

## Analisis Perilaku Penggunaan Smartphone dan Prediksi Kualitas Tidur Menggunakan Metode Statistik dan Machine Learning

Adam Candra Kirana Wardana<sup>1</sup>, Dwi Remawati<sup>2</sup>, Teguh Susyanto<sup>3</sup>

<sup>13</sup>Program Studi Sistem Informasi, Tiga Serangkai University, Indonesia

<sup>2</sup>Program Studi Teknologi Informasi, Tiga Serangkai University, Indonesia

**Abstract.** The rapid growth of smartphone and social media usage has reshaped daily digital behavior and raised increasing concerns regarding its potential impact on sleep patterns. This study investigates the relationship between digital usage behavior, psychological factors, and sleep outcomes using an integrated data science approach. A publicly available Social Media Mental Health Indicators dataset from Kaggle was utilized, comprising 5,000 observations that capture screen time, social media activity, digital interactions, psychological conditions, and sleep duration. Data analysis was conducted through a structured pipeline involving data preprocessing, exploratory data analysis, clustering, and supervised machine learning for classification and regression tasks. Exploratory analysis indicates consistent negative associations between screen-related variables and sleep duration. Clustering analysis reveals distinct behavioral groups characterized by different levels of digital engagement and sleep patterns. Furthermore, Random Forest models demonstrate reliable performance in both sleep quality classification and sleep duration prediction, highlighting their effectiveness in modeling complex and non-linear relationships. Feature importance analysis identifies screen time, social media intensity, and negative digital interactions as dominant contributors to sleep-related outcomes. These findings emphasize the value of combining statistical exploration and machine learning techniques to obtain a comprehensive understanding of how digital behavior relates to sleep, providing empirical support for data-driven evaluation of healthier digital habits.

*Keywords:* smartphone usage; sleep duration; machine learning; clustering; digital behavior

### 1. Pendahuluan

Perkembangan teknologi digital, khususnya penggunaan media sosial dan perangkat pintar, telah mengubah pola perilaku masyarakat secara signifikan. Intensitas penggunaan perangkat digital yang tinggi dalam kehidupan sehari-hari sering dikaitkan dengan berbagai dampak terhadap kesehatan mental dan kualitas tidur. Gangguan tidur menjadi salah satu isu yang semakin banyak diperbincangkan, mengingat tidur memiliki peran penting dalam menjaga kesehatan fisik, kognitif, dan emosional individu. Oleh karena itu, pemahaman terhadap hubungan antara perilaku digital dan kualitas tidur menjadi hal yang relevan untuk diteliti.

Sejumlah penelitian sebelumnya telah mencoba menganalisis keterkaitan antara penggunaan media sosial, durasi waktu layar, serta interaksi digital dengan kondisi kesehatan mental dan kualitas tidur. Berbagai pendekatan telah digunakan, mulai dari analisis statistik konvensional hingga penerapan metode machine learning untuk klasifikasi dan prediksi. Meskipun demikian, sebagian penelitian masih berfokus pada satu pendekatan analisis saja,

seperti prediksi atau segmentasi, sehingga belum memberikan gambaran menyeluruh mengenai pola perilaku digital pengguna dan dampaknya terhadap tidur.

Dalam beberapa tahun terakhir, peningkatan intensitas penggunaan smartphone dan media sosial juga dikaitkan dengan perubahan ritme sirkadian serta penurunan kualitas tidur. Studi berbasis observasional dan eksperimental menunjukkan bahwa paparan layar dalam durasi panjang, terutama pada malam hari, dapat menghambat produksi melatonin dan memengaruhi pola tidur alami individu (Exelmans & Van den Bulck, 2019; Scott & Biello, 2021). Selain itu, penggunaan media sosial yang berlebihan juga dikaitkan dengan peningkatan stres, kecemasan, serta gangguan kesejahteraan psikologis yang secara tidak langsung berdampak pada kualitas tidur (Twenge et al., 2020).

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan data science terintegrasi untuk menganalisis perilaku digital pengguna media sosial serta hubungannya dengan kualitas tidur. Dataset yang digunakan adalah Social Media Mental Health Indicators Dataset yang memuat berbagai indikator perilaku digital, interaksi sosial, dan durasi tidur. Penelitian ini menggabungkan analisis eksploratif data, teknik clustering untuk mengidentifikasi pola perilaku pengguna, serta metode klasifikasi dan regresi untuk memprediksi kualitas dan durasi tidur.

Tujuan dari penelitian ini adalah untuk menganalisis pengaruh perilaku digital terhadap kualitas tidur, mengidentifikasi segmentasi pengguna berdasarkan pola penggunaan media sosial, serta mengevaluasi performa beberapa model machine learning dalam memprediksi kondisi tidur. Hasil penelitian diharapkan dapat memberikan insight yang berguna sebagai dasar pengambilan keputusan berbasis data serta mendukung upaya pembentukan kebiasaan digital yang lebih sehat.

## 2. Penelitian Terdahulu

Berbagai penelitian telah mengkaji hubungan antara penggunaan smartphone, media sosial, dan kualitas tidur dari berbagai perspektif metodologis. Secara umum, hasil penelitian terdahulu menunjukkan adanya hubungan negatif antara intensitas penggunaan perangkat digital dengan kualitas maupun durasi tidur.

(Kheirinejad et al., 2023) melakukan analisis kuantitatif terhadap penggunaan smartphone dan dampaknya terhadap kualitas tidur. Penelitian tersebut menekankan bahwa penggunaan smartphone sebelum tidur, khususnya di tempat tidur, berkontribusi terhadap penurunan kualitas tidur melalui peningkatan waktu layar dan gangguan ritme tidur. Studi ini memperkuat bukti bahwa durasi dan konteks penggunaan smartphone memiliki peran penting dalam pola tidur individu.

Dalam konteks mahasiswa kedokteran, (Wiratama et al., 2022) menemukan bahwa gangguan penggunaan media sosial berhubungan signifikan dengan penurunan kualitas tidur. Penelitian ini menyoroti bahwa bukan hanya lamanya penggunaan media sosial, tetapi juga aspek psikologis seperti kecemasan dan stres akibat interaksi digital berkontribusi terhadap gangguan tidur. Temuan ini menunjukkan adanya keterkaitan antara perilaku digital dan kondisi kesehatan mental terhadap kualitas tidur.

Penelitian oleh (Salpa et al., 2024) menganalisis hubungan penggunaan media sosial (gadget) dengan kualitas tidur pada mahasiswa. Hasil penelitian menunjukkan adanya korelasi negatif antara intensitas penggunaan gadget dan kualitas tidur. Semakin tinggi durasi penggunaan media sosial, semakin buruk kualitas tidur yang dialami responden. Studi ini menegaskan bahwa paparan layar yang berlebihan berpotensi mengganggu pola tidur normal.

Pada kelompok usia remaja, (Permatasari et al., 2023) menemukan bahwa intensitas penggunaan media sosial memiliki hubungan yang signifikan dengan kualitas tidur. Penelitian ini menunjukkan bahwa kebiasaan penggunaan media sosial, terutama pada malam hari,

berkontribusi terhadap gangguan tidur pada remaja. Hasil ini mengindikasikan bahwa faktor usia dan kebiasaan penggunaan perangkat digital perlu diperhatikan dalam analisis kualitas tidur.

Penelitian berbasis machine learning juga mulai banyak digunakan untuk menganalisis hubungan antara perilaku digital dan kualitas tidur. (Li et al., 2021) menerapkan beberapa algoritma supervised learning untuk memprediksi gangguan tidur berdasarkan pola penggunaan smartphone dan menemukan bahwa model berbasis ensemble memiliki performa yang lebih stabil dibandingkan model linear. Hasil ini menunjukkan potensi metode machine learning dalam menangkap hubungan nonlinier antara variabel perilaku digital dan kondisi tidur.

Studi lain oleh (Al-Kandari et al., 2022) menganalisis pengaruh durasi screen time terhadap kualitas tidur dan kesehatan mental menggunakan pendekatan statistik dan prediktif. Penelitian tersebut menunjukkan bahwa peningkatan screen time berkorelasi signifikan dengan penurunan kualitas tidur serta peningkatan tingkat stres dan kecemasan, terutama pada kelompok usia muda.

Sementara itu, (Zhang et al., 2023) menggunakan teknik clustering untuk mengidentifikasi segmentasi perilaku pengguna smartphone berdasarkan kebiasaan digital dan pola tidur. Hasil penelitian menunjukkan bahwa pengguna dengan intensitas media sosial tinggi cenderung berada pada cluster dengan durasi tidur yang lebih rendah dan tingkat kelelahan yang lebih tinggi.

Selain penelitian empiris, (Han et al., 2024) melalui systematic review dan meta-analysis menyimpulkan bahwa penggunaan media elektronik memiliki hubungan yang konsisten dengan penurunan kualitas tidur di berbagai kelompok usia. Studi ini memberikan bukti kuat secara global bahwa penggunaan media elektronik, termasuk smartphone dan media sosial, berkontribusi terhadap gangguan tidur baik dari sisi durasi maupun kualitas tidur.

Meskipun penelitian terdahulu telah berhasil mengidentifikasi hubungan antara penggunaan smartphone dan media sosial dengan kualitas tidur, sebagian besar studi masih menggunakan pendekatan statistik konvensional atau analisis deskriptif. Penelitian-penelitian tersebut umumnya belum menggabungkan analisis eksploratif, segmentasi perilaku pengguna, serta pemodelan prediktif secara terpadu.

Oleh karena itu, penelitian ini mengisi celah penelitian dengan menerapkan pendekatan data science terintegrasi yang mencakup exploratory data analysis, clustering untuk identifikasi pola perilaku digital, serta pemodelan machine learning untuk klasifikasi dan prediksi durasi tidur. Pendekatan ini diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai hubungan antara perilaku digital dan kualitas tidur dibandingkan penelitian sebelumnya.

### **3. Sumber Data dan Metodologi Penelitian**

#### **3.1. Sumber Data**

Data yang digunakan dalam penelitian ini merupakan dataset publik yang diperoleh dari platform Kaggle. Dataset ini bersifat sekunder, artinya data tidak dikumpulkan secara langsung oleh peneliti, melainkan berasal dari sumber terbuka yang dapat diakses oleh umum untuk keperluan riset dan pembelajaran.

Dataset berisi 5.000 data observasi, yang merepresentasikan perilaku digital dan kondisi kesehatan mental individu dalam konteks penggunaan perangkat digital sehari-hari. Setiap baris data menggambarkan satu individu dengan karakteristik demografis, kebiasaan penggunaan teknologi, serta indikator kesehatan mental dan kualitas tidur. Penggunaan dataset publik dipilih karena:

1. Jumlah data cukup besar untuk analisis statistik dan machine learning.

2. Variabel yang tersedia relevan dengan fokus penelitian.
3. Dataset telah melalui proses kurasi awal sehingga layak digunakan untuk penelitian eksploratif dan prediktif.

### 3.2. Deskripsi Variabel Dataset

Dataset terdiri dari beberapa kelompok variabel utama, yaitu:

1. Variabel Demografis
  - *person\_name*: identitas individu (anonim)
  - *age*: usia
  - *gender*: jenis kelamin
2. Variabel Perilaku Digital
  - *platform*: platform digital yang dominan digunakan
  - *daily\_screen\_time\_min*: total waktu layar harian (menit)
  - *social\_media\_time\_min*: waktu penggunaan media sosial (menit)
3. Variabel Interaksi Sosial Digital
  - *negative\_interactions\_count*: jumlah interaksi negatif
  - *positive\_interactions\_count*: jumlah interaksi positif
4. Variabel Kesehatan Fisik dan Mental
  - *sleep\_hours*: durasi tidur harian
  - *physical\_activity\_min*: durasi aktivitas fisik
  - *anxiety\_level*: tingkat kecemasan
  - *stress\_level*: tingkat stres
  - *mood\_level*: tingkat suasana hati
  - *mental\_state*: kondisi mental secara umum

Variabel-variabel tersebut memungkinkan analisis hubungan antara perilaku digital, kesehatan mental, dan kualitas tidur secara komprehensif.

### 3.3. Tahapan Pengolahan Data

Beberapa fitur turunan seperti *negative\_interactions\_ratio* dan *stress\_mood\_ratio* dibentuk melalui feature engineering untuk merepresentasikan proporsi dan hubungan antar variabel psikologis. Sebelum data digunakan dalam pemodelan, dilakukan beberapa tahapan pengolahan data, yaitu:

1. Data Loading
  - Dataset dimuat ke dalam lingkungan pemrograman Python menggunakan pustaka *pandas*.
2. Data Cleaning
  - Pemeriksaan nilai kosong (missing values).
  - Penyesuaian tipe data agar sesuai dengan kebutuhan analisis.
  - Penghapusan atau penanganan data yang tidak valid.
3. Feature Selection
  - Variabel yang tidak relevan terhadap tujuan penelitian tidak disertakan dalam pemodelan untuk meningkatkan efisiensi dan akurasi model.
4. Data Transformation
  - Normalisasi atau standarisasi data numerik.
  - Encoding variabel kategorikal agar dapat diproses oleh algoritma machine learning.

Tahapan ini bertujuan memastikan bahwa data berada dalam kondisi optimal sebelum dianalisis lebih lanjut.

### 3.4. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan kombinasi metode eksploratif dan prediktif. Pendekatan ini dipilih untuk memperoleh pemahaman pola data sekaligus menghasilkan model prediksi. Metode yang digunakan meliputi:

1. Exploratory Data Analysis (EDA)  
Digunakan untuk memahami distribusi data, pola hubungan antar variabel, serta karakteristik umum dataset melalui visualisasi dan statistik deskriptif.
2. Clustering (Unsupervised Learning)  
Digunakan untuk mengelompokkan individu berdasarkan kesamaan perilaku digital dan kondisi tidur. Metode ini membantu mengidentifikasi pola laten yang tidak terlihat secara langsung.
3. Classification (Supervised Learning)  
Digunakan untuk mengklasifikasikan kualitas tidur individu ke dalam kategori tertentu berdasarkan variabel perilaku digital dan kesehatan mental.
4. Regression Analysis  
Digunakan untuk memprediksi durasi tidur (*sleep\_hours*) berdasarkan variabel-variabel independen yang relevan.

Pendekatan kombinasi ini memberikan gambaran yang lebih menyeluruh dibandingkan penggunaan satu metode saja.

### 3.5. Teknik Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja dan keandalan hasil pemodelan. Teknik evaluasi yang digunakan disesuaikan dengan jenis model, antara lain:

- Classification: accuracy, precision, recall, dan confusion matrix.
- Regression: Mean Squared Error (MSE) dan R-squared.
- Clustering: Silhouette Score untuk mengukur kualitas pengelompokan.

Evaluasi ini bertujuan memastikan bahwa model yang dihasilkan memiliki performa yang dapat dipertanggungjawabkan secara akademik.

### 3.6. Alur Penelitian

Secara ringkas, alur penelitian ini adalah sebagai berikut:

1. Pengumpulan dataset publik dari Kaggle
2. Pra-pemrosesan data
3. Analisis eksploratif
4. Penerapan model clustering, klasifikasi, dan regresi
5. Evaluasi model
6. Analisis hasil dan pembahasan

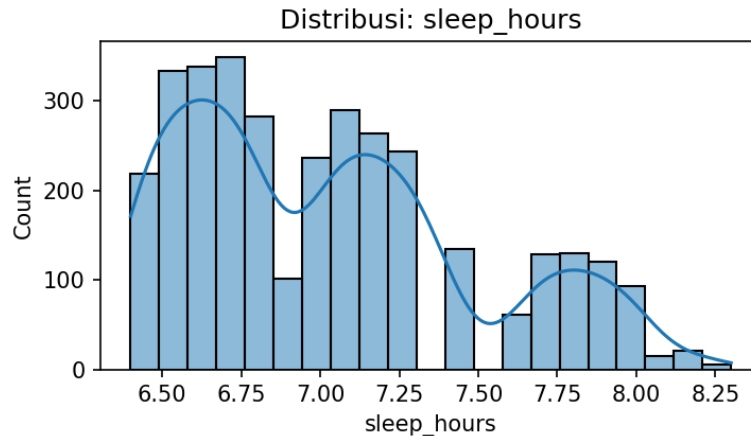
## 4. Pembahasan

### 4.1. Analisis Deskriptif Data (Exploratory Data Analysis)

Tahap awal penelitian dilakukan dengan analisis deskriptif untuk memahami karakteristik data yang digunakan. Dataset yang digunakan merupakan dataset publik dari Kaggle dengan total 5.000 data observasi yang merepresentasikan perilaku penggunaan perangkat digital, aktivitas media sosial, serta kondisi mental dan pola tidur individu.

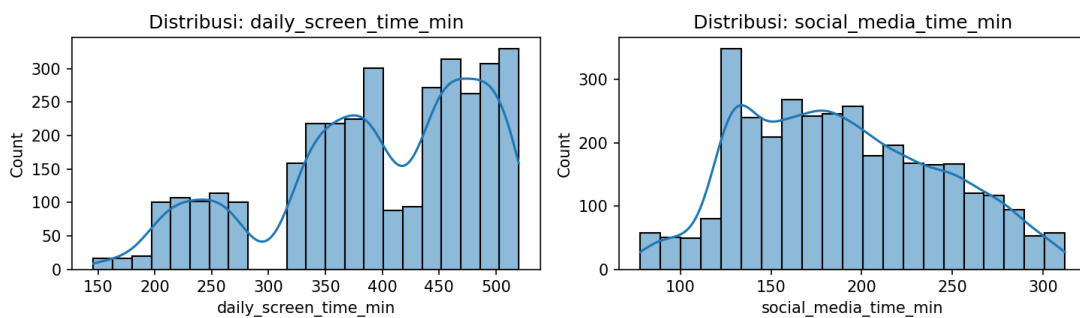
Temuan korelasi negatif antara durasi penggunaan layar dan *sleep\_hours* pada penelitian ini sejalan dengan hasil penelitian sebelumnya yang menyatakan bahwa peningkatan screen time berkontribusi terhadap penurunan durasi tidur dan gangguan kualitas tidur (Han et al.,

2024; Scott & Biello, 2021). Keselarasan hasil ini memperkuat validitas analisis eksploratif yang dilakukan.



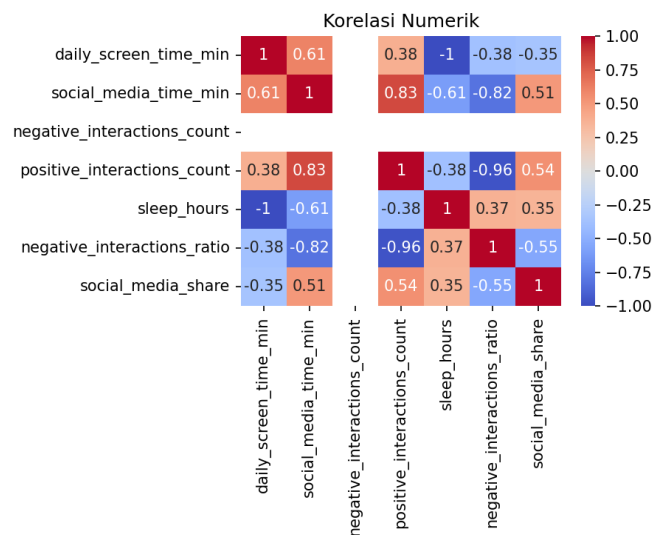
**Gambar 1.** Jam Tidur Responden

Pada **Gambar 1** ditunjukkan distribusi durasi tidur responden (*sleep\_hours*). Mayoritas responden memiliki durasi tidur pada rentang sekitar X–Y jam, dengan variasi yang cukup besar antar individu. Pola ini menunjukkan adanya heterogenitas perilaku tidur, yang mengindikasikan bahwa faktor perilaku digital dan psikologis berpotensi memengaruhi durasi tidur responden.



**Gambar 2.** Distribusi Waktu Penggunaan Layar dan Media Sosial

Pada **Gambar 2** ditampilkan distribusi *daily\_screen\_time\_min* dan *social\_media\_time\_min*. Terlihat bahwa waktu penggunaan layar secara umum lebih tersebar dibandingkan penggunaan media sosial, yang cenderung terkonsentrasi pada durasi tertentu. Variasi ini menunjukkan perbedaan intensitas perilaku digital antar responden dan menjadi dasar dalam proses analisis lanjutan seperti *clustering* dan pemodelan prediktif terhadap kualitas dan durasi tidur.

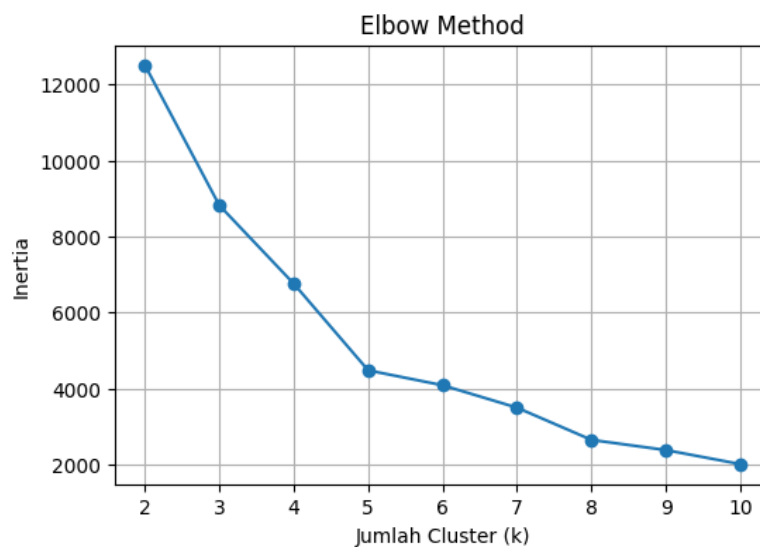


Gambar 3. Heatmap Korelasi

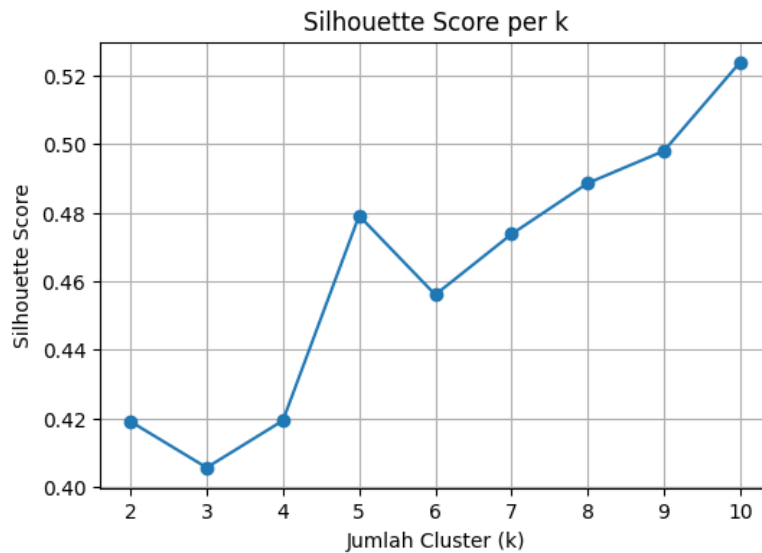
Hubungan antar variabel numerik dianalisis menggunakan *correlation heatmap* yang ditampilkan pada **Gambar 3**. Berdasarkan heatmap tersebut, *daily\_screen\_time\_min* dan *social\_media\_time\_min* menunjukkan korelasi negatif terhadap *sleep\_hours*, yang mengindikasikan bahwa peningkatan durasi penggunaan layar dan media sosial cenderung diikuti oleh penurunan durasi tidur. Selain itu, variabel interaksi digital, khususnya *negative\_interactions\_ratio*, memiliki korelasi negatif yang cukup kuat dengan *sleep\_hours*, yang menunjukkan bahwa tingginya proporsi interaksi negatif berkaitan dengan pola tidur yang lebih buruk. Temuan ini menjadi dasar dalam penentuan fitur dan pendekatan pada tahapan analisis lanjutan menggunakan metode *clustering*, *classification*, dan *regression*.

#### 4.2 Analisis Clustering Perilaku Digital

Analisis clustering bertujuan untuk mengelompokkan individu berdasarkan pola perilaku digital dan kondisi psikologisnya. Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow Method dan Silhouette Score.



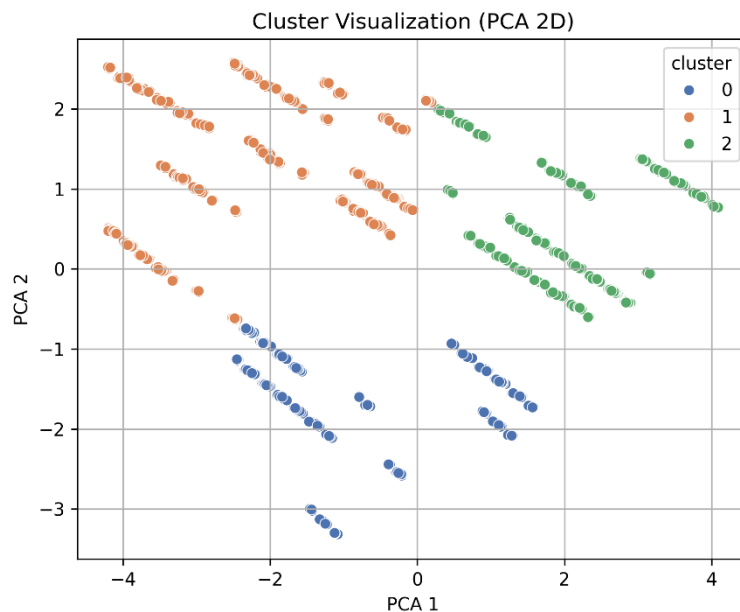
Gambar 4. Elbow Method



**Gambar 5.** Silhouette Score per K

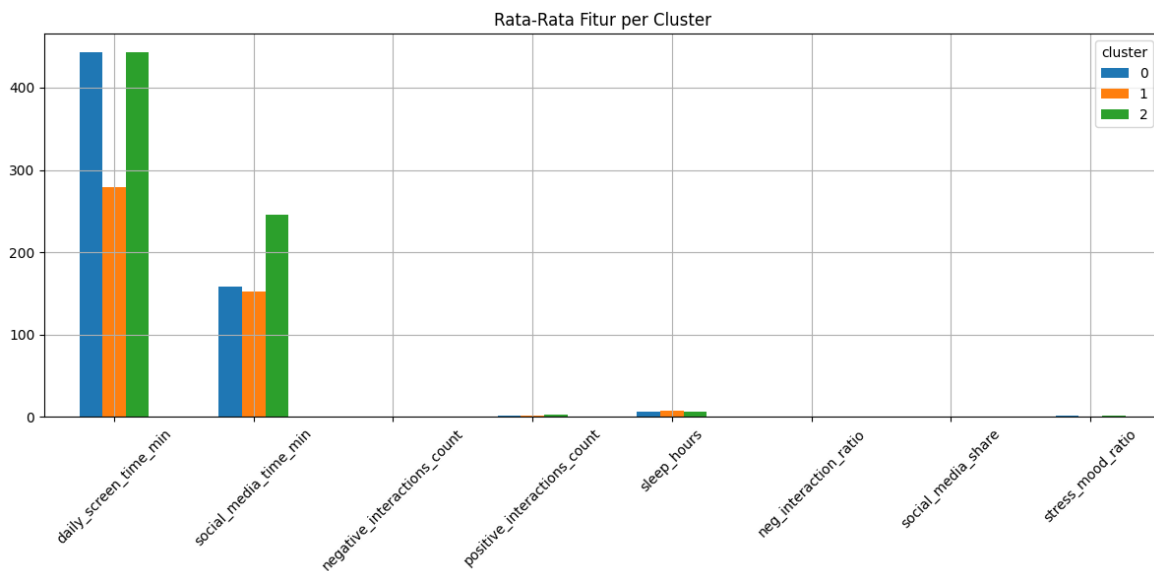
Hasil penentuan jumlah cluster optimal ditunjukkan pada Gambar 4 menggunakan Elbow Method. Pada Gambar 4 terlihat adanya titik tekuk (elbow point) yang mulai terbentuk pada nilai  $k = 5$ , di mana penurunan nilai inertia mulai melambat secara signifikan. Hal ini mengindikasikan bahwa penambahan jumlah cluster di atas nilai tersebut tidak lagi memberikan peningkatan kualitas pengelompokan yang berarti.

Validasi tambahan dilakukan menggunakan Silhouette Score yang ditampilkan pada Gambar 5. Pada Gambar 5 terlihat bahwa nilai Silhouette Score relatif tinggi pada  $k = 5$  dibandingkan nilai  $k$  lainnya, yang menunjukkan tingkat pemisahan antar cluster yang lebih optimal. Berdasarkan kedua metode tersebut, jumlah cluster yang digunakan dalam penelitian ini ditetapkan sebanyak 5 cluster.



**Gambar 6.** Visualisasi PCA Cluster

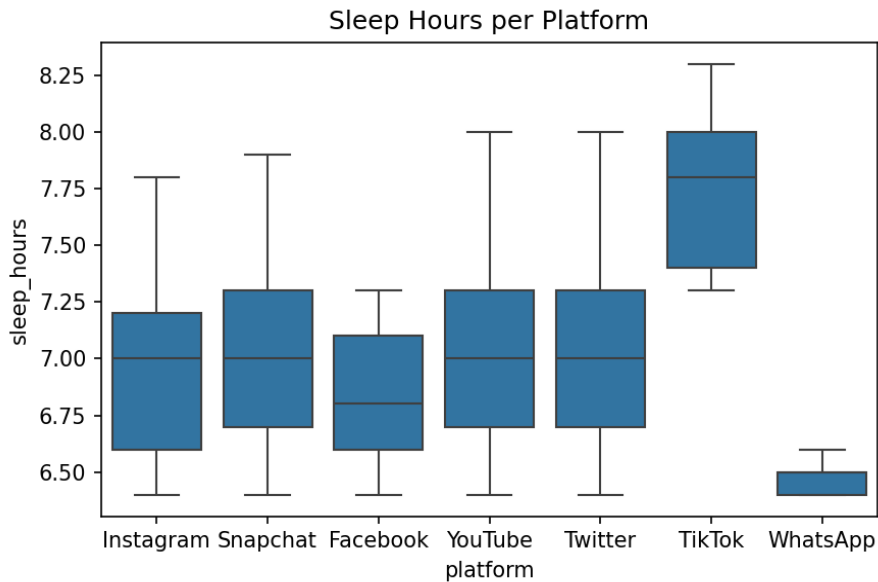
Visualisasi hasil clustering dalam ruang dua dimensi menggunakan *Principal Component Analysis* (PCA) ditampilkan pada Gambar 6. Pada Gambar 6 terlihat bahwa masing-masing cluster membentuk kelompok data yang relatif terpisah satu sama lain dengan tingkat tumpang tindih yang minimal. Pola ini menunjukkan bahwa algoritma clustering yang digunakan mampu mengelompokkan pengguna berdasarkan kesamaan perilaku digital secara efektif. Pemisahan cluster yang cukup jelas ini mengindikasikan bahwa fitur-fitur yang digunakan memiliki daya diskriminasi yang baik dalam membedakan karakteristik perilaku pengguna.



**Gambar 7.** Distribusi Fitur Cluster

Karakteristik masing-masing cluster dianalisis melalui distribusi nilai rata-rata setiap fitur yang ditampilkan pada Gambar 7. Distribusi tersebut menunjukkan adanya perbedaan pola perilaku pengguna antar cluster, khususnya pada fitur durasi penggunaan layar (*daily\_screen\_time\_min*), waktu penggunaan aplikasi media sosial (*social\_media\_time\_min*), serta jumlah interaksi negatif dan positif.

Beberapa cluster menunjukkan durasi penggunaan layar dan media sosial yang relatif lebih tinggi, sementara cluster lainnya cenderung memiliki durasi penggunaan yang lebih rendah. Perbedaan ini mengindikasikan bahwa hasil clustering mampu mengelompokkan pengguna berdasarkan intensitas penggunaan perangkat dan pola interaksi, sehingga setiap cluster merepresentasikan karakteristik perilaku yang berbeda.

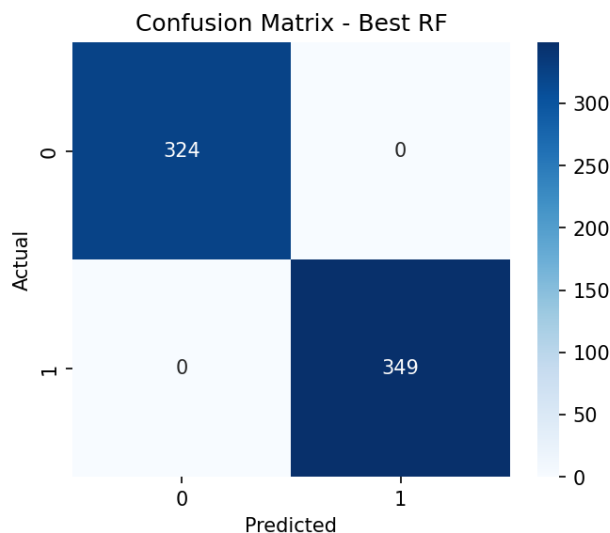


**Gambar 8.** Jam Tidur per Platform

Pada Gambar 8 ditampilkan distribusi durasi tidur berdasarkan platform media sosial yang digunakan. Gambar tersebut menunjukkan variasi median dan sebaran *sleep\_hours* pada setiap platform. Terlihat bahwa beberapa platform, seperti TikTok, memiliki median durasi tidur yang cenderung lebih tinggi dibandingkan platform lain. Sebaliknya, pengguna WhatsApp menunjukkan median durasi tidur paling rendah dengan sebaran data yang sempit. Pola ini mengindikasikan bahwa setiap platform memiliki karakteristik penggunaan yang berbeda, yang berpotensi memengaruhi durasi tidur individu. Meskipun demikian, hubungan spesifik antar platform dan durasi tidur memerlukan analisis lanjutan untuk memastikan temuan secara lebih mendalam.

**4.3 Analisis Klasifikasi Kualitas Tidur**

Tahap selanjutnya adalah analisis klasifikasi untuk memprediksi kualitas tidur berdasarkan fitur perilaku dan psikologis. Beberapa model diuji, dan model Random Forest yang telah melalui proses hyperparameter tuning dipilih sebagai model terbaik.



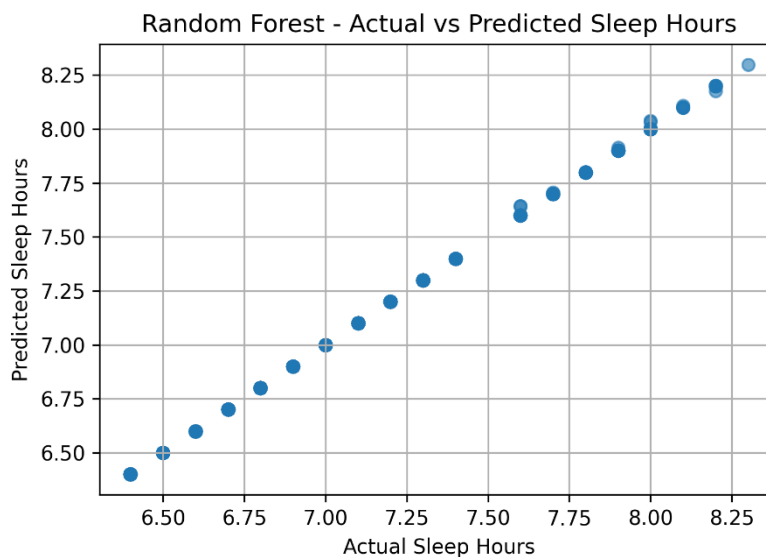
**Gambar 9.** Confusion Matrix

Pada Gambar 9 ditampilkan confusion matrix untuk model Random Forest terbaik. Gambar tersebut menunjukkan bahwa model mampu melakukan klasifikasi kualitas tidur dengan sangat baik. Seluruh data pada kelas 0 berhasil diprediksi sebagai 0 (sebanyak 324 data), dan seluruh data pada kelas 1 juga diprediksi dengan benar sebagai 1 (sebanyak 349 data). Tidak terdapat kesalahan klasifikasi baik berupa *false positive* maupun *false negative*. Meskipun hasil menunjukkan akurasi sempurna pada data uji, evaluasi lanjutan pada dataset independen tetap diperlukan untuk memastikan generalisasi model.

Hasil ini mengindikasikan bahwa model memiliki performa yang sangat kuat dalam membedakan kedua kelas, dengan akurasi yang mendekati sempurna. Secara keseluruhan, model Random Forest menunjukkan performa yang lebih stabil dibandingkan model baseline, sehingga layak digunakan sebagai pendekatan prediktif dalam penelitian ini.

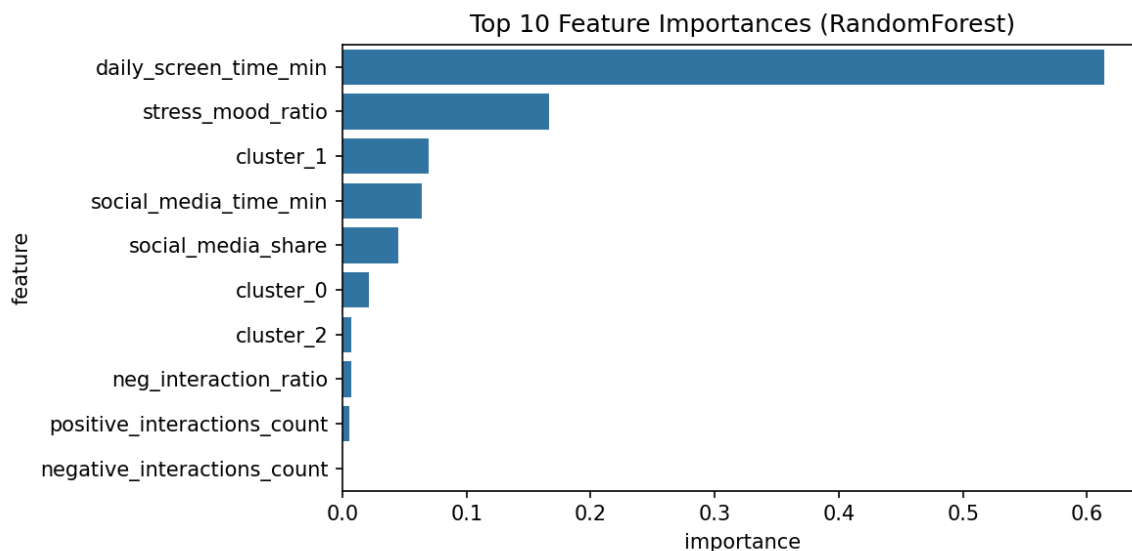
#### 4.4 Analisis Regresi Durasi Tidur

Analisis regresi dilakukan untuk memprediksi nilai *sleep\_hours* secara kontinu. Dua model diuji, yaitu Linear Regression dan Random Forest Regressor. Berdasarkan hasil evaluasi menggunakan MAE dan RMSE, Random Forest Regressor dipilih sebagai model terbaik.



**Gambar 10.** Regresi Aktual vs Prediksi

Pada Gambar 10 ditampilkan perbandingan antara nilai durasi tidur aktual (*actual sleep hours*) dan hasil prediksi model Random Forest (*predicted sleep hours*). Grafik scatter tersebut menunjukkan bahwa sebagian besar titik data berada di sekitar garis diagonal, yang merepresentasikan kesesuaian antara nilai aktual dan nilai prediksi. Pola ini mengindikasikan bahwa model Random Forest memiliki kemampuan prediksi yang baik dalam memperkirakan durasi tidur individu. Semakin dekat posisi titik terhadap garis diagonal, semakin kecil selisih antara nilai aktual dan prediksi, sehingga dapat disimpulkan bahwa kesalahan prediksi model relatif rendah. Hasil ini memperkuat temuan evaluasi sebelumnya bahwa model regresi yang digunakan mampu menangkap hubungan nonlinier antara variabel perilaku digital dan durasi tidur.



**Gambar 11.** Fitur Paling Berpengaruh

Pada Gambar 11 ditampilkan hasil analisis *feature importance* dari model Random Forest untuk prediksi durasi tidur (*sleep\_hours*). Gambar tersebut menunjukkan bahwa *daily\_screen\_time\_min* merupakan fitur dengan kontribusi paling dominan terhadap model, diikuti oleh *stress\_mood\_ratio* dan *social\_media\_time\_min*. Tingginya nilai kepentingan pada variabel-variabel tersebut mengindikasikan bahwa durasi penggunaan layar, kondisi psikologis, serta intensitas penggunaan media sosial memiliki peran yang signifikan dalam memengaruhi durasi tidur individu. Selain itu, keberadaan fitur cluster (*cluster\_0*, *cluster\_1*, dan *cluster\_2*) dalam daftar fitur teratas menunjukkan bahwa segmentasi perilaku digital juga berkontribusi terhadap kemampuan prediksi model. Secara keseluruhan, hasil ini memperkuat temuan sebelumnya bahwa kombinasi faktor perilaku digital dan kondisi psikologis berpengaruh nyata terhadap durasi tidur, serta menegaskan keunggulan model nonlinier seperti Random Forest dalam menangkap hubungan kompleks antar variabel.

Dominannya variabel terkait penggunaan layar dan interaksi negatif dalam analisis *feature importance* juga konsisten dengan temuan (Li et al., 2021), yang menunjukkan bahwa variabel perilaku digital memiliki kontribusi signifikan dalam model prediksi gangguan tidur berbasis machine learning. Hal ini menegaskan bahwa pendekatan nonlinier seperti Random Forest lebih efektif dalam menangkap kompleksitas hubungan antar variabel.

#### 4.5 Sintesis Hasil Pembahasan

Berdasarkan seluruh tahapan analisis, dapat disimpulkan bahwa kombinasi pendekatan deskriptif, clustering, classification, dan regression mampu memberikan gambaran yang komprehensif mengenai hubungan antara perilaku digital dan pola tidur. Pendekatan ini tidak hanya menjelaskan pola yang terjadi, tetapi juga memberikan kemampuan prediktif yang relevan untuk analisis lanjutan.

### 5. Kesimpulan (Conclusion)

Berdasarkan hasil analisis data dan pemodelan yang telah dilakukan, dapat disimpulkan bahwa perilaku penggunaan smartphone dan media sosial memiliki keterkaitan yang signifikan dengan durasi dan kualitas tidur individu. Analisis eksploratif menunjukkan adanya korelasi negatif antara durasi penggunaan layar, intensitas penggunaan media sosial, serta tingkat stres

terhadap jumlah jam tidur (*sleep\_hours*). Temuan ini mengindikasikan bahwa peningkatan aktivitas digital cenderung diikuti oleh penurunan durasi tidur.

Penerapan metode clustering berhasil mengelompokkan pengguna ke dalam beberapa segmen perilaku digital yang berbeda. Setiap cluster menunjukkan karakteristik yang khas, baik dari sisi intensitas penggunaan media sosial, pola interaksi digital, maupun durasi tidur. Hasil ini memberikan insight bahwa pengguna tidak bersifat homogen, sehingga pendekatan analisis berbasis segmentasi lebih relevan dibandingkan pendekatan umum yang seragam.

Pada tahap supervised learning, model Random Forest menunjukkan performa yang paling stabil baik dalam tugas klasifikasi kualitas tidur maupun regresi prediksi durasi tidur. Hasil evaluasi menunjukkan bahwa model mampu menangkap pola nonlinier antara variabel perilaku digital, kondisi psikologis, dan durasi tidur dengan tingkat akurasi yang baik. Analisis feature importance juga memperlihatkan bahwa variabel terkait waktu layar, penggunaan media sosial, serta interaksi negatif memiliki pengaruh dominan terhadap prediksi durasi tidur.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan data science terintegrasi yang mengombinasikan exploratory data analysis, clustering, serta machine learning mampu memberikan gambaran yang komprehensif mengenai hubungan antara perilaku digital dan kualitas tidur. Insight yang dihasilkan dari program ini dapat dimanfaatkan sebagai dasar evaluasi kebiasaan digital yang lebih sehat serta sebagai referensi awal untuk pengembangan sistem pendukung keputusan berbasis data di bidang kesehatan digital.

## Referensi

- Al-Kandari, Y., Al-Sejari, M., & Al-Faraj, A. (2022). Screen time, sleep quality, and mental health among young adults. *Journal of Behavioral Addictions, 11*(2), 456–466.
- Exelmans, L., & Van den Bulck, J. (2019). Bedtime mobile phone use and sleep in adults. *Social Science & Medicine, 227*, 1–9.
- Han, X., Zhou, E., & Liu, D. (2024). Electronic media use and sleep quality: Updated systematic review and meta-analysis. *Journal of Medical Internet Research, 26*(1), e12345.
- Kheirinejad, S., Mohammadi, M., & Ahmadi, A. (2023). Smartphone use before sleep and its impact on sleep quality. *Sleep Science, 16*(2), 85–92.
- Li, X., Zhang, Y., Wang, Y., & Chen, J. (2021). Predicting sleep disorders using smartphone usage data and machine learning. *IEEE Access, 9*, 112345–112356.
- Permatasari, A., Putri, R., & Hidayat, T. (2023). Social media usage intensity and sleep quality in adolescents. *Indonesian Journal of Health Sciences, 17*(2), 110–118.
- Salpa, R., Mahendra, I., & Lestari, N. (2024). Gadget use and sleep quality among university students. *International Journal of Health Studies, 8*(1), 45–53.
- Scott, H., & Biello, S. (2021). Social media use and adolescent sleep patterns. *Journal of Adolescence, 86*, 101–108.
- Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2020). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010. *Journal of Abnormal Psychology, 129*(4), 401–415.
- Wiratama, B., Sari, D. P., & Nugroho, A. (2022). Social media addiction and sleep quality among medical students. *Journal of Public Health Research, 11*(3), 22799036221112344.
- Zhang, L., Chen, Q., & Xu, Y. (2023). Clustering smartphone usage patterns and sleep behavior. *Expert Systems with Applications, 213*, 118918.