Tugas Besar
Pemrograman untuk Data Analitik

Analisis Data Exam Score Prediction

Disusun Oleh

Bryan P. Hutagalung (18222130)
Andang Kurniawan (23524061)
Silvia Rahma (23525021)

PROGRAM STUDI MAGISTER INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG

2025

Daftar Isi

Daftar Gambar

1. Dataset

a. Deksripsi Dataset

Dataset Exam Score Prediction berisi 20.000 entri yang merepresentasikan faktor perilaku akademik, kebiasaan belajar, gaya hidup, dan kondisi ujian. Atribut-atribut seperti jam belajar, kehadiran kelas, kualitas tidur, metode belajar, akses internet, dan tingkat kesulitan ujian digunakan untuk memahami variasi performa akademik siswa. Dataset ini dipilih karena strukturnya yang realistis, mayoritas atribut sudah dalam format numerik/kategorikal yang mudah diproses, serta relevan untuk membangun model prediksi nilai ujian.

b. Deskripsi Teknis Dataset

1) Ukuran Dataset

Jumlah baris 20000
Jumlah kolom 13

2) Daftar dan Deskripsi Atribut

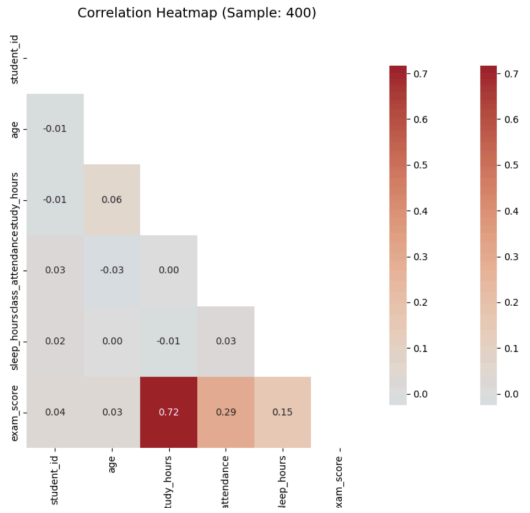
No	Kolom	Tipe Data	Deskripsi
1	student_id	int	ID unik setiap siswa
2	age	int	Usia siswa
3	gender	object	Jenis kelamin (male/female)
4	course	object	Mata pelajaran yang diikuti
5	study_hours	float/int	Waktu belajar per hari
6	class_attendance	float/int	Persentase kehadiran siswa
7	internet_access	object	Akses internet di rumah (yes/no)
8	sleep_hours	float/int	Lama tidur per hari
9	sleep_quality	object	Kualitas tidur (good/poor)
10	study_method	object	Metode belajar (group/self)
11	facility_rating	int	Penilaian fasilitas belajar
12	exam_difficulty	int	Tingkat kesulitan ujian
13	exam_score	int/float	Target variabel, skor ujian

2. Exploratory Data Analysis (Sebelum Pre-Processing)

a. Correlation Heatmap

Heatmap korelasi digunakan untuk melihat hubungan awal antar variabel numerik pada dataset mentah. Perhitungan koefisien Pearson dilakukan pada seluruh fitur numerik, serta diverifikasi melalui animasi yang meningkatkan ukuran sampel secara bertahap untuk memastikan kestabilan pola korelasi.

Hasil visualisasi menunjukkan bahwa study_hours memiliki korelasi paling kuat dengan exam_score (≈ 0,72), diikuti oleh class_attendance dan sleep_hours yang menunjukkan korelasi positif moderat. Variabel lain cenderung memiliki korelasi rendah, sehingga risiko multikolinearitas antar prediktor utama relatif kecil. Pola korelasi juga terlihat stabil pada berbagai ukuran sampel, sehingga temuan ini dapat dianggap cukup robust sebagai dasar analisis lanjutan.



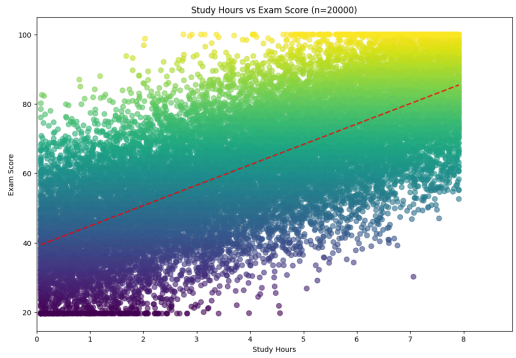
Gambar 1. ***Heatmap korelasi data sebelum preprocessing.***

b. Study Hours vs Exam Score

Untuk mengevaluasi hubungan antara durasi belajar dan performa ujian, dilakukan visualisasi scatter plot antara study_hours dan exam_score yang dilengkapi dengan garis regresi linear. Pendekatan ini digunakan untuk melihat pola sebaran individual sekaligus kecenderungan umum hubungan linear antar variabel.

Hasil visualisasi menunjukkan adanya hubungan positif yang jelas: siswa dengan jam belajar lebih tinggi cenderung memperoleh nilai ujian yang lebih baik. Garis tren linear mempertegas bahwa study_hours merupakan salah satu prediktor kuat terhadap exam_score. Meskipun terdapat variabilitas pada titik data, pola

peningkatan nilai yang konsisten menunjukkan bahwa hubungan ini stabil dan robust, sehingga layak digunakan sebagai dasar dalam tahap pemodelan.

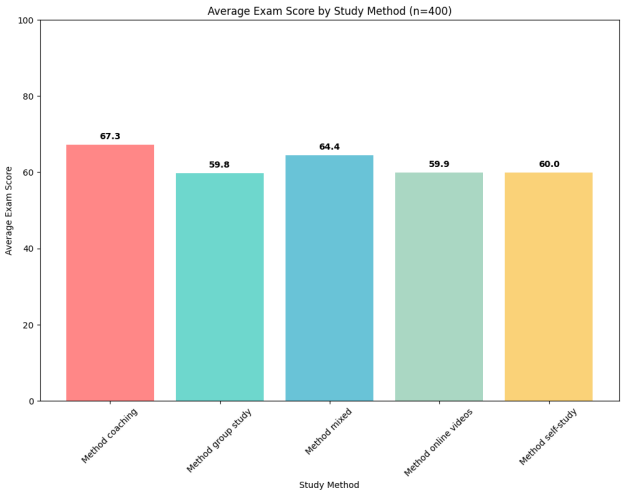


Gambar 2. *Scatter plot study_hours vs exam_score sebelum preprocessing.*

c. Study Method Performance

Perbedaan performa akademik dianalisis menggunakan bar chart yang menampilkan rata-rata exam_score untuk setiap kategori study_method. Visualisasi dibuat menggunakan subset 400 observasi untuk menjaga keterbacaan dan diuji konsistensinya dengan ukuran sampel lain.

Hasilnya menunjukkan bahwa coaching method memiliki rata-rata nilai tertinggi (67,3), mengindikasikan bahwa pendampingan terstruktur memberikan dampak positif pada performa belajar. Metode online videos dan self-study memperlihatkan performa menengah yang stabil, sedangkan group study cenderung menghasilkan nilai yang lebih rendah dibandingkan metode lainnya. Pola ini konsisten pada berbagai ukuran sampel, sehingga perbedaan performa antar metode dapat dianggap robust.

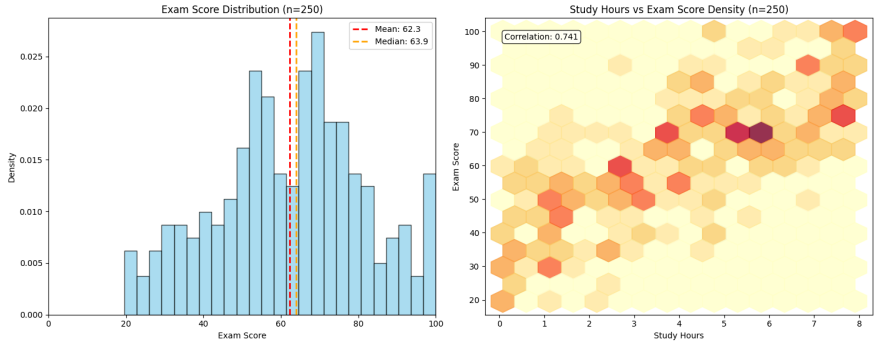


Gambar 3. Rata-rata nilai ujian berdasarkan metode belajar.

d. Distribution Evolution

Untuk memahami pola awal performa akademik, dilakukan visualisasi distribusi exam_score menggunakan histogram serta analisis hubungan study_hours-exam_score menggunakan hexbin plot. Histogram dilengkapi garis mean dan median untuk menilai kecenderungan pusat dan bentuk distribusi, sementara hexbin digunakan untuk menampilkan kepadatan observasi pada dataset besar dan memperlihatkan pola hubungan linear secara lebih jelas.

Hasil visualisasi menunjukkan bahwa distribusi nilai ujian relatif mendekati normal, dengan nilai mean 62,3 dan median 63,9, sehingga tidak terdapat bias ekstrem ke nilai sangat tinggi ataupun sangat rendah. Hexbin plot menampilkan hubungan positif yang kuat antara jam belajar dan nilai ujian, dengan korelasi sebesar 0,741. Area dengan kepadatan warna lebih gelap pada rentang study_hours 4–7 jam dan exam_score 60–80 mengindikasikan kelompok siswa dengan performa optimal. Konsistensi pola ini menunjukkan bahwa hubungan antara jam belajar dan performa ujian cukup stabil dan dapat dijadikan dasar analisis selanjutnya.



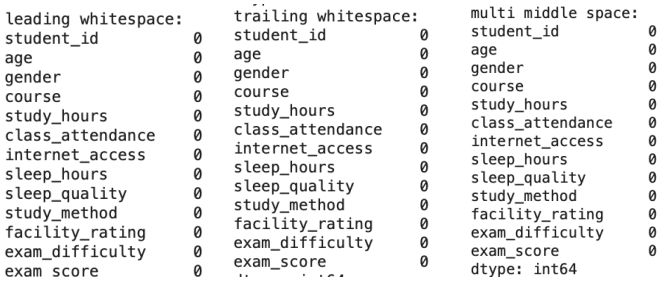
Gambar 4. *Distribusi exam_score dan hubungan awal study_hours-exam_score sebelum preprocessing.*

3. Data Pre-Processing

a. Data Cleaning

1) Cek White Space

Tahap pertama adalah dilakukan pemeriksaan terhadap white space di setiap cell data. Pemeriksaan tersebut dilakukan untuk tiga posisi yaitu di awal teks, akhir teks, dan multi spasi antar kata. Diperoleh hasil bahwa dataset tidak ada mengandung white space.



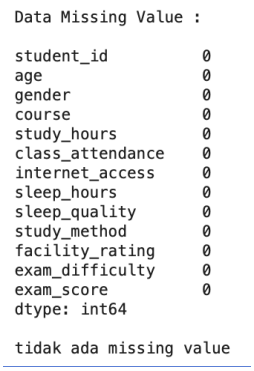
Gambar 5. *Output cek white space.*

2) Cek Missing Value

Cek missing value atau nilai kosong dari setiap cell data. Jika ditemukan missing value, maka penanganan yang dilakukan dengan teknik imputasi sbb :

- Missing value pada kolom numerik diisi dengan nilai dari kolom tersebut
- Missing value pada kolom kategorikal diisi dengan mode dari kolom tersebut.

Diperoleh hasil tidak ada missing value pada dataset.



Gambar *6.Output cek missing value*

3) Cek Data Duplicate

Memeriksa baris duplikat dalam data frame dengan penanganan yaitu dengan menghapus record yang terdeteksi sebagai data duplicated. Dataset tidak mengandung data duplicated

```
Jumlah Duplicates: 0
Tidak ada data duplikat, Tidak dilakukan penghapusan.
```

Gambar 7. Output cek data duplicate

4) Cek Outlier

Mendeteksi data outlier menggunakan teknik *IQR (Interquartile Range).* Data outlier jika melewati batas lower dan upper. Nilai lower dan upper diperoleh dari kuartil bawah dikurangi dengan 1.5 dan kuartil atas ditambah 1.5 kemudian masing-masingnya dikali dengan selisih nilai kuartil atas dan kuartil bawah. Penanganan terhadap data outlier tersebut dengan cara menghapus data-data yang nilai di luar batas tersebut sehingga hanya data yang berada dalam rentang normal yang dipertahankan. Berdasarkan teknik IQR menunjukan bahwa tidak ada data outlier atau noise pada.

```
Cek Data outlier :

Kolom: student_id
  Outlier di bawah lower : 0
  Outlier di atas upper  : 0

Kolom: age
  Outlier di bawah lower : 0
  Outlier di atas upper  : 0

Kolom: study_hours
  Outlier di bawah lower : 0
  Outlier di atas upper  : 0

Kolom: class_attendance
  Outlier di bawah lower : 0
  Outlier di atas upper  : 0

Kolom: sleep_hours
  Outlier di bawah lower : 0
  Outlier di atas upper  : 0

Kolom: exam_score
  Outlier di bawah lower : 0
  Outlier di atas upper  : 0
```

Gambar 8. *Output cek outlier*

b. Data Validation

Melakukan validasi nilai berdasarkan aturan logis. Validasi data dilakukan terhadap data pada fitur gender dengan nilai unik sbb:

- 6695 records : Male → valid
- 6579 records : Female → valid
- 6727 records : Other → invalid (remove)

Penanganan terhadap data invalid pada fitur gender yaitu dengan baris record data tersebut, sehingga ukuran data yang baru valid adalah 13.274 baris record.

c. Feature Creation

Pembuatan fitur baru dilakukan untuk meningkatkan kualitas data dan performa model machine learning, selain itu juga untuk mengungkap hubungan yang

tidak terlihat atau belum tercatat secara eksplisit. Fitur baru yang dibuat yaitu study_efficiency = exam_score / study_hours untuk menangkap hubungan seberapa efektif waktu belajar seseorang dalam menghasilkan nilai ujian dan sleep_study_ratio = sleep_hours / study_hours untuk menggambarkan keseimbangan antara waktu tidur dan waktu belajar.

d. Feature Encoding

Encoding dilakukan untuk mengubah value dari kolom yang bertipe object (nominal ataupun ordinal) menjadi nilai numerik agar dapat diproses oleh model. Teknik encoding data yang digunakan adalah label encoder. Encoding dilakukan pada atribut kategory yaitu gender, course, internet_access, sleep_quality, study_method, facility_rating, dan exam_difficulty. Pertama, pilih kolom yang bertipe objek. Data pada atribut tersebut dikonversi menjadi numerik (integer).

	student_id	age	gender	course	study_hours	class_attendance	internet_access	sleep_hours	sleep_quality	study_method	facility_rating	exam_difficulty	exam_score
0	1	17	1	6	2.78	92.9	1	7.4	2	0	1	1	58.9
2	3	22	1	1	7.88	76.8	1	8.5	2	0	0	2	90.3
4	5	20	0	6	0.89	71.6	1	9.8	2	0	1	2	43.7
5	6	23	1	2	3.48	65.4	1	4.2	1	2	1	2	58.2
6	7	17	0	2	1.35	69.0	1	7.4	0	3	0	1	53.7

Gambar 9. *Encoding data*

e. Perbandingan Data

Data sebelum preprocessing :

	student_id	age	gender	course	study_hours	class_attendance	internet_access	sleep_hours	sleep_quality	study_method	facility_rating	exam_difficulty	exam_score
0	1	17	male	diploma	2.78	92.9	yes	7.4	poor	coaching	low	hard	58.9
1	2	23	other	bca	3.37	64.8	yes	4.6	average	online videos	medium	moderate	54.8
2	3	22	male	b.sc	7.88	76.8	yes	8.5	poor	coaching	high	moderate	90.3
3	4	20	other	diploma	0.67	48.4	yes	5.8	average	online videos	low	moderate	29.7
4	5	20	female	diploma	0.89	71.6	yes	9.8	poor	coaching	low	moderate	43.7

Gambar 10. *Dataframe sebelum preprocessing*

Data setelah preprocessing :

	student_id	age	gender	course	study_hours	class_attendance	internet_access	sleep_hours	sleep_quality	study_method	facility_rating	exam_difficulty	exam_score	study_efficiency	sleep_study_ratio
0	1	17	1	6	2.78	92.9	1	7.4	2	0	1	1	58.9	21.187050	2.661871
2	3	22	1	1	7.88	76.8	1	8.5	2	0	0	2	90.3	11.459391	1.078680
4	5	20	0	6	0.89	71.6	1	9.8	2	0	1	2	43.7	49.101124	11.011236
5	6	23	1	2	3.48	65.4	1	4.2	1	2	1	2	58.2	16.724138	1.206897
6	7	17	0	2	1.35	69.0	1	7.4	0	3	0	1	53.7	39.777778	5.481481

Gambar 11. *Dataframe setelah preprocessing*

4. Exploratory Data Analysis (Sesudah Pre-Processing)

a. Data Overview

Setelah proses preprocessing, dataset menyisakan 13.274 observasi dengan 15 fitur numerik, karena seluruh variabel kategorikal telah melalui proses encoding. Statistik deskriptif memperlihatkan bahwa distribusi fitur utama: study_hours,

class_attendance, sleep_hours, dan exam_score, menjadi lebih stabil setelah pembersihan outlier. Nilai rata-rata menunjukkan pola perilaku belajar yang wajar, yaitu jam belajar sekitar 4 jam per hari, kehadiran kelas 70%, durasi tidur 7 jam, serta nilai ujian rata-rata sekitar 62.

Dua fitur rekayasa (study_efficiency dan sleep_study_ratio) memiliki rentang nilai yang lebih luas, namun tetap informatif dalam menggambarkan efektivitas dan keseimbangan belajar siswa. Secara keseluruhan, dataset pasca-preprocessing berada dalam kondisi bersih dan konsisten, sehingga layak digunakan pada tahap pemodelan prediktif.

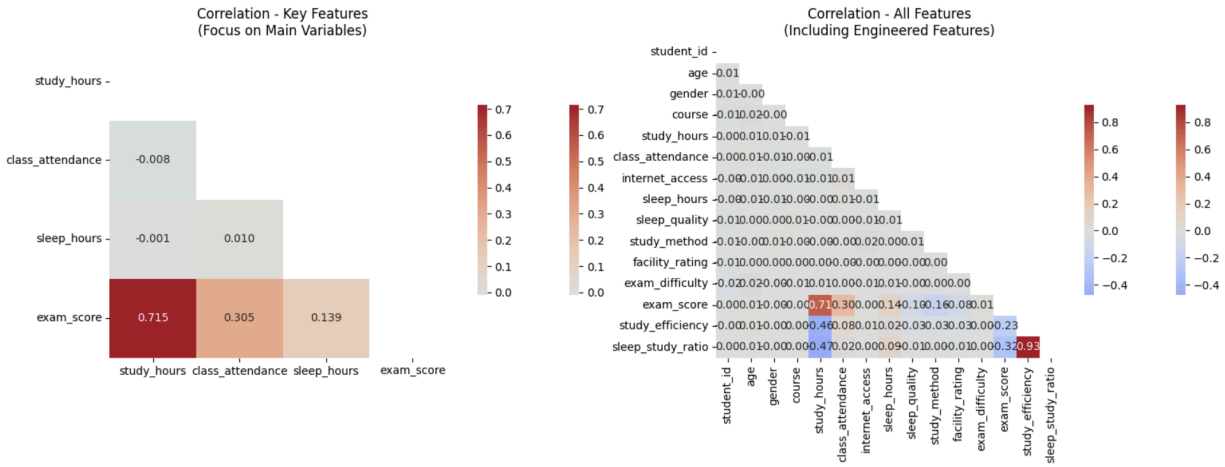
```
Dataset setelah preprocessing:
Shape: (13274, 15)
Total fitur: 15
Fitur numerik: 15
Fitur kategorikal (telah di-encode): 0
```

```
Statistik deskriptif fitur utama:
       study_hours  class_attendance  sleep_hours  exam_score  \
count     13274.00          13274.00     13274.00    13274.00
mean          4.01             70.00         7.01       62.49
std           2.30             17.34         1.73       18.87
min           0.08             40.60         4.10       19.60
25%           2.01             55.00         5.50       48.90
50%           4.05             69.90         7.00       62.60
75%           5.99             85.00         8.50       76.10
max           7.91             99.40         9.90      100.00

       study_efficiency  sleep_study_ratio
count          13274.00           13274.00
mean              33.28               4.86
std               67.63              11.46
min                4.28               0.52
25%               12.15               1.16
50%               15.57               1.73
75%               25.04               3.50
max             1002.50             123.75
```

b. Correlation Matrix Comparison

Analisis korelasi setelah preprocessing menunjukkan bahwa study_hours merupakan prediktor paling kuat terhadap exam_score ($r \approx 0{,}71$), diikuti oleh class_attendance dan sleep_hours dengan korelasi moderat. Pembersihan outlier membuat struktur korelasi lebih stabil dan mudah diinterpretasikan. Dua fitur rekayasa, study_efficiency dan sleep_study_ratio, memperlihatkan pola korelasi baru yang cenderung negatif terhadap nilai ujian, sehingga memberikan perspektif tambahan mengenai keseimbangan belajar dan tidur. Hasil ini menjadi dasar pemilihan fitur pada tahap pemodelan prediktif.

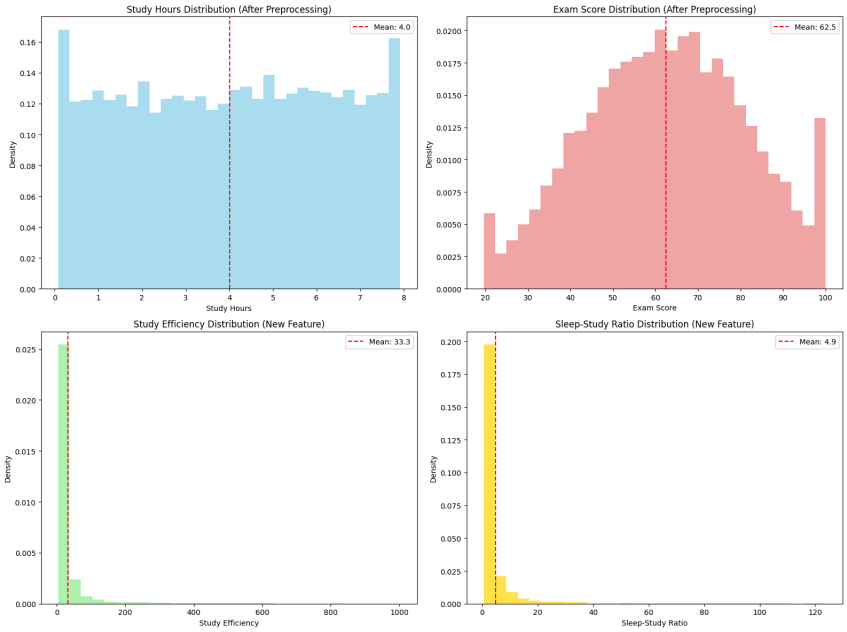


Gambar 12. Matriks korelasi fitur numerik setelah preprocessing.

c. Feature Distribution Analysis

Distribusi empat fitur utama yaitu study_hours, exam_score, study_efficiency, dan sleep_study_ratio, ditampilkan menggunakan histogram setelah proses preprocessing dan penghapusan outlier. Visualisasi ini berfungsi untuk memastikan bahwa persebaran data berada dalam kondisi stabil sebelum memasuki tahap pemodelan.

Setelah pembersihan data, study_hours menunjukkan distribusi yang lebih seragam dan bebas nilai ekstrem, sementara exam_score tampak lebih terpusat dibandingkan kondisi awal. Dua fitur hasil rekayasa, yaitu study_efficiency dan sleep_study_ratio, menunjukkan distribusi yang ter-skew namun tetap memberikan informasi penting terkait efektivitas belajar dan keseimbangan tidur–belajar siswa. Secara keseluruhan, visualisasi ini menegaskan bahwa dataset telah berada dalam kondisi yang layak dan representatif untuk digunakan pada proses modeling.



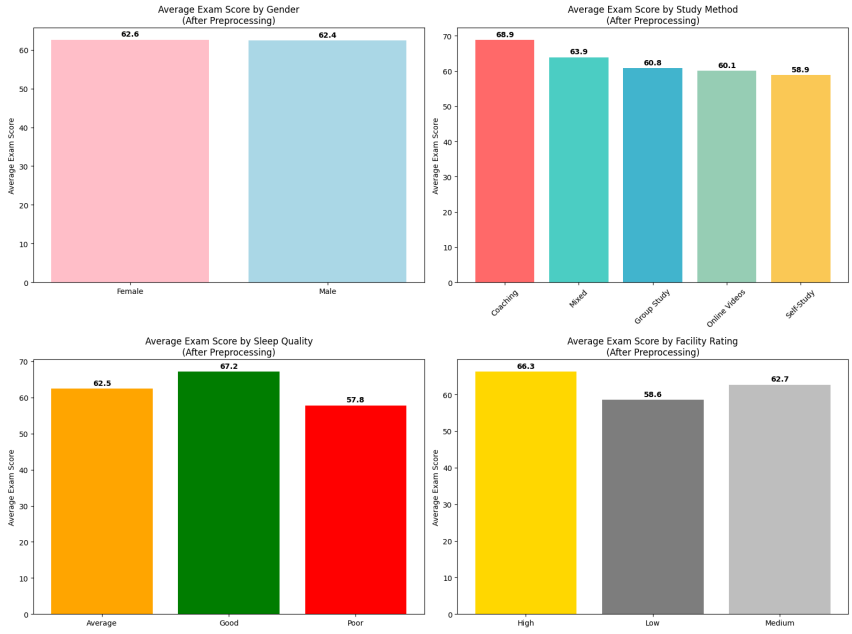
Gambar 13. *Distribusi empat fitur utama setelah preprocessing*

d. Categorical Analysis After Pre-Processing

Analisis terhadap variabel kategorikal dilakukan dengan membandingkan rata-rata exam_score pada setiap kelompok kategori, yang kemudian divisualisasikan menggunakan bar chart. Pendekatan ini memberikan gambaran mengenai perbedaan performa antar kelompok setelah dataset melalui proses preprocessing dan penghapusan outlier.

Hasil analisis menunjukkan bahwa gender tidak memberikan perbedaan performa yang signifikan, di mana nilai rata-rata siswa laki-laki dan perempuan hampir sama. Sebaliknya, study_method memperlihatkan perbedaan paling jelas, dengan metode coaching kembali menjadi kategori dengan nilai tertinggi secara konsisten. Untuk sleep_quality, siswa dengan kualitas tidur good memperoleh nilai tertinggi, diikuti kategori average dan poor. Sementara itu, facility_rating menunjukkan tren bahwa lingkungan belajar dengan fasilitas high berkorelasi terhadap performa akademik yang lebih baik dibandingkan kategori medium maupun low.

Secara keseluruhan, hasil ini mengindikasikan bahwa metode belajar, kualitas tidur, dan kualitas fasilitas memiliki pengaruh nyata terhadap variasi nilai ujian, sedangkan pengaruh gender bersifat minimal.

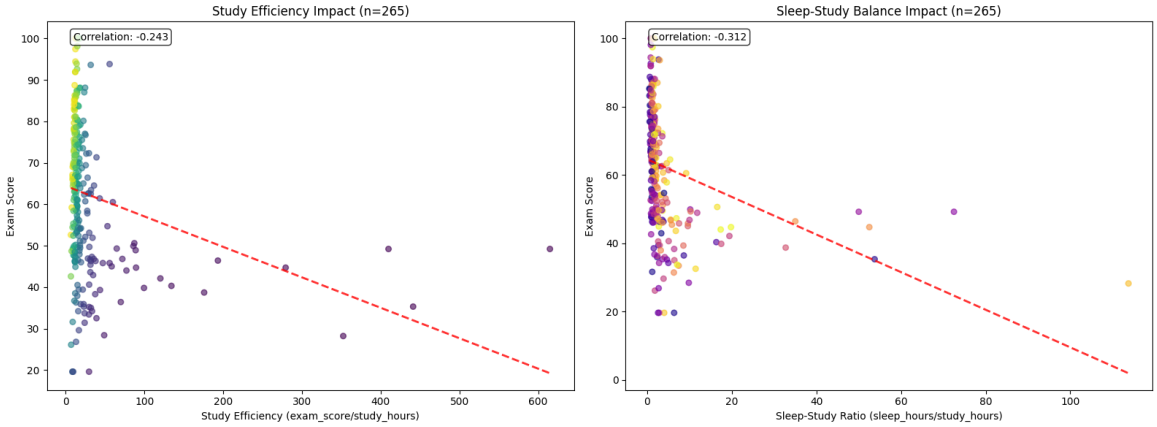


Gambar 14. ***Rata-rata nilai ujian berdasarkan variabel kategorikal setelah preprocessing.***

e. Feature Engineering Impact Analysis

Dua fitur turunan, yaitu study_efficiency dan sleep_study_ratio, ditambahkan untuk menangkap aspek perilaku belajar yang tidak terlihat pada fitur asli. Scatter plot pada Gambar X menunjukkan bahwa keduanya memiliki korelasi negatif moderat terhadap exam_score. Artinya, nilai ujian cenderung menurun ketika rasio efektivitas belajar atau rasio tidur terhadap belajar berada pada nilai ekstrem—suatu pola yang tidak terdeteksi melalui variabel study_hours atau sleep_hours secara langsung.

Temuan ini mengindikasikan bahwa fitur hasil rekayasa mampu memperkaya struktur informasi dalam dataset, terutama dalam merepresentasikan dinamika keseimbangan aktivitas siswa. Dengan demikian, penambahan fitur ini berpotensi meningkatkan performa model prediktif karena menyediakan sinyal tambahan terkait perilaku belajar yang lebih kompleks.



Gambar 15. *Dampak fitur turunan terhadap nilai ujian setelah preprocessing.*

5. Model dan Evaluasi

a. Feature Selection

Proses seleksi fitur dilakukan menggunakan analisis **Matriks Korelasi Pearson** untuk mengidentifikasi variabel yang memiliki hubungan linear terkuat dengan variabel target (exam_score).

1) **Metodologi**
   Fitur dipilih berdasarkan ambang batas (*threshold*) nilai korelasi absolut > 0.1. Variabel student_id dibuang karena tidak relevan secara statistik.
2) **Hasil Seleksi**
   Dari total dataset, terpilih **7 fitur utama**:
   a) study_hours (Korelasi: 0.715): Prediktor terkuat.
   b) sleep_study_ratio (Korelasi: 0.317).
   c) class_attendance (Korelasi: 0.305).
   d) study_efficiency (Korelasi: 0.226).
   e) study_method (Korelasi: 0.157).
   f) sleep_hours (Korelasi: 0.139).
   g) sleep_quality (Korelasi: 0.101).

b. Train Test Split

Data dibagi menjadi himpunan latih (*training set*) dan himpunan uji (*test set*) untuk mengevaluasi generalisasi model pada data yang belum pernah dilihat.

1) Rasio Pembagian
80% Training (10.619 sampel) dan 20% Testing (2.655 sampel).

2) Preprocessing
Dilakukan **Standard Scaling** (StandardScaler) pada fitur independen. Langkah ini menstandarisasi skala data (mean=0, std=1) untuk mencegah bias pada algoritma yang sensitif terhadap magnitudo angka (seperti Linear Regression dan Ridge), serta mempercepat konvergensi model.

c. Model Setup and Training

Eksperimen dilakukan menggunakan empat algoritma regresi dengan karakteristik berbeda untuk mencari performa terbaik:

1) Linear Regression
Sebagai *baseline* model linear sederhana.

2) Ridge Regression
Model linear dengan regularisasi L2 untuk mencegah *overfitting*.

3) **Random Forest Regressor**
   Model *ensemble* berbasis *bagging* (100 *estimators*).
4) **Gradient Boosting Regressor**
   Model *ensemble* berbasis *boosting* (100 *estimators*).

Setiap model dilatih menggunakan data yang telah disakalakan, dan divalidasi menggunakan **5-fold Cross-Validation** untuk memastikan konsistensi performa (rata-rata skor R2 dan standar deviasi).

d. Model Evaluation and Selection

Evaluasi dilakukan berdasarkan metrik R2 (koefisien determinasi), RMSE (*Root Mean Square Error*), dan MAE (*Mean Absolute Error*).

1) **Perbandingan Performa:**
   a) **Linear & Ridge**
      Menunjukkan performa moderat dengan R2 ~0.689 dan error (RMSE) ~10.46. Model ini gagal menangkap pola non-linear yang kompleks.
   b) **Gradient Boosting**
      Performa sangat baik (R2 0.964), namun masih di bawah Random Forest.
   c) **Random Forest:** Menunjukkan performa superior dan nyaris sempurna.
2) **Pemilihan Model Terbaik: Random Forest** dipilih sebagai model final.
   a) **R2 Score: 0.996** (Mampu menjelaskan 99.6% varians data).
   b) **RMSE: 1.16** (Tingkat kesalahan prediksi rata-rata sangat rendah).
   c) **Stabilitas:** CV R2 0.995 +/- 0.001 menunjukkan model sangat stabil dan *robust*.

e. Model Justification and Application

Pemilihan Random Forest didasarkan pada akurasi tinggi dan kemampuan algoritma menangkap hubungan non-linear antar variabel.

1) **Analisis Feature Importance (Random Forest):** Model sangat didominasi oleh dua fitur utama:
   a) **study_hours (0.671)**
      Faktor paling kritis penentu nilai.
   b) **study_efficiency (0.301)**
      Efisiensi belajar memiliki dampak signifikan kedua (fitur hasil *engineering*).
   c) Fitur lain seperti class_attendance dan sleep_quality memiliki kontribusi marginal (< 0.03).
2) **Potensi Implementasi:**

a) **Sistem Peringatan Dini**
   Mengidentifikasi siswa yang berisiko gagal berdasarkan jam belajar dan efisiensi rendah.
b) **Rekomendasi Personal**
   Menyarankan durasi belajar optimal (4–6 jam) bagi siswa yang ingin meningkatkan nilai.
c) **Optimasi Kurikulum**
   Menggeser fokus dari sekadar kehadiran fisik (class_attendance rendah dampaknya) ke metode pengajaran yang meningkatkan efisiensi belajar.

f. Data Insights

Berdasarkan analisis dataset terhadap 13.274 siswa, ditemukan *insight* kunci sebagai berikut:

1) **Kualitas Prediksi**
   Model sangat reliabel dengan *error rate* (MAE) hanya 0.51 poin. Artinya, jika prediksi nilai ujian adalah 80, nilai aslinya kemungkinan besar berada di antara 79.5 hingga 80.5.
2) **Kunci Keberhasilan Siswa:**
   a) **Kuantitas**
      Durasi belajar adalah indikator utama.
   b) **Kualitas**
      study_efficiency membuktikan bahwa *cara* belajar sama pentingnya dengan *lama* belajar.
3) **Rekomendasi Aksi:**
   a) Siswa disarankan menargetkan minimal 4–6 jam belajar mandiri.
   b) Metode *coaching* terbukti memberikan hasil efisiensi terbaik dibandingkan metode pasif.

   c) Keseimbangan tidur tetap diperlukan untuk menjaga rasio sleep_study yang optimal, meskipun dampak langsung jam tidur (sleep_hours) relatif kecil dibandingkan jam belajar.

Tabel Kontribusi

No	Nama	NIM	Tugas Utama
1	Bryan P Hutagalung	18222130	- Menentukan dataset bersama-sama - Mengerjakan Notebook Modeling and Evaluation - Mengerjakan Laporan dan Slide Modeling and Evaluation - Mengoordinasi pembuatan video - Membuat template Slide
2	Andang Kurniawan	23524061	- Menentukan dataset bersama-sama - Mengerjakan Notebook EDA sebelum dan sesudah Pre-Processing, sekaligus Visualisasi Data - Mengerjakan Laporan dan Slide EDA sebelum dan sesudah Pre-Processing sekaligus Visualisasi Data - Berkontribusi pembuatan video
3	Silvia Rahma	23525021	- Mengoordinasi penentuan dataset - Mengerjakan Notebook Data Pre-Processing - Mengerjakan Laporan dan Slide Data Pre-Processing - Berkontribusi pembuatan video - Membuat template Laporan

Link Terkait

Link Github: https://github.com/nathangalung/IF5100_TB_G07
Link Drive Video + Laporan: IF5100_TB_G07