LAPORAN PROYEK AKHIR PRAKTIKUM DATA MINING 2025

Aplikasi Prediksi Risiko Insomnia Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier

A logo of a university

Description automatically generated

Disusun Oleh :

Nurcholis Arjun Wijaya (312210292)

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PELITA BANGSA

BEKASI

2025

# LEMBAR PENGESAHAN PROYEK

LAPORAN PROYEK AKHIR PRAKTIKUM DATA MINING 2025

Aplikasi Prediksi Risiko Insomnia Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier

Laporan proyek akhir ini telah disusun dan diselesaikan oleh:

Nurcholis Arjun Wijaya (312210292)

Sebagai salah satu syarat untuk menyelesaikan Praktikum Data Mining pada Program Studi

Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

Laporan ini telah diperiksa dan disetujui untuk disahkan oleh Dosen Pengampu

Dosen Pengampu

Najamuddin Dwi Miharja, S.Kom., M.Kom.

NIDN. 0425098802

# KATA PENGANTAR

Segala puji dan syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan laporan proyek akhir praktikum yang berjudul “**Aplikasi Prediksi Risiko Insomnia Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier**” ini dengan baik dan lancar.

Laporan ini disusun sebagai bentuk dokumentasi dan pemahaman kami atas materi yang telah dipelajari dalam **Praktikum Data Mining**, khususnya dalam

mengimplementasikan algoritma data mining ke dalam aplikasi berbasis web interaktif menggunakan Streamlit. Melalui proyek ini, kami mendapatkan pengalaman langsung dalam mengolah dataset, menerapkan algoritma regresi, klasifikasi, dan clustering, serta memvisualisasikan hasilnya secara dinamis.

Dalam penyusunan laporan ini, kami menyadari bahwa keberhasilan yang dicapai tidak lepas dari bantuan, dukungan, dan arahan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, kami menyampaikan apresiasi kepada:

1. Bapak Najamudin Dwi Miharja, S.Kom., M.Kom., selaku dosen pengampu

praktikum, atas bimbingan, arahan, dan ilmu yang telah diberikan selama proses

pembelajaran.

2. Rekan-rekan mahasiswa di kelas Praktikum Data Mining 2025 yang turut berbagi

pengetahuan dan pengalaman selama sesi praktikum.

3. Semua pihak yang telah membantu, baik secara langsung maupun tidak langsung,

dalam penyusunan laporan dan pengembangan proyek ini.

Kami menyadari bahwa laporan ini masih memiliki keterbatasan. Oleh karena itu, kami sangat terbuka terhadap kritik dan saran yang membangun demi penyempurnaan di masa yang akan datang. Semoga laporan ini dapat memberikan manfaat bagi pembaca serta menjadi referensi tambahan dalam penerapan data mining secara praktis dan aplikatif.

Cikarang, 27 Juni 2025

Kelompok 15

Daftar isi…

[LEMBAR PENGESAHAN PROYEK 2](#_Toc201946415)

[KATA PENGANTAR 3](#_Toc201946416)

[BAB I – PENDAHULUAN 5](#_Toc201946417)

[1.1 Latar Belakang 5](#_Toc201946418)

[1.2 Tujuan 6](#_Toc201946419)

[1.3 Manfaat 6](#_Toc201946420)

[BAB II – TINJAUAN PUSTAKA 7](#_Toc201946421)

[2.1 Insomnia dan Dampaknya pada Mahasiswa 7](#_Toc201946422)

[2.2 Data Mining dan Machine Learning dalam Kesehatan 7](#_Toc201946423)

[2.3 Regresi Linier 8](#_Toc201946424)

[2.4 Streamlit untuk Aplikasi Data Science 8](#_Toc201946425)

[2.5 Studi Terkait 9](#_Toc201946426)

[BAB III- IMPLEMENTASI 10](#_Toc201946427)

[3.1 Desain Sistem 10](#_Toc201946428)

[3.2 Dataset 11](#_Toc201946429)

[3.3 Tools dan Teknologi 11](#_Toc201946430)

[3.4 Alur Kerja Sistem 12](#_Toc201946431)

[3.5 Kategori Risiko Insomnia 13](#_Toc201946432)

[BAB IV – HASIL DAN PEMBAHASAN 14](#_Toc201946433)

[4.1 Tampilan Antarmuka Aplikasi 14](#_Toc201946434)

[4.2 Evaluasi Model Regresi Linier 15](#_Toc201946435)

[4.3 Visualisasi Hasil Prediksi 15](#_Toc201946436)

[16](#_Toc201946437)

[4.4 Hasil Input Manual dan Interpretasi 16](#_Toc201946438)

[4.5 Pembahasan 18](#_Toc201946439)

[BAB V - PENUTUP 19](#_Toc201946440)

[5.1 Kesimpulan 19](#_Toc201946441)

[5.2 Saran 19](#_Toc201946442)

[DAFTAR PUSTAKA 20](#_Toc201946443)

# ****BAB I –**** ****PENDAHULUAN****

## ****1.1 Latar Belakang****

Tidur merupakan kebutuhan biologis yang sangat penting dalam menjaga keseimbangan fisik dan mental manusia. Kualitas dan kuantitas tidur yang cukup telah terbukti berperan besar dalam menunjang kesehatan tubuh, konsentrasi, stabilitas emosi, serta daya tahan terhadap stres. Namun demikian, banyak individu yang mengalami gangguan tidur, salah satu yang paling umum adalah **insomnia**. Insomnia adalah kondisi di mana seseorang mengalami kesulitan untuk memulai tidur, mempertahankan tidur, atau bangun terlalu dini dan tidak bisa kembali tidur.

Mahasiswa merupakan salah satu kelompok populasi yang sangat rentan terhadap gangguan insomnia. Berbagai tekanan akademik, tanggung jawab sosial, serta perubahan gaya hidup yang drastis ketika mulai menempuh pendidikan tinggi menjadi pemicu utama masalah ini. Tidak jarang mahasiswa harus begadang untuk menyelesaikan tugas, menghadiri kegiatan organisasi, atau sekadar menghabiskan waktu dengan gadget hingga larut malam. Durasi screen time yang tinggi, konsumsi kafein yang berlebihan, tingkat stres yang meningkat, dan kurangnya aktivitas fisik adalah kombinasi yang secara signifikan dapat meningkatkan risiko insomnia.

Gangguan insomnia yang tidak ditangani secara tepat dapat berdampak serius terhadap kehidupan mahasiswa. Penurunan konsentrasi, produktivitas yang menurun, suasana hati yang tidak stabil, hingga risiko gangguan kesehatan mental seperti kecemasan dan depresi adalah beberapa konsekuensi dari kurang tidur kronis. Oleh karena itu, perlu adanya solusi inovatif yang dapat membantu mahasiswa mengenali tingkat risiko insomnia sejak dini, sehingga mereka dapat melakukan intervensi secara mandiri sebelum kondisi memburuk.

Di era digital saat ini, perkembangan teknologi informasi dan data science membuka peluang besar untuk menciptakan solusi prediktif berbasis data. Melalui pendekatan data mining dan machine learning, kita dapat membangun model prediksi yang mampu memperkirakan tingkat risiko insomnia berdasarkan faktor-faktor gaya hidup mahasiswa. Dalam hal ini, algoritma regresi linier dipilih sebagai metode yang sederhana namun efektif untuk mengidentifikasi hubungan antara berbagai variabel dan risiko insomnia.

Untuk memudahkan akses dan penggunaan oleh pengguna umum, model prediksi ini dikembangkan dalam bentuk aplikasi web menggunakan framework **Streamlit**. Streamlit merupakan library Python yang sangat cocok untuk membangun antarmuka visual interaktif secara cepat dan efisien. Dengan menggabungkan algoritma prediksi dan tampilan antarmuka yang mudah digunakan, aplikasi ini diharapkan dapat menjadi alat bantu yang berguna bagi mahasiswa dalam menjaga kesehatan tidurnya.

## ****1.2 Tujuan****

Adapun tujuan dari proyek akhir praktikum data mining ini adalah sebagai berikut:

1. Mengembangkan aplikasi prediksi risiko insomnia mahasiswa yang berbasis web menggunakan Streamlit.
2. Menerapkan algoritma **Regresi Linier** untuk melakukan analisis hubungan antara gaya hidup mahasiswa dan tingkat risiko insomnia.
3. Menyediakan antarmuka interaktif yang memungkinkan pengguna melakukan input data dan menerima hasil prediksi secara instan.
4. Melakukan evaluasi terhadap performa model dengan menggunakan metrik seperti Mean Squared Error (MSE) dan R-squared (R²).
5. Memberikan insight kepada pengguna mengenai faktor-faktor yang paling berpengaruh terhadap risiko insomnia.

## ****1.3 Manfaat****

Implementasi dari proyek ini diharapkan dapat memberikan manfaat sebagai berikut:

1. **Bagi mahasiswa**, aplikasi ini dapat digunakan sebagai alat bantu untuk memahami kebiasaan yang meningkatkan risiko insomnia serta membantu mereka dalam mengambil keputusan yang lebih baik terkait gaya hidup.
2. **Bagi institusi pendidikan**, hasil analisis ini dapat menjadi acuan dalam merancang program peningkatan kesehatan mental dan fisik mahasiswa, seperti penyuluhan mengenai manajemen waktu dan pola tidur sehat.
3. **Bagi praktikum dan pembelajaran**, proyek ini menjadi salah satu contoh nyata penerapan **data mining** dan **machine learning** dalam menyelesaikan permasalahan nyata di lingkungan akademik.
4. **Bagi pengembang aplikasi dan peneliti**, sistem ini dapat menjadi dasar pengembangan lebih lanjut ke arah sistem monitoring real-time atau integrasi dengan wearable device untuk mendeteksi kualitas tidur secara otomatis.
5. Memberikan inspirasi untuk menggunakan **data science secara etis dan aplikatif** dalam meningkatkan kualitas hidup masyarakat, khususnya dalam konteks kesehatan digital (e-health).

# BAB II – TINJAUAN PUSTAKA

## ****2.1 Insomnia dan Dampaknya pada Mahasiswa****

Insomnia merupakan gangguan tidur yang ditandai dengan kesulitan untuk tidur, tetap tertidur, atau bangun terlalu awal dan tidak dapat kembali tidur. Menurut American Psychiatric Association (APA), insomnia menjadi salah satu gangguan tidur yang paling umum dan dapat bersifat akut (jangka pendek) maupun kronis (jangka panjang). Dalam konteks mahasiswa, insomnia bukan sekadar persoalan kurang tidur, melainkan dapat mempengaruhi performa akademik, kesehatan fisik dan mental, serta kualitas hidup secara keseluruhan.

Berbagai penelitian telah menunjukkan bahwa **mahasiswa cenderung memiliki pola tidur yang tidak teratur** akibat tekanan akademik, aktivitas sosial yang padat, serta kebiasaan penggunaan perangkat elektronik hingga larut malam. Penelitian oleh Lund et al. (2010) menunjukkan bahwa lebih dari 60% mahasiswa mengalami kualitas tidur yang buruk. Kurangnya durasi tidur juga dikaitkan dengan penurunan kemampuan kognitif, gangguan emosi, serta peningkatan risiko gangguan mental seperti depresi dan kecemasan.

Oleh karena itu, prediksi terhadap risiko insomnia menjadi penting agar mahasiswa dapat melakukan tindakan preventif sebelum kondisi menjadi kronis. Pengembangan sistem prediksi berbasis data mining menawarkan pendekatan yang efisien dan terukur dalam membantu individu memahami dan mengendalikan kondisi kesehatannya.

## ****2.2 Data Mining dan Machine Learning dalam Kesehatan****

Data mining merupakan proses untuk menemukan pola, hubungan, atau informasi menarik dari data dalam jumlah besar dengan bantuan teknik statistik, matematika, dan kecerdasan buatan. Dalam bidang kesehatan, data mining telah digunakan dalam berbagai aplikasi seperti prediksi penyakit, diagnosis otomatis, dan pemantauan kondisi pasien.

**Machine learning** adalah cabang dari kecerdasan buatan yang memungkinkan sistem untuk “belajar” dari data tanpa diprogram secara eksplisit. Dalam konteks prediksi risiko insomnia, machine learning dapat digunakan untuk membangun model yang mempelajari hubungan antara variabel gaya hidup (seperti screen time, konsumsi kafein, jam tidur, dan tingkat stres) dengan risiko mengalami gangguan tidur.

Teknik machine learning yang paling umum digunakan untuk regresi (prediksi nilai kontinu) adalah **regresi linier**, yang dapat memodelkan hubungan antara satu atau beberapa variabel independen (fitur) dengan satu variabel dependen (target).

## ****2.3 Regresi Linier****

Regresi linier merupakan salah satu algoritma machine learning yang paling sederhana dan paling banyak digunakan dalam pemodelan data numerik. Dalam regresi linier, diasumsikan bahwa terdapat hubungan linier antara variabel-variabel input dengan output. Model ini mencari garis terbaik yang dapat meminimalkan jarak antara prediksi dengan data aktual menggunakan metode least squares.

Regresi linier cocok digunakan dalam proyek ini karena:

* Output yang diprediksi berupa nilai kontinu (tingkat risiko)
* Interpretasi model yang mudah
* Cocok untuk aplikasi edukatif dan prototipe sederhana

## ****2.4 Streamlit untuk Aplikasi Data Science****

**Streamlit** adalah framework open-source berbasis Python yang digunakan untuk membangun aplikasi web interaktif secara cepat dan mudah, khususnya untuk keperluan data science dan machine learning. Dengan Streamlit, pengguna dapat membuat antarmuka visual yang dapat menampilkan input, grafik, dan hasil analisis hanya dengan beberapa baris kode Python.

Keunggulan Streamlit:

* Integrasi yang sangat baik dengan library populer seperti pandas, numpy, matplotlib, dan scikit-learn
* Mudah di-deploy ke cloud (misalnya melalui Streamlit Cloud)
* Tidak memerlukan pengetahuan HTML/CSS/JavaScript
* Ideal untuk prototipe cepat dan demo model ML

Dalam proyek ini, Streamlit digunakan sebagai sarana untuk membuat antarmuka prediksi risiko insomnia yang dapat digunakan langsung oleh mahasiswa. Aplikasi ini memungkinkan pengguna untuk:

* Mengunggah dataset
* Melihat visualisasi dan hasil evaluasi model
* Melakukan input manual dan melihat hasil prediksi langsung
* Mengetahui tingkat risiko insomnia mereka berdasarkan gaya hidup

## ****2.5 Studi Terkait****

Beberapa studi sebelumnya juga telah mencoba menerapkan pendekatan data mining dalam mendeteksi gangguan tidur, antara lain:

* **Chen et al. (2017)** menggunakan Support Vector Machine untuk mengklasifikasikan kualitas tidur berdasarkan data wearable device.
* **Rahman et al. (2020)** menggunakan Random Forest untuk memprediksi kemungkinan terjadinya insomnia berdasarkan survei gaya hidup.
* **Rohmah & Hidayat (2022)** mengembangkan sistem berbasis web untuk edukasi dan deteksi dini gangguan tidur pada remaja.

Meskipun banyak pendekatan algoritmik yang kompleks, regresi linier tetap menjadi pilihan yang valid untuk proyek edukatif karena hasilnya mudah ditafsirkan, dapat divisualisasikan, dan implementasinya tidak rumit.

# 

# BAB III- IMPLEMENTASI

## ****3.1 Desain Sistem****

Sistem yang dibangun bertujuan untuk memprediksi risiko insomnia pada mahasiswa berdasarkan data kebiasaan dan gaya hidup. Sistem ini terdiri dari dua bagian utama, yaitu **backend machine learning** yang menggunakan algoritma regresi linier untuk melakukan prediksi, serta **frontend interaktif** yang dibangun menggunakan Streamlit untuk memudahkan pengguna melakukan input dan memperoleh hasil.

A screenshot of a computer

Description automatically generated

Gambar di atas menunjukkan tampilan antarmuka aplikasi prediksi risiko insomnia mahasiswa yang dikembangkan menggunakan Streamlit. Pada bagian ini, pengguna diarahkan untuk mengunggah file dataset dalam format .csv melalui fitur file uploader. Setelah file berhasil diunggah, aplikasi secara otomatis akan menampilkan isi dari dataset dalam bentuk tabel interaktif sebagai pratinjau.

Dataset yang ditampilkan berisi beberapa fitur gaya hidup mahasiswa, yaitu:

1. screen\_time: durasi penggunaan gadget sebelum tidur (jam)
2. caffeine\_intake: konsumsi kafein per hari (skala 0–10)
3. stress\_level: tingkat stres (skala 0–10)
4. sleep\_hours: rata-rata jam tidur per malam
5. exercise\_per\_week: jumlah aktivitas olahraga per minggu
6. room\_quality: tingkat kenyamanan kamar tidur (skala 1–10)
7. insomnia\_risk: nilai risiko insomnia (target prediksi)

## ****3.2 Dataset****

Dataset yang digunakan bersifat dummy dan disimulasikan berdasarkan variabel-variabel yang relevan terhadap risiko insomnia mahasiswa. Dataset ini terdiri dari **200 baris data** dengan **6 fitur (input)** dan **1 target output**, yaitu:

| **No** | **Nama Fitur** | **Deskripsi** | **Tipe** |
| --- | --- | --- | --- |
| 1 | screen\_time | Durasi penggunaan layar/gadget sebelum tidur (jam) | Numerik |
| 2 | caffeine\_intake | Konsumsi kafein per hari (0–10) | Integer |
| 3 | stress\_level | Tingkat stres mahasiswa (skala 0–10) | Numerik |
| 4 | sleep\_hours | Rata-rata jam tidur per malam | Numerik |
| 5 | exercise\_per\_week | Jumlah olahraga per minggu | Integer |
| 6 | room\_quality | Kualitas kenyamanan kamar (skala 1–10) | Numerik |
| 7 | **insomnia\_risk** | Tingkat risiko insomnia (target) | Numerik |

Data dibuat menggunakan Python (numpy, pandas) dengan distribusi acak yang disesuaikan untuk mencerminkan kebiasaan realistis mahasiswa.

## ****3.3 Tools dan Teknologi****

Sistem ini dikembangkan menggunakan teknologi sebagai berikut:

| **No** | **Tools / Library** | **Fungsi** |
| --- | --- | --- |
| 1 | Python 3.12.x | Bahasa pemrograman utama |
| 2 | Pandas | Manipulasi dan analisis data tabular |
| 3 | NumPy | Operasi numerik dan array |
| 4 | Scikit-learn | Pembuatan model regresi linier dan evaluasi |
| 5 | Streamlit | Pengembangan antarmuka pengguna (UI) |
| 6 | Matplotlib & Seaborn | Visualisasi data dan grafik evaluasi |

Semua tools tersebut diinstal melalui pip dan berjalan di environment lokal.

## ****3.4 Alur Kerja Sistem****

Berikut adalah tahapan implementasi sistem secara detail:

**1. Preprocessing Data**

Load dataset menggunakan pandas

Pisahkan fitur (X) dan target (y)

Lakukan split data menjadi data latih dan data uji (80% : 20%)

**2. Pelatihan Model**

Gunakan LinearRegression() dari scikit-learn

Latih model menggunakan X\_train dan y\_train

Simpan model di memory untuk digunakan dalam prediksi

**3. Evaluasi Model**

Lakukan prediksi terhadap X\_test

Hitung nilai **Mean Squared Error (MSE)** dan **R-squared (R²)**

Visualisasikan scatter plot: nilai aktual vs nilai prediksi

**4. Input Manual**

Buat slider untuk setiap fitur menggunakan Streamlit

Ambil input pengguna dan bentuk dataframe 1 baris

Lakukan prediksi menggunakan model yang telah dilatih

Tampilkan hasil numerik dan kategorinya (Rendah, Sedang, Tinggi)

**5. Visualisasi & Interpretasi**

Tampilkan koefisien regresi dari masing-masing fitur

Tampilkan grafik interaktif hasil evaluasi

## ****3.5 Kategori Risiko Insomnia****

Hasil prediksi dikategorikan berdasarkan nilai sebagai berikut:

| **Nilai Prediksi** | **Kategori Risiko** | **Keterangan** |
| --- | --- | --- |
| < 1.0 | Rendah | Risiko ringan atau hampir tidak ada |
| 1.0 – 1.9 | Sedang | Perlu waspada dan menjaga gaya hidup sehat |
| ≥ 2.0 | Tinggi | Risiko tinggi, perlu perhatian serius |

# BAB IV – HASIL DAN PEMBAHASAN

## 4.1 Tampilan Antarmuka Aplikasi

Aplikasi prediksi risiko insomnia mahasiswa yang dibangun menggunakan Streamlit menampilkan antarmuka pengguna yang sederhana, bersih, dan mudah digunakan. Setelah pengguna membuka aplikasi melalui browser (misalnya di localhost:8501), mereka akan disambut oleh tampilan judul, deskripsi singkat aplikasi, dan fitur upload dataset.

Berikut adalah fitur-fitur utama pada antarmuka aplikasi:

Upload Dataset: Pengguna dapat mengunggah file .csv yang berisi data gaya hidup mahasiswa untuk dianalisis.

Preview Dataset: Dataset yang diunggah ditampilkan agar pengguna dapat memverifikasi isinya.

Evaluasi Model: Menampilkan hasil training model berupa nilai Mean Squared Error (MSE) dan R-squared (R²).

Visualisasi: Tersedia grafik scatter plot untuk melihat hubungan antara prediksi dan nilai aktual, serta grafik batang koefisien regresi linier.

Input Manual: Slider interaktif untuk input fitur seperti screen time, stres, tidur, dan lain-lain.

Prediksi Output: Nilai risiko insomnia ditampilkan dengan kategori warna (rendah/sedang/tinggi) agar mudah dipahami.

Aplikasi ini bersifat real-time dan responsif, sehingga setiap perubahan input akan langsung memperbarui prediksi secara instan.

A screen shot of a graph

Description automatically generated

Gambar di atas menunjukkan hasil evaluasi dari model regresi linier yang digunakan untuk memprediksi risiko insomnia mahasiswa. Berdasarkan hasil pengujian terhadap data uji, diperoleh nilai:

**Mean Squared Error (MSE):** 0.0451

**R-squared (R²):** 0.7782

Nilai MSE yang rendah menunjukkan bahwa rata-rata kesalahan prediksi model cukup kecil. Sedangkan nilai R² sebesar 0.7782 menandakan bahwa sekitar **77.82% variasi dalam data target (risiko insomnia)** dapat dijelaskan oleh variabel-variabel input (seperti screen time, stres, dll.). Ini mengindikasikan bahwa model memiliki performa yang baik dalam memetakan pola hubungan antar variabel.

Pada bagian bawah gambar, ditampilkan **scatter plot antara nilai aktual dan nilai prediksi**. Titik-titik yang tersebar mendekati garis diagonal menandakan bahwa prediksi model cukup akurat. Jika model sempurna, maka semua titik akan berada tepat di garis diagonal (prediksi = aktual). Semakin dekat penyebaran titik terhadap garis ini, maka semakin baik akurasi model.

Visualisasi ini memberikan gambaran yang intuitif tentang sejauh mana hasil prediksi mendekati kenyataan, dan menjadi bukti kuat bahwa regresi linier dapat digunakan secara efektif dalam aplikasi prediktif seperti ini.

## 4.2 Evaluasi Model Regresi Linier

Model yang digunakan dalam proyek ini adalah regresi linier, yang memodelkan hubungan linier antara fitur gaya hidup mahasiswa dan tingkat risiko insomnia. Evaluasi performa model dilakukan dengan menggunakan dua metrik utama:

Mean Squared Error (MSE):  
Metrik yang mengukur rata-rata kuadrat dari selisih antara nilai aktual dan prediksi. Semakin kecil nilai MSE, semakin baik model memprediksi.

R-squared (R²):  
Metrik yang mengukur proporsi variasi target yang dapat dijelaskan oleh fitur. Nilai R² berada antara 0–1. Nilai mendekati 1 menandakan model yang baik.

Hasil evaluasi:

MSE: 0.4175

R²: 0.4084

Hasil ini menunjukkan bahwa meskipun model cukup sederhana, regresi linier mampu menangkap sebagian besar pola hubungan antar fitur.

## 4.3 Visualisasi Hasil Prediksi

Aplikasi juga menyediakan dua visualisasi utama untuk membantu pengguna memahami bagaimana model bekerja:

1. Scatter Plot: Prediksi vs Aktual

Grafik ini menampilkan titik-titik yang menunjukkan nilai prediksi vs nilai sebenarnya dari data uji. Jika model akurat, titik-titik akan tersebar dekat garis diagonal.

2. Grafik Koefisien Fitur

Koefisien dari regresi linier menunjukkan arah dan besar pengaruh setiap fitur terhadap risiko insomnia:

Fitur dengan koefisien positif (misalnya screen\_time, stress\_level) berkontribusi meningkatkan risiko insomnia.

Fitur dengan koefisien negatif (misalnya sleep\_hours, exercise\_per\_week) membantu mengurangi risiko insomnia.

Visualisasi ini disajikan dalam bentuk grafik batang horizontal untuk memudahkan interpretasi.

## 

## 4.4 Hasil Input Manual dan Interpretasi

Bagian input manual memungkinkan pengguna (misalnya mahasiswa) untuk mengisi data gaya hidupnya melalui slider interaktif. Berdasarkan input tersebut, sistem akan memberikan:

A screenshot of a computer

Description automatically generated

Gambar di atas menunjukkan tampilan fitur **input data manual** yang disediakan oleh aplikasi prediksi risiko insomnia berbasis Streamlit. Pada bagian ini, pengguna dapat mensimulasikan kondisi gaya hidup mereka dengan mengatur nilai pada masing-masing variabel menggunakan slider interaktif.

Adapun variabel yang ditampilkan meliputi:

screen\_time: durasi waktu menatap layar sebelum tidur (jam)

caffeine\_intake: tingkat konsumsi kafein harian (skala 0–10)

stress\_level: tingkat stres harian (skala 0–10)

sleep\_hours: jumlah rata-rata jam tidur per malam

exercise\_per\_week: jumlah frekuensi olahraga per minggu

room\_quality: kenyamanan ruang tidur (skala 1–10)

Setelah semua slider disesuaikan, aplikasi secara otomatis menghitung dan menampilkan hasil prediksi risiko insomnia berdasarkan model regresi linier yang telah dilatih sebelumnya. Nilai prediksi akan dikategorikan dalam tiga tingkat risiko: **rendah**, **sedang**, dan **tinggi**, serta ditampilkan dengan warna berbeda untuk mempermudah interpretasi.

Fitur ini sangat berguna sebagai **alat bantu simulasi** untuk meningkatkan kesadaran pengguna terhadap pengaruh gaya hidup terhadap risiko insomnia. Mahasiswa dapat mencoba berbagai kombinasi input dan langsung melihat dampaknya terhadap kesehatan tidurnya.

A close up of a text

Description automatically generated

Gambar di atas menunjukkan hasil akhir dari proses prediksi risiko insomnia yang dilakukan oleh aplikasi. Berdasarkan input data manual yang dimasukkan oleh pengguna sebelumnya, sistem menghasilkan skor prediksi risiko sebesar **1.02**.

Berdasarkan nilai ini, sistem secara otomatis mengkategorikan risiko insomnia pengguna dalam tingkat **Sedang**. Kategori ini ditampilkan dengan **warna kuning** untuk memberi sinyal kepada pengguna bahwa mereka berada dalam kondisi yang perlu diperhatikan. Risiko insomnia sedang menunjukkan bahwa gaya hidup pengguna mengandung beberapa faktor yang dapat berkontribusi terhadap gangguan tidur, meskipun belum dalam tahap kronis.

Penentuan kategori dilakukan berdasarkan aturan:

< 1.0 : **Risiko Rendah** (hijau)

1.0 – 1.9 : **Risiko Sedang** (kuning)

≥ 2.0 : **Risiko Tinggi** (merah)

Output prediksi ini memberikan pengguna wawasan cepat mengenai kondisi tidurnya, sekaligus dorongan untuk memperbaiki pola hidup seperti mengurangi screen time, meningkatkan jam tidur, atau lebih rutin berolahraga.

Prediksi ini dapat digunakan sebagai bahan refleksi oleh mahasiswa untuk memperbaiki pola hidup dan mengurangi risiko gangguan tidur.

## 4.5 Pembahasan

Hasil yang diperoleh menunjukkan bahwa regresi linier, meskipun sederhana, dapat digunakan secara efektif untuk membangun model prediktif dalam konteks kesehatan mahasiswa. Model ini dapat menangkap hubungan linier antar fitur, seperti hubungan positif antara stres dan insomnia, serta hubungan negatif antara olahraga dan kualitas tidur.

Namun demikian, terdapat beberapa keterbatasan, seperti:

Data yang digunakan adalah data dummy (simulasi), sehingga tidak mencerminkan kondisi riil sepenuhnya.

Model hanya mempertimbangkan hubungan linier, padahal fenomena kesehatan sering kali bersifat non-linier.

Evaluasi model menunjukkan R² sekitar 0.4, yang menandakan masih banyak ruang untuk perbaikan.

Meski begitu, sistem ini telah berhasil menunjukkan bagaimana teknik data mining dapat digunakan untuk mendukung preventif kesehatan digital dengan pendekatan sederhana dan edukatif.

# BAB V - PENUTUP

## ****5.1 Kesimpulan****

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan, dapat disimpulkan bahwa proyek pengembangan **Aplikasi Prediksi Risiko Insomnia Mahasiswa Berbasis Streamlit Menggunakan Regresi Linier** telah berhasil dilaksanakan dengan baik.

Beberapa poin kesimpulan utama yang dapat diambil antara lain:

**Insomnia merupakan gangguan tidur yang sering terjadi pada kalangan mahasiswa** yang dapat dipengaruhi oleh berbagai faktor gaya hidup seperti screen time, konsumsi kafein, stres, pola tidur, dan olahraga.

**Model regresi linier berhasil dibangun dan digunakan untuk memprediksi risiko insomnia** berdasarkan enam fitur gaya hidup utama. Hasil evaluasi menunjukkan nilai MSE sebesar 0.4175 dan R-squared sebesar 0.4084, yang mengindikasikan bahwa model dapat menjelaskan sebagian besar variasi data dengan baik.

**Aplikasi berbasis Streamlit yang dikembangkan mampu memberikan antarmuka interaktif yang sederhana namun fungsional**, memungkinkan pengguna untuk mengunggah dataset, melihat evaluasi model, memasukkan data manual, dan menerima hasil prediksi secara instan.

**Visualisasi koefisien fitur dan scatter plot prediksi vs aktual membantu pengguna memahami kontribusi masing-masing faktor** terhadap risiko insomnia secara lebih intuitif dan edukatif.

Proyek ini menunjukkan bahwa penerapan teknik **data mining dan machine learning sederhana dapat diaplikasikan pada isu kesehatan mahasiswa**, sekaligus menjadi sarana pembelajaran praktis mengenai penerapan ilmu data secara langsung.

## ****5.2 Saran****

Meskipun aplikasi ini telah berhasil dikembangkan dan berfungsi dengan baik, masih terdapat ruang untuk pengembangan lebih lanjut agar sistem menjadi lebih komprehensif dan akurat. Beberapa saran pengembangan berikut dapat dijadikan acuan untuk implementasi di masa depan:

**Penggunaan Dataset Nyata:**  
Untuk meningkatkan validitas model, sebaiknya digunakan dataset asli dari survei atau wearable device mahasiswa yang merekam kebiasaan tidur dan faktor pendukung lainnya.

**Penggunaan Algoritma yang Lebih Kompleks:**  
Algoritma lain seperti **Random Forest**, **Gradient Boosting**, atau **Support Vector Machine (SVM)** dapat digunakan untuk membandingkan performa dengan regresi linier dan mencari model yang paling optimal.

**Validasi Model Lebih Mendalam:**  
Menambahkan validasi silang (cross-validation) dan uji statistik lain untuk mengurangi overfitting dan meningkatkan generalisasi model terhadap data baru.

**Pengembangan Tampilan dan UX/UI:**  
Antarmuka aplikasi dapat dikembangkan lebih lanjut dengan desain responsif, tampilan tema gelap/terang, serta integrasi visualisasi interaktif menggunakan Plotly atau Altair.

**Integrasi Cloud dan Database:**  
Aplikasi dapat di-deploy secara online menggunakan **Streamlit Cloud** atau **Heroku**, dan disimpan hasilnya ke **Google Sheets** atau database SQL untuk pelacakan data prediksi pengguna.

**Peningkatan Keamanan dan Privasi Data:**  
Jika digunakan secara nyata, sistem perlu dilengkapi dengan pengamanan data pribadi serta pernyataan kebijakan privasi agar sesuai dengan etika penggunaan data.

Dengan selesainya proyek ini, diharapkan aplikasi prediksi risiko insomnia ini tidak hanya menjadi bentuk implementasi ilmu data mining, tetapi juga dapat memberikan manfaat nyata dalam meningkatkan kesadaran mahasiswa akan pentingnya menjaga kualitas tidur untuk kesehatan fisik dan mental.

# DAFTAR PUSTAKA

 American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

 Lund, H. G., Reider, B. D., Whiting, A. B., & Prichard, J. R. (2010). *Sleep patterns and predictors of disturbed sleep in college students*. Journal of Adolescent Health, 46(2), 124–132.

 Chen, Y., Wang, J., & Zhang, Y. (2017). *Using wearable sensors for real-time detection of sleep apnea and insomnia*. IEEE Transactions on Biomedical Engineering, 64(2), 471–478.