterapkan untuk memprediksi nilai IPK berdasarkan variabel gaya hidup. Hasil evaluasi model dilakukan menggunakan metrik R^2 , MAE, dan RMSE, yang selanjutnya dianalisis untuk menilai sejauh mana model mampu memprediksi IPK dengan akurat.

3.4 *Preprocessing*

Preprocessing merupakan langkah awal yang penting dalam pengolahan data sebelum model machine learning diterapkan. Tahapan ini bertujuan untuk membersihkan dan mempersiapkan data sebelum digunakan dalam pemodelan (Demir, 2023). Pada penelitian ini, proses *preprocessing* mencakup pembersihan data, transformasi data, seleksi fitur.

3.4.1 Pembersihan Data

Pembersihan data merupakan langkah krusial dalam tahap *preprocessing* untuk memastikan kualitas dataset yang digunakan. Proses ini meliputi penanganan nilai yang hilang (*missing values*) dengan menerapkan metode yang sesuai, seperti imputasi rata-rata (*mean imputation*) atau interpolasi. Selain itu, data duplikat akan

dihapus untuk menghindari potensi bias, dan data yang tidak valid akan diperiksa serta diperbaiki agar tidak memengaruhi integritas hasil analisis.

3.4.2 Transformasi Data

Transformasi data adalah proses mengubah data menjadi format yang lebih sesuai untuk analisis. Pada penelitian ini, transformasi dilakukan dengan Teknik Standarisasi (*Z-Score Scaling*) untuk memastikan bahwa setiap fitur dalam dataset memiliki skala yang seragam. Standarisasi adalah metode transformasi data yang mengubah nilai data sehingga memiliki rata-rata 0 dan standar deviasi 1. Teknik ini biasanya digunakan ketika data mengikuti distribusi normal dan membutuhkan penyamaan skala antar fitur agar analisis lebih akurat. Berikut untuk rumus yang digunakan:

$$X' = \frac{X - \mu}{\sigma} \quad (3, 1)$$

Keterangan:

X' = nilai setelah standarisasi

X = nilai asli

μ = rata-rata

σ = standar deviasi

3.4.3 Split Data

Split data merupakan proses membagi dataset menjadi beberapa bagian, biasanya terdiri dari *training set* dan *testing set* (Botache et al., 2023). Pada penelitian ini, training set ditetapkan sebesar 80% dari keseluruhan data dan digunakan untuk membangun model, sedangkan testing set sebesar 20% digunakan untuk mengevaluasi performa model. Tujuan dari pembagian ini adalah untuk

melatih model menggunakan sebagian data (*training set*) dan menguji kinerjanya menggunakan data yang belum pernah dilihat sebelumnya (*testing set*).

3.5 Implementasi *Random Forest*

Pada penelitian ini, penerapan algoritma *Random Forest Regression* dimulai dengan beberapa tahap penting yang bertujuan untuk meningkatkan keberagaman model dan mengurangi risiko *overfitting*. *Random Forest* adalah metode ensemble yang menggunakan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat dan stabil. Tahapan-tahapan implementasi algoritma ini adalah sebagai berikut:

1. *Bootstrap Sampling*: Pada tahap ini, dataset utama dibagi menjadi beberapa subset dengan teknik resampling acak dengan pengembalian (*with replacement*) agar model memiliki variasi dan mengurangi risiko *overfitting*.

Tabel 3.2 Contoh hasil Bootstrap sampling

<i>Study_Hours_Per_Day</i>	<i>Extracurricular_Hours_Per_Day</i>	<i>Sleep_Hours_Per_Day</i>	<i>Social_Hours_Per_Day</i>	<i>Physical_Activity_Hours_Per_Day</i>	<i>GPA</i>
6.9	3.8	8.7	2.8	1.8	2.99
5.3	3.5	8	4.2	3	2.75
5.1	3.9	9.2	1.2	4.6	2.67
6.5	2.1	7.2	1.7	6.5	2.88
5.2	3.6	6.3	4	4.9	2.82
6	2.1	8	0.3	7.6	2.85
8	0.7	5.3	5.7	4.3	3.08
8.4	1.8	5.6	3	5.2	3.2
5.2	3.6	6.3	4	4.9	2.82
7.7	0.7	9.8	4.5	1.3	2.76

2. *Pemilihan Fitur Secara Acak*: Sebelum melakukan pemilihan split, algoritma memilih sejumlah fitur secara acak dari total fitur yang tersedia. Hal ini dikenal dengan *feature randomness*. Jumlah fitur yang dipilih ditentukan oleh rumus:

$$M = \sqrt{n} \quad (3, 2)$$

Keterangan:

M = jumlah maksimum fitur yang dipilih secara acak untuk menentukan split terbaik di setiap node.

n = jumlah total fitur dalam dataset.

Tabel 3.3 Contoh pemilihan fitur secara acak

<i>Study_Hours_Per_Day</i>	<i>Extracurricular_Hours_Per_Day</i>
6.9	3.8
5.3	3.5
5.1	3.9
6.5	2.1
5.2	3.6
6	2.1
8	0.7
8.4	1.8
5.2	3.6
7.7	0.7

3. Pemilihan Split Terbaik: Setelah fitur acak dipilih, di setiap node pohon, berbagai nilai threshold diuji untuk membagi data menjadi dua bagian. Split terbaik dipilih berdasarkan kriteria *Mean Squared Error* (MSE), yang dihitung dengan rumus berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2 \quad (3, 3)$$

Keterangan:

y_i adalah nilai aktual.

\hat{y} adalah nilai rata-rata dalam node.

n adalah jumlah sampel dalam node.

Misal pada perhitungan ini kita menggunakan fitur *Study_Hours_Per_Day*

dengan nilai threshold nya adalah 7.85. Berikut untuk detail perhitungannya:

Tabel 3.4 Contoh tabel split di 7.85

<i>Study Hours Per Day</i>	IPK	Rata-rata (\bar{Y})
≤ 7.85	2.67, 2.82, 2.82, 2.75, 2.85, 2.88, 2.99, 2.76	2.82
> 7.85	3.08, 3.20	3.14

MSE(≤ 7.85):

MSE_1

$$\begin{aligned}
 &= \frac{(2.67 - 2.82)^2 + (2.82 - 2.82)^2 + (2.82 - 2.82)^2 + (2.75 - 2.82)^2 +}{8} \\
 &= \frac{(2.85 - 2.82)^2 + (2.88 - 2.82)^2 + (2.99 - 2.82)^2 + (2.76 - 2.82)^2}{8} \\
 &= \frac{0.0225 + 0 + 0 + 0.0049 + 0.0009 + 0.0036 + 0.0289 + 0.0036}{8} \\
 &= 0.00804375
 \end{aligned}$$

MSE(> 7.85):

$$\begin{aligned}
 MSE_2 &= \frac{(3.08 - 3.14)^2 + (3.20 - 3.14)^2}{2} \\
 &= \frac{0.0036 + 0.0036}{2} = 0.0036
 \end{aligned}$$

Total MSE split 7.85:

$$MSE_{7.85} = \frac{8}{10} \times 0.00804375 + \frac{2}{10} \times 0.0036 = 0.007155$$

4. Membangun *Decision Tree*: Berdasarkan split terbaik yang telah dipilih pada setiap node, pohon keputusan mulai dibangun secara rekursif. Proses ini terus dilakukan di setiap cabang pohon (node) hingga tercapai kriteria berhenti.

5. Agregasi Prediksi: Setelah semua pohon keputusan dalam *Random Forest* dilatih, hasil prediksi dari setiap pohon digabungkan dengan mengambil rata-rata dari semua prediksi:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(X) \quad (3, 4)$$

Keterangan:

T jumlah pohon dalam *Random Forest*.
 $h_t(X)$ adalah prediksi dari pohon ke- t .

Sebagai contoh, misalnya disini kita memiliki lima tree yang memberikan prediksi 3.52, 3.20, 2.80, 2.32, dan 2.76, kemudian akan dijumlahkan semua dan diambil rata-ratanya sebagai prediksi akhir. Berikut untuk perhitungannya:

$$\frac{3.52 + 3.20 + 2.80 + 2.32 + 2.76}{5} = \frac{14.60}{5} = 2.92$$

Jadi, prediksi akhir dari semua *decision tree* tersebut adalah 2.92.

3.6 Evaluasi Model

Evaluasi model dalam pembelajaran mesin bertujuan untuk mengukur sejauh mana model mampu melakukan prediksi dengan akurat. Dalam konteks regresi, evaluasi dilakukan dengan membandingkan nilai prediksi dengan nilai aktual dari data uji. Beberapa metrik yang umum digunakan adalah:

1. *R Squared* (R^2)

R Squared (R^2) adalah presentase untuk mengukur seberapa baik model menjelaskan variabilitas data target. Nilai R^2 berada pada rentang 0 hingga 1, di mana semakin mendekati 1 menunjukkan model semakin baik.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \quad (3, 5)$$

Keterangan:

y_i = nilai sebenarnya
 \hat{y} = nilai prediksi
 \bar{y} = rata-rata nilai sebenarnya

2. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah presentase untuk mengukur rata-rata kesalahan absolut antara prediksi model dan nilai sebenarnya.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}| \quad (3, 6)$$

Keterangan:

n = jumlah sampel
 y_i = nilai sebenarnya
 \hat{y} = nilai prediksi

3. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) adalah presentase untuk menghitung akar kuadrat dari rata-rata selisih kuadrat antara nilai aktual dan prediksi. Metode ini lebih sensitif terhadap nilai error yang besar.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (3, 7)$$

3.7 Skenario Pengujian

Untuk mengevaluasi performa model *Random Forest Regression* dalam memprediksi IPK mahasiswa berdasarkan data pola gaya hidup harian, penelitian ini akan menggunakan beberapa pengujian yaitu sebagai berikut:

a) Pengujian Pengaruh *Hyperparameter*

Dilakukan pengujian terhadap beberapa *hyperparameter* untuk mengetahui kombinasi parameter terbaik yang memberikan prediksi paling akurat. Setiap parameter memiliki sejumlah nilai atau metode yang digunakan sebagai konfigurasi. Nilai-nilai tersebut diperoleh berdasarkan hasil terbaik yang telah diuji dalam penelitian sebelumnya (Baita et al., 2023). Berikut ini merupakan konfigurasi yang digunakan untuk masing-masing parameter.

Tabel 3.5 Parameter yang digunakan

Parameter	Nilai yang digunakan	Keterangan
<i>n_estimators</i>	[100, 200, 300]	Jumlah pohon keputusan
<i>max_depth</i>	[10, 20,30]	Kedalaman maksimum pohon
<i>min_samples_split</i>	[2, 5, 10]	Minimum sampel untuk membagi node
<i>min_samples_leaf</i>	[1, 2, 4]	Minimum sampel pada daun
<i>max_features</i>	[sqrt, log2]	Fitur yang dipilih secara acak saat split
<i>Bootstrap</i>	[True, False]	Sampel pelatihan untuk setiap pohon

Proses ini akan dilakukan menggunakan *GridSearchCV* untuk mencari kombinasi optimal. Kemudian akan dievaluasi berdasarkan tiga metrik utama yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, *R-squared (R²)*.

b) Pengujian Berdasarkan Kombinasi Fitur Gaya Hidup

Untuk mengetahui kontribusi masing-masing variabel gaya hidup terhadap prediksi IPK, dilakukan pengujian dengan beberapa kombinasi fitur input sebagai berikut:

Tabel 3.6 Fitur yang digunakan untuk pengujian

Model	Kombinasi Fitur
A	<i>Study_Hours_Per_Day</i> (Teratur) dan <i>Sleep_Hours_Per_Day</i> (Teratur)
B	<i>Study_Hours_Per_Day</i> (Acak) dan <i>Extracurricular_Hours_Per_Day</i>
C	<i>Social_Hours_per_Day</i> dan <i>Physical_Activity_Hours_per_Day</i>
D	<i>Sleep_Hours_Per_Day</i> (Terbalik) dan <i>Social_Hours_per_Day</i>
E	<i>Sleep_Hours_Per_Day</i> (Acak) dan <i>Physical_Activity_Hours_Per_Day</i>
F	<i>Extracurricular_Hours_Per_Day</i> dan <i>Social_Hours_Per_Day</i>
G	<i>Study_Hours_Per_Day</i> (Terbalik) dan <i>Sleep_Hours_Per_Day</i> (Terbalik)

Setiap fitur diatas akan diuji dengan *Random Forest* menggunakan *5-Fold Cross Validation* dan dievaluasi menggunakan metrik MAE yang bertujuan untuk mengidentifikasi kombinasi fitur terbaik terhadap prediksi IPK mahasiswa.

c) Pengujian Fitur Penting (*Feature Importance*)

Pengujian fitur paling penting atau *Feature Importance* dilakukan untuk mengetahui fitur gaya hidup mana yang paling berpengaruh terhadap prediksi nilai IPK. Dalam penelitian ini, metode *Random Forest* digunakan untuk menghitung *feature importance* berdasarkan kontribusi masing-masing fitur dalam mengurangi nilai kesalahan (*impurity*) saat proses pembentukan pohon keputusan.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah *Student Lifestyle Dataset* yang diperoleh dari platform *Kaggle* dan telah dimodifikasi untuk meningkatkan relevansi dan kualitas data. Dataset ini memuat 2000 entri yang merepresentasikan pola gaya hidup harian mahasiswa, mencakup durasi belajar, tidur, aktivitas sosial, kegiatan ekstrakurikuler, dan aktivitas fisik. Dataset ini dipilih karena memuat variabel-variabel gaya hidup yang dianggap berpengaruh terhadap pencapaian akademik mahasiswa, khususnya Indeks Prestasi Kumulatif (IPK). Berikut tampilan dataset yang digunakan.

Tabel 4.1 Dataset

<i>Student_ID</i>	<i>Study_Hours_Per_Day</i>	<i>Krt_Study</i>	<i>Extracurricular_Hours_Per_Day</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	<i>Social_Hours_Per_Day</i>	<i>Physical_Activity_Hours_Per_Day</i>	<i>GPA</i>
1	6.9	2	3.8	8.7	1	2.8	1.8	2.99
2	5.3	2	3.5	8	1	4.2	3	2.75
3	5.1	2	3.9	9.2	2	1.2	4.6	2.67
4	6.5	1	2.1	7.2	2	1.7	6.5	2.88
5	8.1	2	0.6	6.5	2	2.2	6.6	3.51
6	6	1	2.1	8	2	0.3	7.6	2.85
7	8	1	0.7	5.3	1	5.7	4.3	3.08
8	8.4	2	1.8	5.6	3	3	5.2	3.2
9	5.2	2	3.6	6.3	2	4	4.9	2.82
10	7.7	3	0.7	9.8	3	4.5	1.3	2.76

4.2 Data Preprocessing

Pada tahap ini, data diproses untuk memudahkan analisis lebih lanjut, khususnya dalam proses encoding dan pemodelan. Proses *preprocessing* mencakup pembersihan data untuk mengatasi nilai hilang dan duplikat, serta normalisasi

menggunakan *Z-Score Scaling* guna menyeragamkan skala antar fitur. Hasil dari setiap tahap preprocessing ditampilkan pada bagian berikut.

4.2.1 Data Cleaning

Proses *preprocessing* ini dimulai dengan pembersihan data, yaitu menghapus nilai duplikat dan menangani *missing values* menggunakan metode interpolasi rata-rata agar data menjadi lebih konsisten dan bebas dari ketidakteraturan yang dapat memengaruhi hasil prediksi. Berikut hasil setelah data cleaning.

Tabel 4.2 Hasil data cleaning

No.	Column	Missing Value	Dtype
1	<i>Student_ID</i>	2000 non-null	int64
2	<i>Study_Hours_Per_Day</i>	2000 non-null	float64
3	<i>Krt_Study</i>	2000 non-null	int64
4	<i>Extracurricular_Hours_Per_Day</i>	2000 non-null	float64
5	<i>Sleep_Hours_Per_Day</i>	2000 non-null	float64
6	<i>Krt_Sleep</i>	2000 non-null	int64
7	<i>Social_Hours_Per_Day</i>	2000 non-null	float64
8	<i>Physical_Activity_Hours_Per_Day</i>	2000 non-null	float64
9	<i>GPA</i>	2000 non-null	float64

Berdasarkan hasil data cleaning diatas, dapat diketahui bahwa tidak terdapat missing value pada seluruh kolom. Setiap atribut memiliki 2000 entri non-null, yang menunjukkan bahwa tidak ada data yang hilang atau kosong dalam keseluruhan dataset. Hal ini mengindikasikan bahwa data tidak memerlukan proses imputasi atau pengisian nilai hilang, sehingga dapat langsung digunakan dalam tahap preprocessing berikutnya.

4.2.2 Data Normalisasi (*Z-Score Scaling*)

Langkah selanjutnya adalah *transformasi* data, di mana seluruh fitur numerik dalam dataset dilakukan proses normalisasi menggunakan metode *Z-Score Scaling*.

Normalisasi ini bertujuan untuk menyamakan skala antar fitur agar tidak ada atribut yang mendominasi model akibat perbedaan rentang nilai. Dengan cara ini, setiap fitur input diubah sehingga memiliki nilai rata-rata 0 dan standar deviasi 1. Berikut adalah hasil data normalisasi.

Tabel 4.3 Hasil normalisasi *Z-score scaling*

<i>Student_ID</i>	<i>Study_Hours_Per_Day</i>	<i>Extracurricular_Hours_Per_Day</i>	<i>Sleep_Hours_Per_Day</i>	<i>Social_Hours_Per_Day</i>	<i>Physical_Activity_Hours_Per_Day</i>	<i>GPA</i>
1	-0.404487	1.566246	0.820734	0.056543	-1.005896	2.99
2	-1.528451	1.306633	0.341473	0.885882	-0.528470	2.75
3	-1.668947	1.652784	1.163063	-0.891273	0.108097	2.67
4	-0.685478	0.095105	-0.206253	-0.595080	0.864021	2.88
5	0.438487	-1.202961	-0.685514	-0.298888	0.903806	3.51
6	-1.036717	0.095105	0.341473	-1.424419	1.301661	2.85
7	0.368239	-1.116423	-1.507103	1.774459	-0.011259	3.08
8	0.649230	-0.164508	-1.301706	0.175020	0.346810	3.2
9	-1.598699	1.393171	-0.822445	0.767405	0.227453	2.82
10	0.157496	-1.116423	1.573858	1.063597	-1.204823	2.76

Hasil normalisasi menunjukkan bahwa setiap fitur gaya hidup mahasiswa telah berada pada skala yang terstandarisasi di sekitar nol. Nilai negatif mengindikasikan aktivitas di bawah rata-rata, sedangkan nilai positif menunjukkan aktivitas di atas rata-rata. Hal ini mencerminkan tingkat penyimpangan masing-masing atribut terhadap rata-rata populasi. Sementara itu, kolom GPA tetap menggunakan nilai asli karena berfungsi sebagai variabel target dalam pemodelan.

4.2.3 Split Data

Proses split data dilakukan dengan membagi dataset menjadi dua bagian, yaitu 80% data latih (*training data*) dan 20% data uji (*testing data*). Dari total 2000 data mahasiswa, sebanyak 1600 data digunakan untuk melatih model, sementara 400 data sisanya digunakan untuk menguji performa model yang telah dibangun.

Pembagian ini dilakukan secara acak untuk menjaga distribusi data tetap representatif. Hasil pembagian ini memastikan bahwa model dapat belajar dari sebagian besar data dan dievaluasi secara objektif menggunakan data yang belum pernah dilihat sebelumnya.

4.3 Metode *Random Forest*

Setelah melalui tahapan preprocessing dan pembagian data, algoritma *Random Forest Regression* diterapkan untuk memprediksi Indeks Prestasi Kumulatif (IPK) mahasiswa berdasarkan lima variabel gaya hidup utama, yaitu *Study_Hours_Per_Day*, *Sleep_Hours_Per_Day*, *Extracurricular_Hours_Per_Day*, *Social_Hours_Per_Day*, dan *Physical_Activity_Hours_Per_Day*. Berikut adalah hasil setiap Langkah-langkah algoritma *Random Forest Regression*.

4.3.1 *Bootstrap Sampling*

Pada tahap ini, dilakukan proses *bootstrap sampling* sebagai bagian dari pembentukan model *Random Forest*. Teknik ini mengambil sampel secara acak dengan pengembalian (*with replacement*) dari data latih, sehingga setiap pohon dalam *Random Forest* dilatih dengan subset data yang berbeda. Berikut adalah hasil dari salah satu *bootstrap sampling*.

Tabel 4.4 Hasil salah satu *Bootstrap sampling*

<i>Study_Hours_Per_Day</i>	<i>Extracurricular_Hours_Per_Day</i>	<i>Sleep_Hours_Per_Day</i>	<i>Social_Hours_Per_Day</i>	<i>Physical_Activity_Hours_Per_Day</i>	<i>GPA</i>
8.1	0.4	7.0	4.3	4.2	3.24
6.3	0.1	7.7	5.1	4.8	3.22
9.5	2.5	5.6	5.6	0.8	3.13
8.9	3.0	9.0	1.6	1.5	3.42
9.2	1.8	8.2	1.1	3.7	3.60
7.2	2.3	7.9	2.0	4.6	3.06
7.5	1.2	7.1	3.9	4.3	2.96
6.0	0.8	5.2	1.9	10.1	3.01
9.1	1.1	9.4	0.5	3.9	3.58
6.0	1.7	6.1	0.0	10.2	2.87

4.3.2 Feature Randomness

Pada tahap ini, fitur yang terpilih untuk digunakan sebagai input dalam proses pembentukan *decision tree* adalah *Sleep_Hours_Per_Day* dan *Study_Hours_Per_Day*. Pemilihan dilakukan secara acak sesuai dengan mekanisme algoritma *Random Forest*, yang bertujuan untuk meningkatkan keberagaman antar pohon dan mengurangi risiko *overfitting*. Kedua fitur ini dianggap mewakili aspek penting dari gaya hidup mahasiswa yang berpotensi memengaruhi nilai IPK. Fitur-fitur terpilih kemudian digunakan untuk menentukan split terbaik dalam pembentukan struktur pohon keputusan.

4.3.3 Split Terbaik

Pada tahap ini, proses pemilihan split terbaik dilakukan dengan menggunakan kriteria *Mean Squared Error* (MSE). Berikut adalah hasil pemilihan split terbaik.

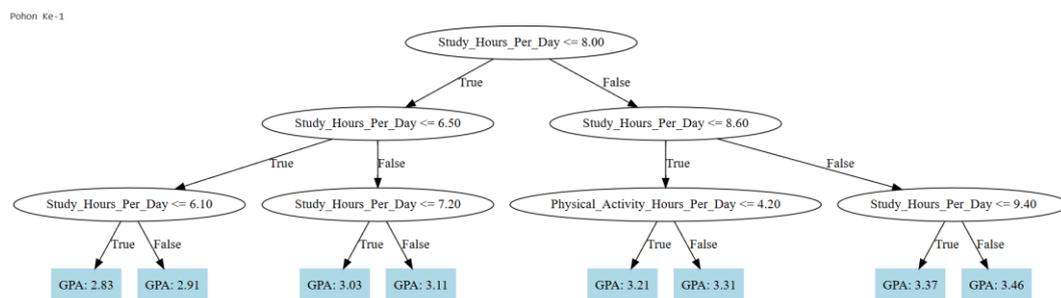
Tabel 4.5 Hasil salah satu split terbaik

Hasil Split Terbaik	
Fitur	<i>Study_Hours_Per_Day</i>
Threshold	8.0
MSE	0.0521

Berdasarkan hasil evaluasi, fitur terbaik untuk pemisahan data adalah *Study_Hours_Per_Day* dengan threshold 8.0. Split ini menghasilkan nilai MSE terkecil sebesar 0.0521, yang menunjukkan bahwa variasi nilai IPK dalam masing-masing kelompok setelah pemisahan sangat rendah. Oleh karena itu, fitur ini digunakan sebagai dasar pembentukan struktur awal pada pohon keputusan.

4.3.4 Membangun *Decision Tree*

Berdasarkan hasil dari perhitungan split terbaik diatas, berikut adalah bentuk pohon keputusan dari pemisahan node split terbaik tersebut.



Gambar 4.1 Hasil salah satu *decision tree*

Gambar 4.1 menunjukkan struktur salah satu pohon keputusan (*decision tree*) yang dihasilkan oleh algoritma *Random Forest*. Fitur utama yang digunakan untuk pemisahan adalah *Study_Hours_Per_Day*, dengan beberapa cabang yang mengarah pada nilai ambang berbeda, seperti 6.10, 6.50, 7.20, 8.00, 8.60, dan 9.40. Ini menunjukkan bahwa durasi belajar per hari merupakan faktor dominan dalam memprediksi IPK mahasiswa.

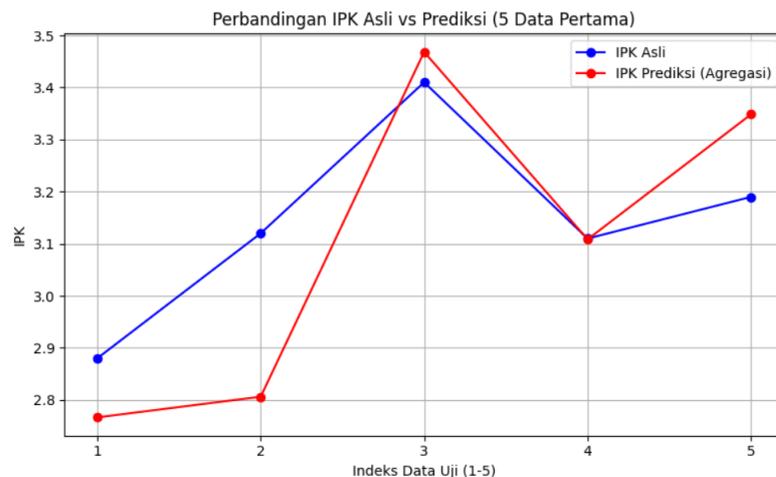
4.3.5 Agregasi Prediksi

Setelah seluruh *decision tree* dalam *Random Forest* selesai dibentuk, tahap akhir adalah melakukan agregasi prediksi. Berikut adalah hasil agregasi prediksinya.

Tabel 4.6 Hasil agregasi prediksi

<i>Student_ID</i>	<i>Study_Hours_Per_Day</i>	<i>Extracurricular_Hours_Per_Day</i>	...	<i>GPA</i>	<i>DT_1</i>	<i>DT_2</i>	<i>DT_3</i>	<i>Prediksi Akhir</i>
1861	6.20	4.00	...	2.88	2.98	2.85	2.96	2.93
354	5.90	2.20	...	3.12	2.82	2.75	2.75	2.77
1334	9.60	2.30	...	3.41	3.45	3.46	3.43	3.44
906	7.20	0.50	...	3.11	3.10	2.97	3.05	3.04
1290	9.00	3.60	...	3.19	3.34	3.35	3.33	3.34

Tabel di atas menunjukkan hasil prediksi IPK mahasiswa yang diperoleh dari beberapa *decision tree* dalam model *Random Forest*. Kolom *DT_1*, *DT_2*, dan *DT_3* menunjukkan hasil prediksi IPK oleh masing-masing pohon, sedangkan kolom *Prediksi Akhir* merupakan hasil agregasi dalam bentuk rata-rata dari ketiga nilai tersebut. Berikut merupakan tampilan visual perbandingan nilai asli dan hasil agregasi prediksi dari ke-5 data diatas.



Gambar 4.2 Hasil visual agregasi prediksi

Gambar 4.2 menunjukkan perbandingan antara nilai IPK asli dan hasil prediksi agregasi dari model *Random Forest* pada 5 data uji pertama. Garis berwarna biru merepresentasikan IPK aktual (*ground truth*), sedangkan garis merah menunjukkan hasil prediksi dari model. Dari grafik terlihat bahwa pola prediksi model cukup mengikuti tren IPK asli, meskipun terdapat sedikit deviasi pada beberapa titik, seperti data ke-3 dan ke-5. Secara umum, model mampu menangkap pola hubungan antara fitur gaya hidup dan IPK mahasiswa, dengan hasil prediksi yang cukup mendekati nilai aktual.

4.3.6 Evaluasi Model

Berdasarkan hasil implementasi model algoritma *Random forest* dalam hal prediksi diatas, yang mana kemudian dilakukan evaluasi menggunakan metrik R^2 , MAE, dan RMSE. Hasil evaluasi model tersebut dapat dilihat sebagai berikut.

Tabel 4.7 Hasil evaluasi model

Hasil Evaluasi Model	
R^2	0.5210
MAE	0.1689
RMSE	0.0447

Berdasarkan Tabel 4.7, model *Random Forest* menghasilkan nilai R^2 sebesar 0.5210, yang menunjukkan bahwa sekitar 52% variabilitas nilai IPK dapat dijelaskan oleh model melalui variabel gaya hidup yang digunakan. Nilai MAE (*Mean Absolute Error*) sebesar 0.1689 menunjukkan bahwa rata-rata selisih antara nilai prediksi dan nilai aktual IPK adalah sekitar 0.17 poin. Sementara itu, RMSE (*Root Mean Squared Error*) sebesar 0.0447 mengindikasikan tingkat kesalahan prediksi yang relatif rendah secara keseluruhan. Hasil ini menunjukkan bahwa

model memiliki performa yang cukup baik dalam memprediksi IPK mahasiswa berdasarkan pola gaya hidup harian mereka.

4.4 Hasil Pengujian *Hyperparameter*

Berdasarkan rincian konfigurasi parameter yang terdapat pada Tabel 3.9, jumlah total kombinasi parameter yang diuji mencapai 486 kombinasi. Pengujian ini dilakukan menggunakan grid search untuk menemukan kombinasi parameter yang optimal. Proses ini melibatkan evaluasi berbagai kombinasi parameter model, dengan tujuan untuk meningkatkan akurasi prediksi IPK mahasiswa. Berikut merupakan hasil dari pencarian kombinasi parameter terbaik dengan grid search.

Tabel 4.8 Hasil kombinasi parameter terbaik

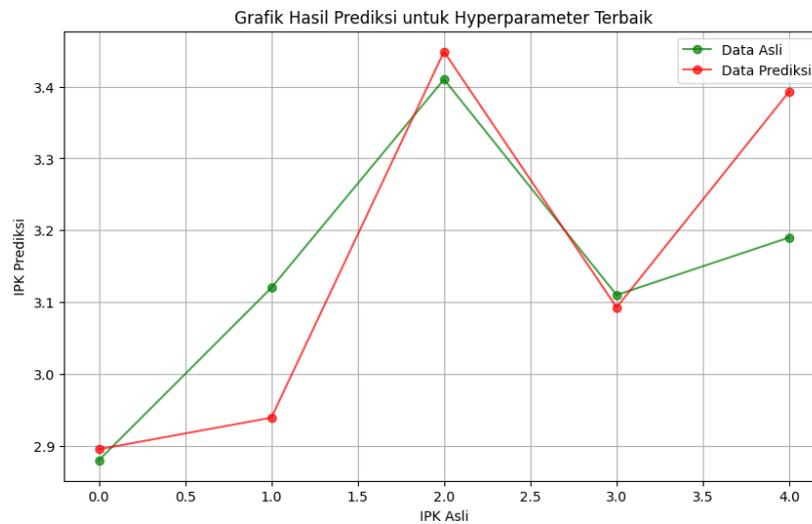
Parameter	Nilai Terbaik
<i>n_estimators</i>	300
<i>max_depth</i>	10
<i>min_samples_split</i>	10
<i>min_samples_leaf</i>	4
<i>max_features</i>	Sqrt
<i>Bootstrap</i>	True

Selanjutnya, model dilatih kembali berdasarkan kombinasi konfigurasi parameter terbaik untuk memastikan bahwa model dapat berfungsi secara optimal dengan seluruh data yang tersedia. Berikut ini adalah hasil prediksi yang diperoleh menggunakan nilai parameter terbaik tersebut.

Tabel 4.9 Hasil prediksi parameter terbaik

<i>Student_ID</i>	<i>Study_Hours</i>	<i>Krt_Study</i>	<i>Extracurricular_Hours</i>	<i>Sleep_Hours</i>	<i>Krt_Sleep</i>	<i>Social_Hours</i>	<i>Physical_Hours</i>	<i>GPA</i>	<i>Prediksi</i>
1861	6.2	1	4	6	1	3.5	4.3	2.88	2.90
354	5.9	1	2.2	9.9	3	5.6	0.4	3.12	2.93
1334	9.6	3	2.3	8.3	3	3.5	0.3	3.41	3.45
906	7.2	2	0.5	7.6	2	2.1	6.6	3.11	3.09
1290	9	1	3.6	9.9	1	0.3	1.2	3.19	3.39

Tabel 4.10 menunjukkan hasil prediksi IPK mahasiswa menggunakan kombinasi konfigurasi parameter terbaik. Pada tabel ini, setiap baris mewakili data mahasiswa dengan informasi mengenai *Student ID*, *Study Hours*, *Kriteria Study*, *Extracurricular Hours*, *Sleep Hours*, *Kriteria Sleep*, *Social Hours*, *Physical Hours*, dan IPK asli yang diikuti dengan IPK Prediksi yang dihasilkan oleh model. Misalnya, pada mahasiswa dengan Student ID 1861, yang memiliki durasi belajar 6.2 jam per hari dan GPA asli 2.88, model memprediksi nilai IPK sebesar 2.90, yang cukup mendekati nilai asli. Begitu juga dengan data lainnya, seperti Student ID 354 dengan durasi belajar 5.9 jam, yang memiliki GPA asli 3.12, dan diprediksi menjadi 2.93. Secara keseluruhan, hasil prediksi menunjukkan bahwa model *Random Forest* mampu memberikan prediksi yang cukup akurat berdasarkan konfigurasi parameter terbaik. Berikut untuk tampilan grafik IPK asli dan IPK prediksi menggunakan nilai *hyperparameter* terbaik tersebut.



Gambar 4.3 Grafik hasil prediksi parameter terbaik

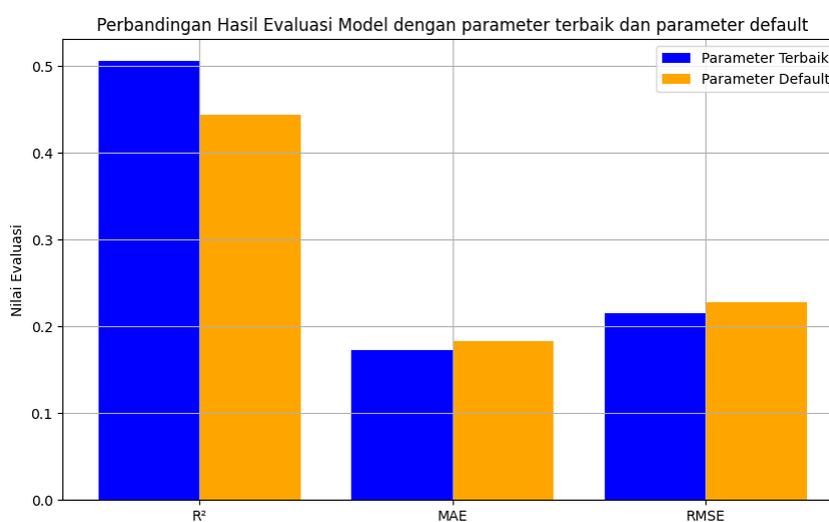
Berikut ini merupakan hasil perbandingan evaluasi model menggunakan metrik R^2 , MAE, dan RMSE antara perbandingan dari parameter default dan parameter terbaik pada *Random Forest*. Untuk hasil lebih detailnya, dapat dilihat dalam tabel 4.11 berikut.

Tabel 4.10 Hasil evaluasi parameter terbaik dan parameter default

Metrik	Parameter Terbaik	Parameter Default
R^2	0.5057	0.4435
MAE	0.1729	0.1836
RMSE	0.2148	0.2279

Tabel 4.11 menunjukkan hasil perbandingan evaluasi model menggunakan metrik R^2 , MAE, dan RMSE antara parameter terbaik dan parameter default pada model *Random Forest*. Berdasarkan hasil yang ditampilkan, nilai R^2 untuk parameter terbaik adalah 0.5057, lebih tinggi dibandingkan dengan 0.4435 pada parameter default, menunjukkan bahwa model dengan parameter terbaik dapat menjelaskan lebih banyak variabilitas data. Untuk MAE, nilai parameter terbaik adalah 0.1729, lebih rendah dibandingkan dengan 0.1836 pada parameter default,

yang menunjukkan kesalahan prediksi lebih kecil pada parameter terbaik. Begitu pula dengan RMSE, yang memiliki nilai 0.2148 pada parameter terbaik, lebih rendah dari 0.2279 pada parameter default, menunjukkan bahwa model dengan parameter terbaik lebih akurat dalam memprediksi IPK mahasiswa. Berikut grafik yang menunjukkan perbandingan evaluasi model antara parameter terbaik dan parameter default tersebut.



Gambar 4.4 Grafik perbandingan evaluasi model

4.5 Hasil Pengujian Kombinasi Fitur

Pengujian kombinasi fitur gaya hidup untuk prediksi IPK mahasiswa dilakukan menggunakan model *Random Forest* dengan *5-Fold Cross Validation*. Setiap kombinasi pada Tabel 3.10 dievaluasi berdasarkan nilai MAE untuk menilai kontribusinya terhadap akurasi prediksi. Pengujian ini menggunakan parameter terbaik hasil *gridsearch* sebelumnya. Tujuan dari pengujian ini adalah untuk mengidentifikasi kombinasi fitur paling efektif dalam meningkatkan prediksi IPK mahasiswa. Hasil evaluasi digunakan untuk menentukan kombinasi fitur yang paling menurunkan nilai MAE dan meningkatkan akurasi model.

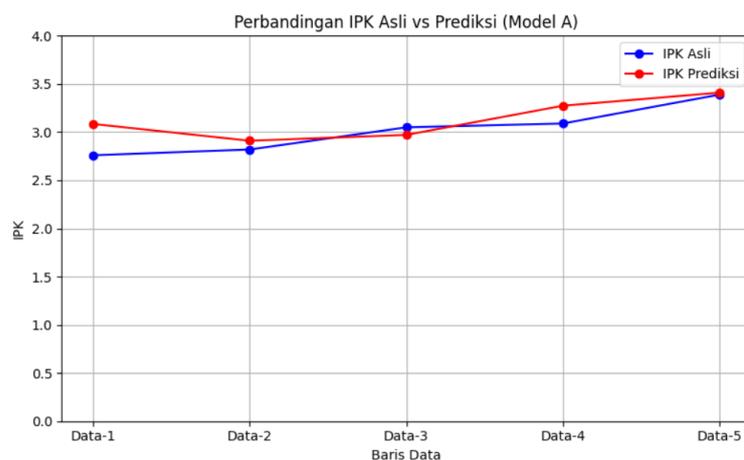
4.5.1 Hasil Kombinasi Model A

Pada model A ini, pengujian dilakukan dengan kombinasi fitur *Study_Hours_Per_Day* (Teratur) dan *Sleep_Hours_Per_Day* (Teratur). Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.11.

Tabel 4.11 Hasil prediksi model A

<i>Indeks</i>	<i>Study_Hours_Per_Day</i>	<i>Krt_Study</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	IPK Asli	IPK Prediksi
10	7.7	3	9.8	3	2.76	3.10
13	6.4	3	5.7	3	2.82	2.90
31	6.2	3	6.8	3	3.05	2.97
48	9.3	3	5.8	3	3.09	3.27
58	9	3	7.2	3	3.39	3.40

Berdasarkan Tabel 4.11, sebagian besar hasil prediksi IPK menunjukkan peningkatan. Pada data indeks 10, 13, 48, dan 58 mengalami peningkatan sekitar (+0.01-0.34 poin), sementara indeks 31 mengalami penurunan (-0.08 poin). Gambar 4.5 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.5 Hasil visualisasi model A

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.12.

Tabel 4.12 Hasil evaluasi model A

Model	Fitur	MAE
A	<i>Study_Hours_Per_Day</i> (Teratur) dan <i>Sleep_Hours_Per_Day</i> (Teratur)	0.138892

Berdasarkan hasil yang tertera pada Tabel 4.12, evaluasi menunjukkan bahwa nilai MAE (*Mean Absolute Error*) yang diperoleh adalah sebesar 0.138892. Model A menggunakan kombinasi fitur *Study_Hours_Per_Day* (Teratur) dan *Sleep_Hours_Per_Day* (Teratur), dengan nilai 3 pada kedua fitur tersebut pada data yang digunakan.

4.5.2 Hasil Kombinasi Model B

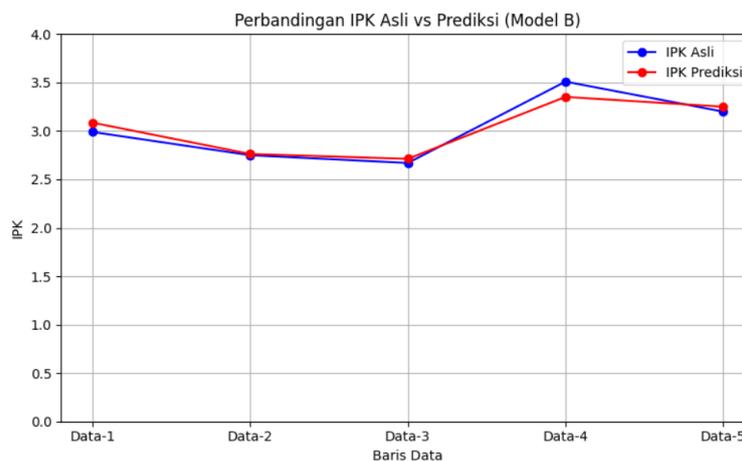
Pada model B ini, pengujian dilakukan dengan kombinasi fitur *Study_Hours_Per_Day* (Acak) dan *Extracurricular_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.13.

Tabel 4.13 Hasil Prediksi Model B

Indeks	<i>Study_Hours_Per_Day</i>	<i>Krt_Study</i>	<i>Extracurricular_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
1	6.9	2	3.8	2.99	3.01
2	5.3	2	3.5	2.75	2.77
3	5.1	2	3.9	2.67	2.79
5	8.1	2	0.6	3.51	3.25
8	8.4	2	1.8	3.20	3.22

Berdasarkan Tabel 4.13, peningkatan prediksi IPK hanya terjadi secara ringan. Hasil prediksi diatas menunjukkan indeks 1,2,3, dan 8 mengalami peningkatan antara (+0.02-0.12 poin), sedangkan indeks 5 justru mengalami penurunan (-0.26 poin). Secara umum, model ini menghasilkan kenaikan yang

kecil dan tidak stabil. Gambar 4.6 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.6 Hasil visualisasi model B

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.14.

Tabel 4.14 Hasil evaluasi model B

Model	Fitur	MAE
B	<i>Study_Hours_Per_Day</i> (Acak) dan <i>Extracurricular_Hours_Per_Day</i>	0.140464

Berdasarkan hasil yang ditampilkan pada Tabel 4.14, Hasil evaluasi menunjukkan nilai MAE (*Mean Absolute Error*) sebesar 0.140464, yang mana model B menggunakan kombinasi fitur *Study_Hours_Per_Day* (Acak) dengan nilai = 2 pada data.

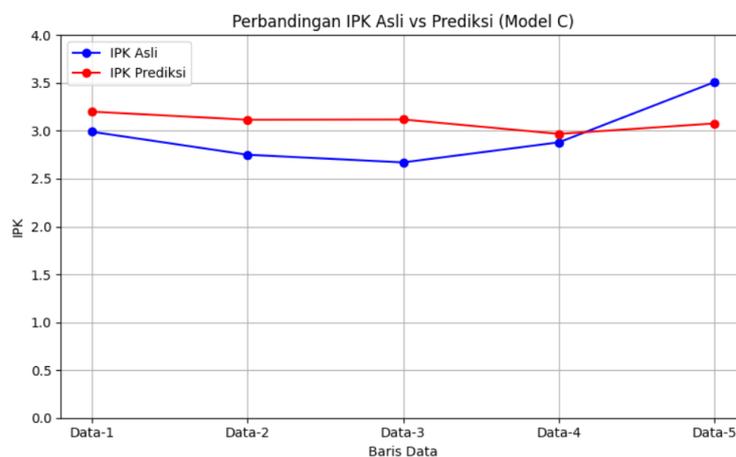
4.5.3 Hasil Kombinasi Model C

Pada model C ini, pengujian dilakukan dengan menggabungkan fitur *Social_Hours_Per_Day* dan *Physical_Activity_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.15.

Tabel 4.15 Hasil prediksi model C

<i>Indeks</i>	<i>Social_Hours_Per_Day</i>	<i>Physical_Activity_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
1	2.8	1.8	2.99	3.18
2	4.2	3	2.75	3.11
3	1.2	4.6	2.67	3.11
4	1.7	6.5	2.88	2.96
5	2.2	6.6	3.51	3.07

Tabel 4.15 menunjukkan bahwa semua data mengalami peningkatan IPK prediksi, kecuali pada indeks 5. Prediksi diatas menunjukkan indeks 1 sampai 4 mengalami peningkatan antara (+0.08-0.44 poin), sedangkan indeks 5 justru turun (-0.44 poin). Secara keseluruhan, peningkatan yang terjadi pada sebagian besar data cukup signifikan. Gambar 4.7 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.7 Hasil visualisasi model C

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.16.

Tabel 4.16 Hasil evaluasi model C

Model	Fitur	MAE
C	<i>Social_Hours_Per_Day</i> dan <i>Physical_Activity_Hours_Per_Day</i>	0.192672

Berdasarkan hasil yang ditampilkan pada Tabel 4.16, Hasil evaluasi menunjukkan nilai MAE (*Mean Absolute Error*) dari fitur tersebut adalah 0.192672.

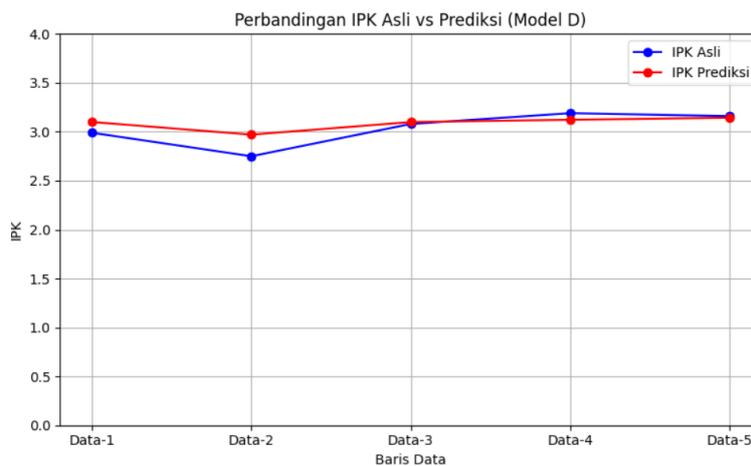
4.5.4 Hasil Kombinasi Model D

Pada model D ini, pengujian dilakukan dengan kombinasi fitur *Sleep_Hours_Per_Day* (Terbalik) dan *Social_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.17.

Tabel 4.17 Hasil prediksi model D

<i>Indeks</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	<i>Social_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
1	8.7	1	2.8	2.99	3.09
2	8	1	4.2	2.75	2.95
7	8	1	0.3	3.08	3.11
18	8.9	1	2	3.19	3.14
19	5.1	1	0.6	3.16	3.13

Berdasarkan Tabel 4.17 menunjukkan sebagian data mengalami peningkatan IPK prediksi, kecuali pada indeks 18 dan 19. Pada indeks 18 turun sedikit dari 3.19 ke 3.14 (-0.05 poin), sedangkan indeks 19 turun (-0.03 poin). Model ini menunjukkan hasil yang cenderung stabil dengan selisih kecil. Gambar 4.8 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.8 Hasil visualisasi model D

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.18.

Tabel 4.18 Hasil evaluasi model D

Model	Fitur	MAE
D	<i>Sleep_Hours_Per_Day</i> (Terbalik) dan <i>Social_Hours_Per_Day</i>	0.210519

Berdasarkan hasil yang ditampilkan pada Tabel 4.18, Hasil evaluasi menunjukkan nilai MAE (*Mean Absolute Error*) sebesar 0.210519, yang mana model D menggunakan kombinasi fitur *Sleep_Hours_Per_Day* (Terbalik) dengan nilai = 1 pada data.

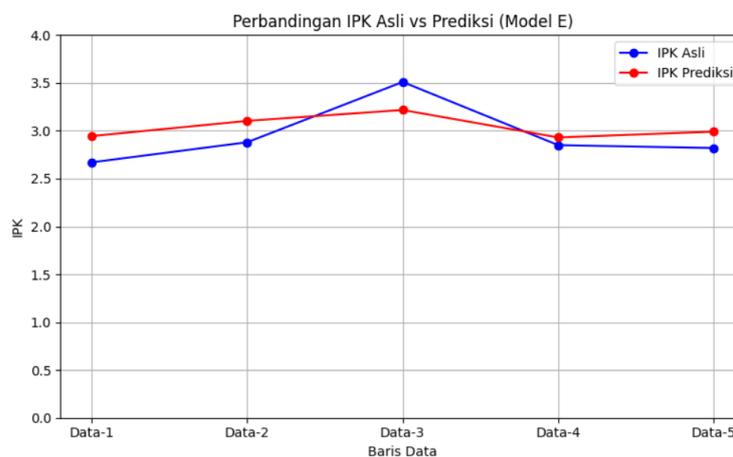
4.5.5 Hasil Kombinasi Model E

Pada model E ini, pengujian dilakukan dengan kombinasi fitur *Sleep_Hours_Per_Day* (Acak) dan *Physical_Activity_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.19.

Tabel 4.19 Hasil prediksi model E

<i>Indeks</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	<i>Physical_Activity_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
3	9.2	2	4.6	2.67	2.89
4	7.2	2	6.5	2.88	3.07
5	6.5	2	6.6	3.51	3.12
6	8	2	7.6	2.85	3.00
9	6.3	2	4.9	2.82	3.03

Tabel 4.19 menunjukkan peningkatan pada sebagian besar data. Pada indeks 3, 4, 6, dan 9 mengalami kenaikan yang cukup besar antara (+0.15-0.22 poin), sedangkan indeks 5 mengalami penurunan cukup signifikan (-0.39 poin).. Gambar 4.9 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.9 Hasil visualisasi model E

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.21.

Tabel 4.20 Hasil evaluasi model E

Model	Fitur	MAE
E	<i>Sleep_Hours_Per_Day (Terbalik)</i> dan <i>Physical_Activity_Hours_Per_Day</i>	0.190862

Berdasarkan hasil yang ditampilkan pada Tabel 4.20, Hasil evaluasi menunjukkan nilai MAE (*Mean Absolute Error*) sebesar 0.190862, yang mana model E menggunakan kombinasi fitur *Sleep_Hours_Per_Day* (Acak) dengan nilai = 2 pada data.

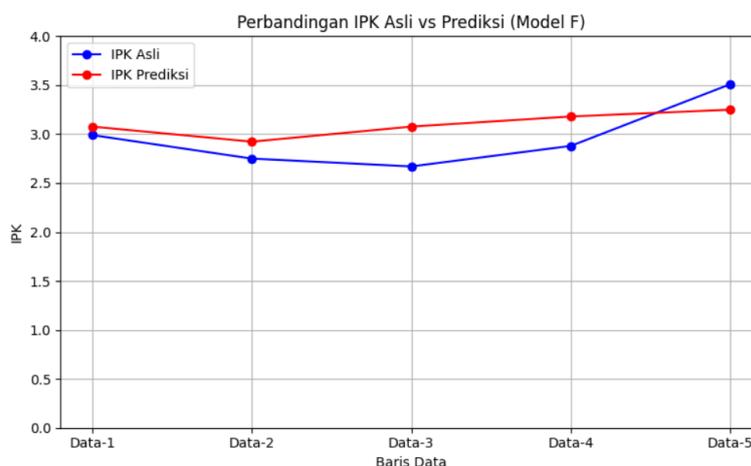
4.5.6 Hasil Kombinasi Model F

Pada model F ini, pengujian dilakukan dengan menggabungkan fitur *Extracurricular_Hours_Per_Day* dan *Social_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.21.

Tabel 4.21 Hasil prediksi model F

<i>Indeks</i>	<i>Extracurricular_Hours_Per_Day</i>	<i>Social_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
1	3.8	2.8	2.99	3.07
2	3.5	4.2	2.75	2.91
3	3.9	1.2	2.67	3.09
4	2.1	1.7	2.88	3.17
5	0.6	2.2	3.51	3.24

Pada Tabel 4.21 menunjukkan semua data mengalami peningkatan. Pada indeks 1 sampai 4 mengalami peningkatan antara (+0.08-0.42 poin), sedangkan indeks 5 mengalami penurunan (-0.27 poin). Meskipun satu data menurun, rata-rata menunjukkan hasil prediksi yang lebih tinggi. Gambar 4.10 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.10 Hasil visualisasi model F

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.22.

Tabel 4.22 Hasil evaluasi model F

Model	Fitur	MAE
F	<i>Extracurricular_Hours_Per_Day</i> dan <i>Social_Hours_Per_Day</i>	0.216275

Berdasarkan hasil yang ditampilkan pada Tabel 4.22, Hasil evaluasi menunjukkan nilai MAE (*Mean Absolute Error*) dari fitur tersebut adalah 0.216275.

4.5.7 Hasil Kombinasi Model G

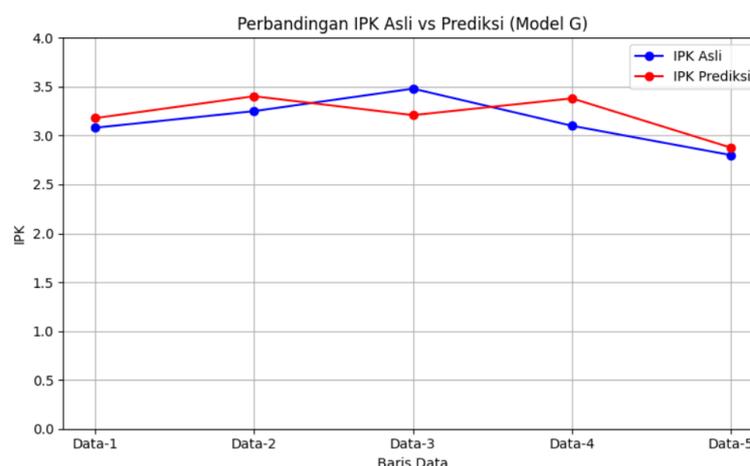
Pada model G ini, pengujian dilakukan dengan kombinasi fitur *Study_Hours_Per_Day* (Terbalik) dan *Sleep_Hours_Per_Day* (Terbalik). Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.23.

Tabel 4.23 Hasil prediksi model G

Indeks	<i>Study_Hours_Per_Day</i>	<i>Krt_Study</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	IPK Asli	IPK Prediksi
7	8	1	5.3	1	3.08	3.17
30	9.9	1	8.4	1	3.25	3.40

41	8.3	1	8.6	1	3.48	3.21
56	8.8	1	8.1	1	3.10	3.37
60	5.8	1	7.7	1	2.80	2.87

Tabel 4.23 menunjukkan Sebagian besar hasil prediksi mengalami peningkatan. Hasil menunjukkan pada indeks 7, 30, 56 dan 60 mengalami peningkatan antara (+0.09-0.27 poin). Namun, indeks 41 mengalami penurunan (-0.27 poin). Hasil menunjukkan kombinasi ini cukup efektif, meskipun tidak sepenuhnya konsisten. Gambar 4.11 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.11 Hasil visualisasi model G

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.24.

Tabel 4.24 Hasil evaluasi model G

Model	Fitur	MAE
G	<i>Study_Hours_Per_Day</i> (Terbalik) dan <i>Sleep_Hours_Per_Day</i> (Terbalik)	0.146818

Berdasarkan hasil yang tertera pada Tabel 4.25, evaluasi menunjukkan bahwa nilai MAE (*Mean Absolute Error*) yang diperoleh adalah sebesar 0.146818. Model G menggunakan kombinasi fitur *Study_Hours_Per_Day* (Terbalik) dan *Sleep_Hours_Per_Day* (Terbalik), dengan nilai 1 pada kedua fitur tersebut pada data yang digunakan.

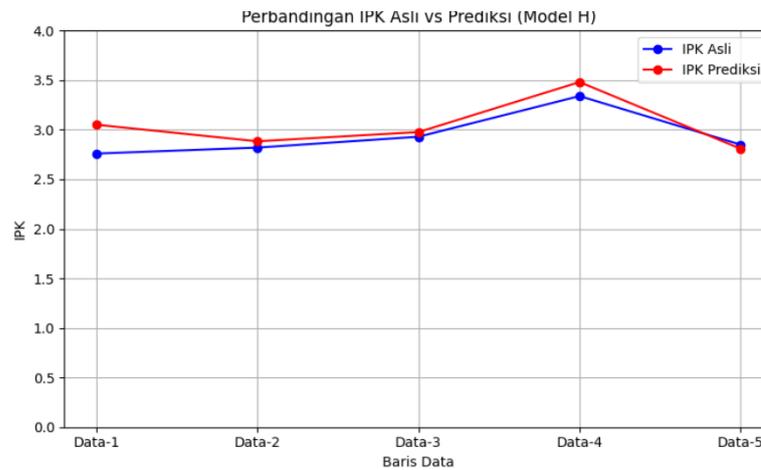
4.5.8 Hasil Kombinasi Model H

Pada model H ini, pengujian dilakukan dengan kombinasi fitur *Study_Hours_Per_Day* (Teratur) dan *Physical_Activity_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.25.

Tabel 4.25 Hasil prediksi model H

<i>Indeks</i>	<i>Study_Hours_Per_Day</i>	<i>Krt_Study</i>	<i>Physical_Activity_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
10	7.7	3	1.3	2.76	3.05
13	6.4	3	4.9	2.82	2.88
20	6.6	3	8.6	2.93	2.98
22	9.6	3	0.4	3.34	3.48
24	6.1	3	1.9	2.85	2.81

Tabel 4.25 menunjukkan sebagian besar data mengalami peningkatan IPK. Data pada indeks 10, 13, 20 dan 22 mengalami peningkatan signifikan antara (+0.05-0.29 poin), sedangkan satu-satunya penurunan terjadi pada indeks 24, dengan selisih -0.04 poin. Secara keseluruhan, model ini menunjukkan kecenderungan positif dalam prediksi IPK berdasarkan pola belajar teratur dan aktivitas fisik. Gambar 4.12 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.12 Hasil visualisasi model H

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.26.

Tabel 4.26 Hasil evaluasi model H

Model	Fitur	MAE
H	<i>Study_Hours_Per_Day</i> (Teratur) dan <i>Physical_Activity_Hours_Per_Day</i>	0.136139

Berdasarkan hasil yang tertera pada Tabel 4.26, evaluasi menunjukkan bahwa nilai MAE (*Mean Absolute Error*) yang diperoleh cukup rendah yaitu sebesar 0.136139. Model H menggunakan kombinasi fitur *Study_Hours_Per_Day* (Teratur), dengan nilai 2 pada data yang digunakan.

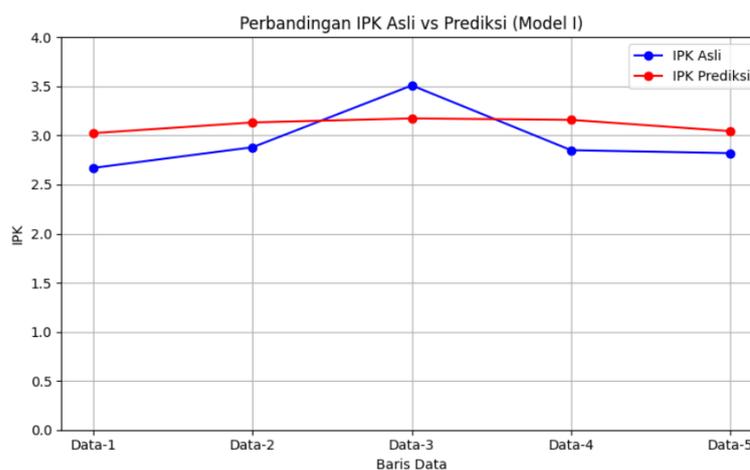
4.5.9 Hasil Kombinasi Model I

Pada model I ini, pengujian dilakukan dengan kombinasi fitur *Sleep_Hours_Per_Day* (Acak) dan *Extracurricular_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.27.

Tabel 4.27 Hasil prediksi model I

<i>Indeks</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	<i>Extracurricular_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
3	9.2	2	3.5	2.67	3.02
4	7.2	2	3.9	2.88	3.13
5	6.5	2	2.1	3.51	3.17
6	8	2	0.6	2.85	3.16
9	6.3	2	3.6	2.82	3.04

Pada tabel 4.27 hasil prediksi menunjukkan bahwa empat dari lima data mengalami peningkatan nilai IPK. Indeks 3, 4, 6, dan 9 mengalami kenaikan sekitar (+0.22-0.35 poin), sedangkan pada indeks 5, terjadi penurunan (-0.34 poin). Secara keseluruhan, model ini menunjukkan tren prediksi IPK yang meningkat, khususnya pada mahasiswa dengan pola tidur acak dan partisipasi dalam kegiatan ekstrakurikuler. Gambar 4.13 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.13 Hasil visualisasi model I

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.28.

Tabel 4.28 Hasil evaluasi model I

Model	Fitur	MAE
I	<i>Sleep_Hours_Per_Day</i> (Acak) dan <i>Extracurricular_Hours_Per_Day</i>	0.217579

Berdasarkan hasil yang tertera pada Tabel 4.28, evaluasi menunjukkan bahwa nilai MAE (*Mean Absolute Error*) yang diperoleh adalah 0.217579. Model I menggunakan kombinasi fitur *Sleep_Hours_Per_Day* (Acak), dengan nilai 2 pada data yang digunakan.

4.5.10 Hasil Kombinasi Model J

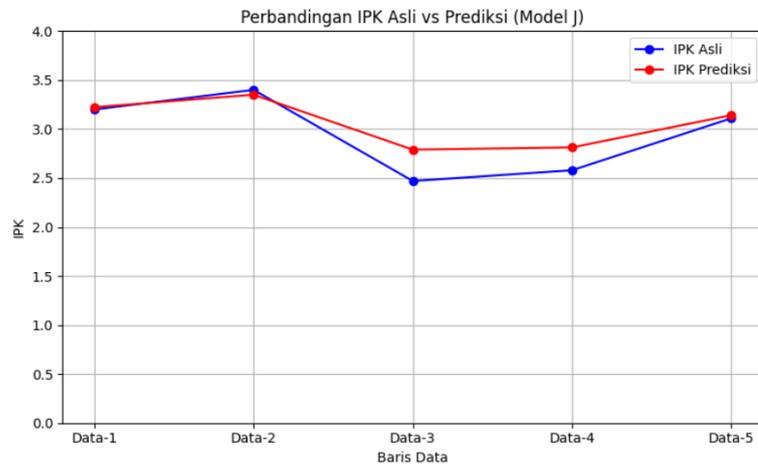
Pada model J ini, pengujian dilakukan dengan kombinasi fitur *Study_Hours_Per_Day* (Acak), *Sleep_Hours_Per_Day* (Teratur), dan *Social_Hours_Per_Day*. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.29.

Tabel 4.29 Hasil prediksi model J

Indeks	<i>Study_Hours_Per_Day</i>	<i>Krt_Study</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	<i>Social_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
8	8.4	2	5.6	3	3	3.20	3.22
15	8.9	2	6.8	3	0.7	3.40	3.35
33	5.2	2	8.4	3	0.1	2.47	2.79
37	5.5	2	9.5	3	3.4	2.58	2.81
54	7.3	2	6.4	3	3.5	3.11	3.14

Tabel 4.23 menunjukkan sebagian besar data mengalami peningkatan IPK. Indeks 8, 33, 37, dan 54 mengalami peningkatan signifikan antara (+0.02-0.32 poin). Satu-satunya penurunan terjadi pada indeks 15, yaitu (-0.05 poin). Secara keseluruhan, model ini menunjukkan hasil prediksi yang cukup stabil dan cenderung meningkat, terutama pada mahasiswa dengan variasi belajar yang acak namun pola tidur teratur dan aktivitas sosial yang terjaga. Gambar 4.14 dibawah ini

merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.14 Hasil visualisasi model J

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.30.

Tabel 4.30 Hasil evaluasi model J

Model	Fitur	MAE
J	<i>Study_Hours_Per_Day</i> (Acak), <i>Sleep_Hours_Per_Day</i> (Teratur), dan <i>Social_Hours_Per_Day</i>	0.129427

Berdasarkan hasil yang tertera pada Tabel 4.30, evaluasi menunjukkan bahwa nilai MAE (*Mean Absolute Error*) yang diperoleh adalah 0.129427. Model J menggunakan kombinasi fitur *Study_Hours_Per_Day* (Acak) dan *Sleep_Hours_Per_Day* (Teratur), dengan nilai 2 untuk acak dan 3 untuk teratur pada data yang digunakan.

4.5.11 Hasil Kombinasi Model K

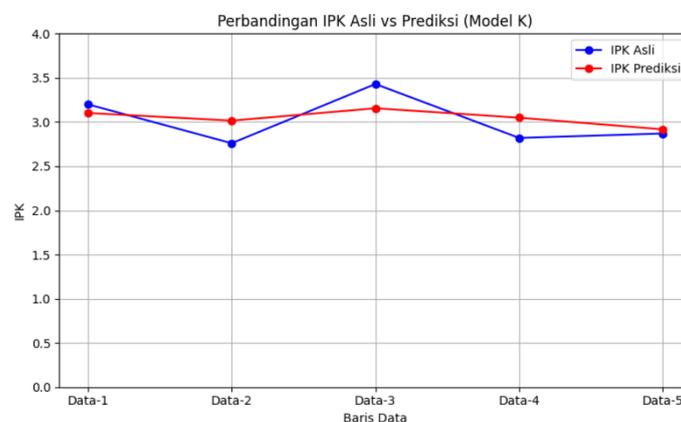
Pada model K ini, pengujian dilakukan dengan kombinasi fitur *Sleep_Hours_Per_Day* (Teratur), *Extracurricular_Hours_Per_Day* dan

Social_Hours_Per_Day. Hasil prediksi yang diperoleh dapat dilihat pada Tabel 4.31.

Tabel 4.31 Hasil prediksi model I

<i>Indeks</i>	<i>Sleep_Hours_Per_Day</i>	<i>Krt_Sleep</i>	<i>Extracurricular_Hours_Per_Day</i>	<i>Social_Hours_Per_Day</i>	IPK Asli	IPK Prediksi
8	5.6	3	1.8	3	3.20	3.10
10	9.8	3	0.7	4.5	2.76	3.02
11	8	3	3.6	2.5	3.43	3.16
13	5.7	3	2.2	4.8	2.82	3.05
14	8.5	3	3.3	4.4	2.87	2.92

Tabel 4.31 menunjukkan adanya variasi hasil antara peningkatan dan penurunan. Data indeks 10 mengalami peningkatan yang cukup besar dari 2.76 menjadi 3.02 (+0.26 poin), diikuti oleh indeks 13 yang naik sebesar +0.23 poin, dan indeks 14 naik ringan sebesar +0.05 poin. Namun, dua data mengalami penurunan, yakni indeks 8 (-0.10 poin) dan indeks 11 (-0.27 poin). Secara keseluruhan, meskipun model ini menunjukkan peningkatan pada sebagian data, hasil prediksinya tidak sepenuhnya stabil, kemungkinan karena interaksi kompleks antara aktivitas sosial dan ekstrakurikuler terhadap IPK. Gambar 4.11 dibawah ini merupakan hasil visualisasi antara data asli dan data prediksi hasil kombinasi model dari data diatas.



Gambar 4.15 Hasil visualisasi model K

Setelah memperoleh hasil prediksi, dilakukan pengukuran performa model menggunakan MAE, sebagaimana ditampilkan pada Tabel 4.32.

Tabel 4.32 Hasil evaluasi model K

Model	Fitur	MAE
K	<i>Sleep_Hours_Per_Day</i> (Teratur), <i>Extracurricular_Hours_Per_Day</i> dan <i>Social_Hours_Per_Day</i>	0.200322

Berdasarkan hasil yang tertera pada Tabel 4.32, evaluasi menunjukkan bahwa nilai MAE (*Mean Absolute Error*) yang diperoleh adalah sebesar 0.146818. Model K menggunakan kombinasi fitur *Sleep_Hours_Per_Day* (Teratur), dengan nilai 3 pada data yang digunakan.

Setelah semua kombinasi fitur dilakukan pengujian, perbandingan hasil akhir MAE dari setiap kombinasi fitur dilakukan guna mencari kombinasi fitur terbaik pada model. Berikut adalah hasil akhir MAE pada setiap model.

Tabel 4.33 Perbandingan hasil evaluasi setiap model

Model	Rata-rata MAE
A	0.138892
B	0.140464
C	0.192672
D	0.210519
E	0.190862
F	0.216275
G	0.146818
H	0.136139
I	0.217579
J	0.129427
K	0.200322

Tabel 4.33 menyajikan perbandingan hasil evaluasi rata-rata MAE dari setiap model yang diuji. Model J menghasilkan nilai MAE paling rendah sebesar 0.129427, menunjukkan akurasi prediksi yang paling baik dibandingkan model lainnya. Disusul oleh Model H dengan MAE sebesar 0.136139, kemudian Model A

sebesar 0.138892, dan Model B sebesar 0.140464. Model G menyusul dengan nilai MAE sebesar 0.146818, sementara model-model lainnya memiliki tingkat kesalahan yang lebih tinggi seperti Model C (0.192672), Model E (0.190862), dan Model K (0.200322). Dua model dengan nilai MAE tertinggi adalah Model F (0.216275) dan Model I (0.217579), yang menunjukkan tingkat akurasi prediksi yang paling rendah di antara seluruh model.

4.6 Hasil Pengujian *Feature Importance*

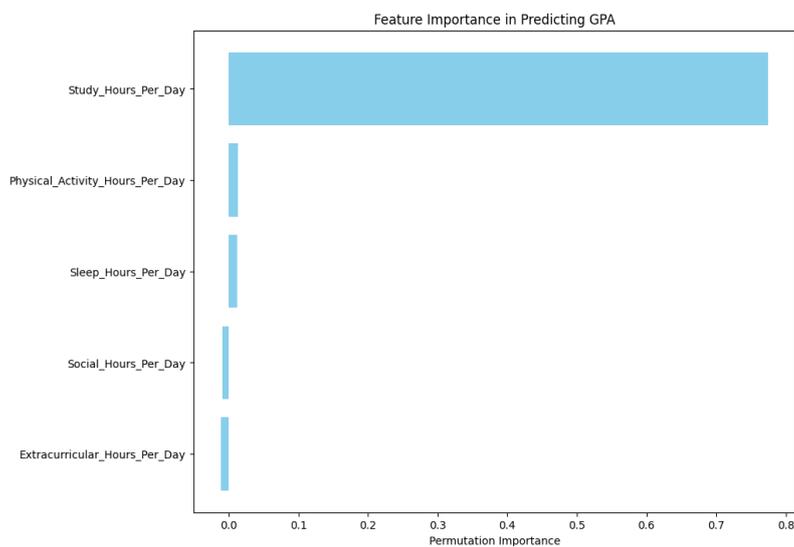
Pengujian ini dilakukan untuk mengetahui seberapa besar pengaruh masing-masing fitur gaya hidup terhadap hasil prediksi IPK mahasiswa menggunakan model *Random Forest*. Metode *Permutation Importance* bekerja dengan cara mengacak nilai setiap fitur dan mengukur sejauh mana performa model menurun. Jika performa model turun drastis ketika sebuah fitur diacak, berarti fitur tersebut penting. Berikut adalah hasil dari *Feature Importance*.

Tabel 4.34 Hasil *Feature Importance*

Fitur	Rata-rata <i>Importance</i>	Standar Deviasi
<i>Study_Hours_Per_Day</i>	0.775065	0.089594
<i>Physical_Activity_Hours_Per_Day</i>	0.013126	0.007897
<i>Sleep_Hours_Per_Day</i>	0.012265	0.007480
<i>Social_Hours_Per_Day</i>	-0.009407	0.004117
<i>Extracurricular_Hours_Per_Day</i>	-0.011674	0.006357

Berdasarkan Tabel 4.26, fitur *Study_Hours_Per_Day* memiliki nilai rata-rata importance tertinggi sebesar 0.775, menunjukkan bahwa jam belajar per hari adalah faktor paling berpengaruh dalam prediksi IPK. Fitur lain seperti *Physical_Activity_Hours_Per_Day* dan *Sleep_Hours_Per_Day* memiliki pengaruh kecil, sedangkan *Social_Hours_Per_Day* dan *Extracurricular_Hours_Per_Day*

justru berdampak negatif terhadap performa model. Ini menunjukkan bahwa waktu belajar memiliki peran dominan dalam mempengaruhi hasil prediksi IPK. Visualisasi dari hasil feature importance dapat dilihat pada Gambar 4.12 berikut.



Gambar 4.16 Hasil Visual *Feature Importance*

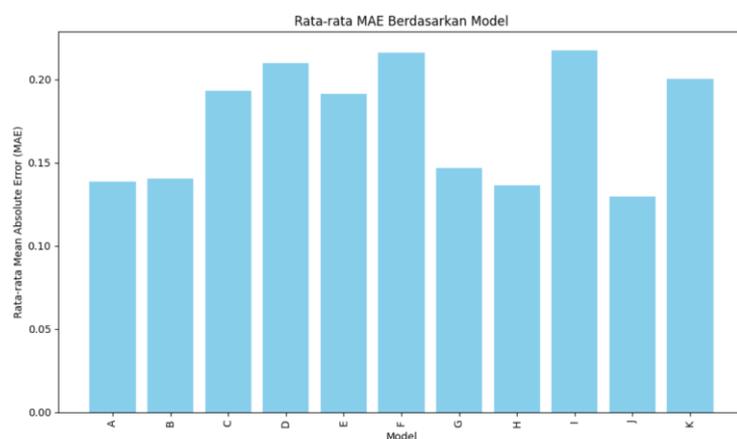
4.7 Pembahasan

Penelitian ini bertujuan untuk mengevaluasi kinerja model *Random Forest* dalam memprediksi Indeks Prestasi Kumulatif (IPK) mahasiswa berdasarkan pola gaya hidup harian mereka. Data yang digunakan berasal dari 2000 mahasiswa dengan berbagai variabel gaya hidup, seperti durasi belajar, durasi tidur, kegiatan ekstrakurikuler, interaksi sosial, dan aktivitas fisik. Berbagai pengujian dilakukan untuk menguji akurasi model, termasuk penyesuaian *hyperparameter*, uji variasi kombinasi fitur, dan pengujian fitur penting. Evaluasi ini bertujuan untuk menilai sejauh mana model dapat memprediksi IPK dengan akurat.

Untuk meningkatkan akurasi model dalam prediksi IPK, dilakukan penyesuaian *hyperparameter* menggunakan metode *gridsearch* guna mencari kombinasi parameter yang optimal. Hasil tuning menunjukkan bahwa penggunaan

parameter seperti $n_estimators = 300$, $max_depth = 10$, $min_samples_split = 10$, $min_samples_leaf = 4$, $max_features = Sqrt$, dan $Bootstrap = True$ terbukti meningkatkan kinerja model secara signifikan. Hasil prediksi pada pengujian dengan parameter terbaik menunjukkan bahwa model menghasilkan prediksi yang sangat mendekati nilai asli, seperti pada Student ID 1861 dengan prediksi IPK 2.90, yang sangat mendekati IPK asli 2.88.

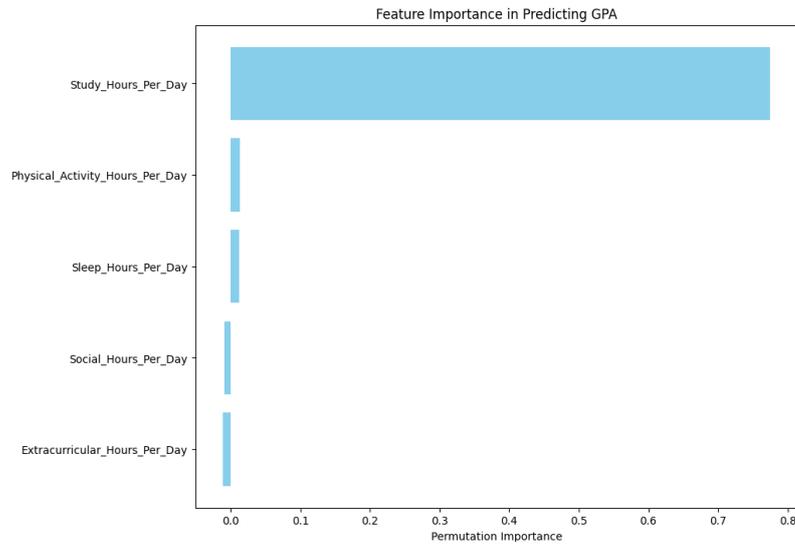
Evaluasi lebih lanjut dilakukan dengan menguji berbagai kombinasi fitur gaya hidup untuk mengetahui fitur mana yang paling berpengaruh terhadap prediksi IPK. Pengujian dilakukan menggunakan *5-Fold Cross Validation* dengan metrik *Mean Absolute Error* (MAE). Hasil pengujian menunjukkan bahwa kombinasi fitur *Study_Hours_Per_Day* (Acak), *Sleep_Hours_Per_Day* (Teratur), dan *Social_Hours_Per_Day* pada Model J menghasilkan nilai MAE terendah sebesar 0.129427, yang menunjukkan bahwa kombinasi fitur ini paling efektif dalam memprediksi IPK mahasiswa. Sebaliknya, kombinasi fitur yang melibatkan *Extracurricular_Hours_Per_Day* dan *Social_Hours_Per_Day* pada Model F memberikan hasil yang kurang efektif dengan nilai MAE tertinggi sebesar 0.216275.



Gambar 4.17 Hasil evaluasi MAE setiap model kombinasi

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa Model J, yang mengombinasikan jam belajar acak, tidur yang teratur, dan aktivitas sosial mahasiswa, merupakan model dengan performa prediksi terbaik. Hal ini mengindikasikan bahwa kombinasi antara fleksibilitas belajar, pola tidur yang stabil, dan interaksi sosial yang teratur dapat menjadi indikator penting dalam pencapaian akademik. Secara keseluruhan, model *Random Forest* menunjukkan potensi yang besar dalam memprediksi IPK mahasiswa berdasarkan gaya hidup mereka. Temuan ini dapat dijadikan dasar untuk merancang strategi yang lebih tepat dalam membantu mahasiswa mengelola waktu dan kebiasaan mereka untuk mencapai prestasi akademik yang optimal.

Selain itu, hasil analisis *feature importance* menunjukkan bahwa *Study_Hours_Per_Day* merupakan fitur paling berpengaruh dalam prediksi IPK, dengan nilai rata-rata importance sebesar 0.775, jauh di atas fitur lainnya. Sementara itu, fitur seperti *Physical_Activity_Hours_Per_Day* dan *Sleep_Hours_Per_Day* memiliki pengaruh rendah, dan fitur *Social_Hours_Per_Day* serta *Extracurricular_Hours_Per_Day* bahkan menunjukkan kontribusi negatif terhadap performa model. Visualisasi hasil *feature importance* ditampilkan pada gambar 4.14 berikut.



Gambar 4.18 Feature Importance

4.8 Integrasi Penelitian

Dalam Al-Qur'an, konsep prediksi dalam Islam bukanlah hal yang asing. Hal ini sejalan dengan prinsip yang terdapat dalam Surat Ar-Rum ayat 2-5, yang berbunyi:

عَلَيْتِ الرُّومُ ﴿٢﴾ فِي أَدْنَى الْأَرْضِ وَهُمْ مِنْ بَعْدِ غَلَبِهِمْ سَعْلَبُونَ ﴿٣﴾ فِي بَضْعِ سِنِينَ ۗ لِلَّهِ الْأَمْرُ مِنْ قَبْلُ وَمِنْ بَعْدُ ۗ وَيَوْمَئِذٍ يُفْرِخُ الْمُؤْمِنُونَ ﴿٤﴾ يَنْصُرُ اللَّهُ ۗ يَنْصُرُ مَنْ يَشَاءُ ۗ وَهُوَ الْعَزِيزُ الرَّحِيمُ ﴿٥﴾

“Bangsa Romawi telah dikalahkan, di negeri yang terdekat dan mereka setelah kekalahannya itu akan menang dalam beberapa tahun (lagi). Milik Allahlah urusan sebelum dan setelah (mereka menang). Pada hari (kemenangan bangsa Romawi) itu bergembiralah orang-orang mukmin karena pertolongan Allah. Dia menolong siapa yang Dia kehendaki. Dia MahaPerkasa lagi Maha Penyayang.” (QS. Ar-Rum: 2-5)

Dalam ayat diatas, Allah SWT berfirman bahwa bangsa Romawi yang sebelumnya kalah akan mendapatkan kemenangan dalam beberapa tahun ke depan. Prediksi kemenangan tersebut tampak mustahil secara kasat mata pada saat itu,

karena secara kekuatan militer bangsa Romawi telah kalah. Namun, melalui wahyunya, Allah memberikan kabar gembira tentang masa depan yang belum terjadi. Hal ini menjadi bukti bahwa prediksi terhadap suatu kejadian di masa depan merupakan bagian dari pengetahuan yang dapat dicapai oleh manusia dalam batas kemampuannya, sebagaimana Allah juga menunjukkan bahwa realitas tidak selalu dapat dipahami hanya dari kondisi saat ini.

Hal ini sejalan dengan penelitian ini, di mana algoritma *machine learning* seperti *Random Forest* digunakan untuk memprediksi Indeks Prestasi Kumulatif (IPK) mahasiswa. Melalui data gaya hidup harian mahasiswa, sistem dapat memperkirakan IPK yang mungkin dicapai, sebagai bentuk usaha ilmiah manusia dalam memahami pola dan tren yang ada. Dengan demikian, prediksi dalam konteks teknologi dapat dipandang sebagai bentuk ikhtiar manusia yang sejalan dengan nilai-nilai Islam, selama tetap menyadari bahwa pengetahuan yang hakiki tetap milik Allah SWT.

Selain itu, prinsip dalam Surat Al-Qamar ayat 4-5 mengingatkan kita akan pentingnya kejujuran dan keterbukaan dalam setiap aspek kehidupan, yang berbunyi:

لَقَدْ جَاءَهُمْ مِنَ الْأَنْبَاءِ مَا فِيهِ مُرْدَجَةٌ ﴿٤﴾ حِكْمَةٌ بَلِغَةٌ ۗ فَمَا تُغْنِي الْتُّذُرُ ﴿٥﴾

“Sungguh, benar-benar telah datang kepada mereka beberapa berita yang di dalamnya terdapat ancaman (terhadap orang-orang kafir). (Berita-berita itu adalah) hikmah yang sempurna, tetapi peringatan-peringatan itu tidak berguna (bagi mereka).” (QS. Al-Qamar: 4-5)

Dalam ayat ini, Allah menegaskan bahwa setiap peringatan dan kebenaran yang dibawa oleh para rasul adalah hakikat yang nyata dan tidak boleh

diputarbalikkan. Kejujuran adalah fondasi dalam menuntut ilmu dan menyampaikan hasil penelitian. Oleh karena itu, dalam penelitian ini penulis menyampaikan secara terbuka bahwa data yang digunakan berasal dari dataset publik yang diunduh dari platform *Kaggle*, yang bersifat sintetis atau bukan data nyata dari lapangan. Meskipun demikian, data ini tetap valid untuk tujuan simulasi dan pengembangan model prediksi IPK mahasiswa. Dengan menyampaikan hal ini secara jujur, penulis berusaha menerapkan prinsip integritas ilmiah yang sejalan dengan ajaran Islam.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model *Random Forest* dapat digunakan secara efektif untuk memprediksi IPK mahasiswa berdasarkan pola gaya hidup harian mereka. Model ini menunjukkan performa yang cukup baik dengan nilai evaluasi akhir R^2 sebesar 0.5287, MAE sebesar 0.1683, dan RMSE sebesar 0.0440. Artinya, model mampu menjelaskan sekitar 52% variasi data IPK dengan tingkat kesalahan prediksi yang relatif kecil.

Penelitian ini juga berhasil mengidentifikasi faktor-faktor gaya hidup yang berpengaruh terhadap prediksi IPK, seperti durasi belajar, durasi tidur, aktivitas fisik, interaksi sosial, dan keterlibatan dalam kegiatan ekstrakurikuler. Hasil *feature importance* menunjukkan bahwa *Study_Hours_Per_Day* merupakan fitur paling dominan dalam model, disusul oleh *Sleep_Hours_Per_Day*, sementara fitur seperti *Social_Hours_Per_Day* dan *Extracurricular_Hours_Per_Day* memiliki kontribusi yang rendah atau bahkan negatif terhadap akurasi prediksi.

Evaluasi kombinasi fitur menunjukkan bahwa model paling akurat diperoleh dari kombinasi *Study_Hours_Per_Day* (Acak), *Sleep_Hours_Per_Day* (Teratur), dan *Social_Hours_Per_Day*, sebagaimana diterapkan pada Model J, dengan nilai MAE terendah sebesar 0.1294. Hasil ini mengindikasikan bahwa fleksibilitas dalam pola belajar, dikombinasikan dengan kebiasaan tidur yang teratur serta keterlibatan sosial, merupakan indikator yang paling berpengaruh dalam memprediksi prestasi

akademik mahasiswa. Penyesuaian hyperparameter menggunakan metode grid search juga berhasil meningkatkan performa model. Konfigurasi terbaik diperoleh dengan parameter $n_estimators = 300$, $max_depth = 10$, $min_samples_split = 10$, $min_samples_leaf = 4$, $max_features = sqrt$, dan $bootstrap = True$.

Berdasarkan temuan ini, mahasiswa disarankan untuk menjaga pola belajar dan tidur yang konsisten sebagai bagian dari strategi peningkatan prestasi akademik. Meskipun aktivitas sosial dan ekstrakurikuler penting untuk keseimbangan hidup, pengaruhnya terhadap IPK tidak sekuat dua faktor utama tersebut dalam konteks model prediktif ini.

5.2 Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang dapat diberikan untuk penelitian selanjutnya serta implementasi lebih lanjut terkait penggunaan algoritma *Random Forest* dalam memprediksi IPK mahasiswa berdasarkan pola gaya hidup harian:

1. Menambahkan variabel lain yang dapat mempengaruhi kinerja akademik mahasiswa.
2. Menambahkan metode lain sebagai perbandingan dan eksplorasi untuk menemukan hasil yang terbaik.
3. Dapat mencoba bahasa pemrograman lain menggunakan paket *Random forest*.
4. Menggunakan lebih banyak data, agar model dapat menghasilkan prediksi yang lebih baik karena memiliki informasi yang lebih bervariasi.

DAFTAR PUSTAKA

- Adriansyah, M. A., & Muhliansyah, M. (2022). Improving Personal Life in Mental Health to Build Work-Life Balance in The Era of Digitalization. *Jurnal Aisyah: Jurnal Ilmu Kesehatan*, 7(4), 1173–1182. <https://doi.org/10.30604/jika.v7i4.1387>
- Akbar, A., Indra, Z., Andriyani, Y., & Melia, T. (2023). Implementation Of The Random Forest Method For Predicting Students'length Of Study. *Journal of Statistical Methods and Data Science*, 1(2). <https://doi.org/10.31258/jsmds.v1i2.15>
- Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. *Interactive Learning Environments*, 31(6), 3360–3379. <https://doi.org/10.1080/10494820.2021.1928235>
- Baita, A., Prasetyo, I. A., & Cahyono, N. (2023). Hyperparameter Tuning On Random Forest For Diagnose Covid-19. *JIKO (Jurnal Informatika Dan Komputer)*, 6(2). <https://doi.org/10.33387/jiko.v6i2.6389>
- Bakar, A. A., Osman, M. M., & Hitam, M. (2020). Personality and Lifestyle Interprets External Condition to Environmental Behaviours. *Planning Malaysia*, 18. <https://doi.org/10.21837/pm.v18i11.709>
- Behr, A., Giese, M., Tegui K, H. D., & Theune, K. (2020). Early Prediction of University Dropouts – A Random Forest Approach. *Jahrbücher Für Nationalökonomie Und Statistik*, 240(6), 743–789. <https://doi.org/10.1515/jbnst-2019-0006>
- Botache, D., Dingel, K., Huhnstock, R., Ehresmann, A., & Sick, B. (2023). Unraveling the Complexity of Splitting Sequential Data: Tackling Challenges in Video and Time Series Analysis. *arXiv Preprint arXiv:2307.14294*. <https://doi.org/10.48550/arXiv.2307.14294>
- Brivio, F., Viganò, A., Paterna, A., Palena, N., & Greco, A. (2023). Narrative review and analysis of the use of “Lifestyle” in health psychology. *International Journal of Environmental Research and Public Health*, 20(5), 4427. <https://doi.org/10.3390/ijerph20054427>
- Brown, D. M. Y., Porter, C., Hamilton, F., Almanza, F., Narvid, C., Pish, M., & Arizabalo, D. (2022). Interactive Associations between Physical Activity and Sleep Duration in Relation to Adolescent Academic Achievement.

International Journal of Environmental Research and Public Health, 19(23), 15604. <https://doi.org/10.3390/ijerph192315604>

- Buijs, V. L., Jeronimus, B. F., Lodder, G. M., Riediger, M., Luong, G., & Wrzus, C. (2023). Interdependencies between family and friends in daily life: Personality differences and associations with affective well-being across the lifespan. *European Journal of Personality*, 37(2), 154–170. <https://doi.org/10.1177/08902070211072745>
- Demir, S. (2023). Bibliometrics Analysis on Using Machine Learning Algorithms in Teacher Education Researches. *Journal of Research in Didactical Sciences*. <https://doi.org/10.51853/jorids/14202>
- Elian, V., Ozon, E. A., Rachminov, Y. C., Varlas, V., & Ghica, M. (2023). Lifestyle, Physical Activity And Nutritional Habits In A Group Of Specialists In Diabetes. *Farmacia*, 71(6), 1274–1281. <https://doi.org/10.31925/farmacia.2023.6.18>
- Evstatiev, B. I., & Gabrovska-Evstatieva, K. G. (2021). A review on the methods for big data analysis in agriculture. *IOP Conference Series: Materials Science and Engineering*, 1032(1), 012053. <https://doi.org/10.1088/1757-899X/1032/1/012053>
- Falát, L., & Piscová, T. (2022). Predicting GPA of University Students with Supervised Regression Machine Learning Models. *Applied Sciences*, 12(17), 8403. <https://doi.org/10.3390/app12178403>
- Himanshu J. Dhide Tycs, Sneha S. Choudhary Sycs, & Dr. Akshata A. Nayak. (2025). The Impact of Sleep Patterns on the Academic Performance of Undergraduate Students: A Behavioral and Cognitive Study. *The Voice of Creative Research*, 7(2), 410–418. <https://doi.org/10.53032/tvcr/2025.v7n2.50>
- Huang, C. (2024). Teaching Evaluation Research of Professional Courses Based on Multinomial Random Forest With Improved Grassberger Entropy. *IEEE Access*, 12, 91154–91163. <https://doi.org/10.1109/ACCESS.2024.3421542>
- Huynh-Cam, T.-T., Chen, L.-S., & Huynh, K.-V. (2022). Learning Performance of International Students and Students with Disabilities: Early Prediction and Feature Selection through Educational Data Mining. *Big Data and Cognitive Computing*, 6(3), 94. <https://doi.org/10.3390/bdcc6030094>
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics*, 10(11), 1942. <https://doi.org/10.3390/math10111942>

- Lone, R. A. (2021). Study habits and academic performance among students: A systematic review. *International Journal of Multidisciplinary*, 6(5), 132–135. <https://doi.org/10.31305/rrijm.2021.v06.i05.019>
- Nunik, F. (2021). Peranan IPK dan Soft Skill dalam Proses Rekrutmen dan Seleksi Fresh Graduate: Tanggapan dan Harapan Perusahaan Peserta Job Fair di UKSW. *Jibaku: Jurnal Ilmiah Bisnis, Manajemen Dan Akuntansi*, 1(1), 42–47. <https://doi.org/10.35473/jibaku.v1i1.955>
- Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). Predicting student success using big data and machine learning algorithms. *International Journal of Emerging Technologies in Learning (iJET)*, 17(12), 236–251. <https://doi.org/10.3991/ijet.v17i12.30259>
- Papadogiannis, I., Pouloupoulos, V., Platis, N., Vassilakis, C., Lepouras, G., & Wallace, M. (2023). First Grade GPA as a Predictor of Later Academic Performance in High School. *Knowledge*, 3(3), 513–524. <https://doi.org/10.3390/knowledge3030033>
- Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 729–735. <https://doi.org/10.1109/ICSSIT48917.2020.9214160>
- Sakız, H., Özdaş, F., Göksu, İ., & Ekinci, A. (2021). A longitudinal analysis of academic achievement and its correlates in higher education. *SAGE Open*, 11(1), 21582440211003085. <https://doi.org/10.1177/21582440211003085>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/BJML/2024/007>
- Savargiv, M., Masoumi, B., & Keyvanpour, M. R. (2021). A new random forest algorithm based on learning automata. *Computational Intelligence and Neuroscience*, 2021(1), 5572781. <https://doi.org/10.1155/2021/5572781>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
- Seoane, H. A., Moschetto, L., Orliacq, F., Orliacq, J., Serrano, E., Cazenave, M. I., Vigo, D. E., & Perez-Lloret, S. (2020). Sleep disruption in medicine students and its relationship with impaired academic performance: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 53, 101333. <https://doi.org/10.1016/j.smrv.2020.101333>

- Shafie, S. N. M., Shahri, M. S., Izuddi, N. N. N. I., Shukri, N. M., Aziz, N. A., Amran, A., & Nafi, M. N. A. (2022). *Association between Lifestyle-Related Behaviors and Academic Performance among Students*. 82(1), 105.
<https://doi.org/10.3390/proceedings2022082105>
- Sirgy, M. J. (2024). Balanced Life. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 370–373). Springer.
<https://doi.org/10.1007/s10902007-9074-1>
- Sterner, E. A. (2020). Impact of academic libraries on grade point average (GPA): A review. *Performance Measurement and Metrics*, 22(1), 71–85.
<https://doi.org/10.1108/PMM-01-2020-0004>
- Teferi, G. (2020). The effect of physical activity on academic performance and mental health: Systematic review. *Am. J. Sci. Eng. Technol*, 5, 131–136.
<https://doi.org/10.11648/J.AJSET.20200503.12>
- Vives, L., Cabezas, I., Vives, J. C., Reyes, N. G., Aquino, J., C ndor, J. B., & Altamirano, S. F. S. (2024). Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks. *IEEE Access*, 12, 5882–5898.
<https://doi.org/10.1109/ACCESS.2024.3350169>
- Zhao, L., Chen, K., Song, J., Zhu, X., Sun, J., Caulfield, B., & Namee, B. M. (2021). Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data. *IEEE Access*, 9, 5453–5465.
<https://doi.org/10.1109/ACCESS.2020.3002791>