



Pengelompokkan Tingkat Stres Remaja Terhadap Jam Tidur dan Aktivitas Media Sosial Menggunakan Metode K-Means

Ivana Naomi Simanjuntak¹, Abdi Rahim Damanik²

¹Teknik Informatika, STIKOM Tunas Bangsa, Indonesia

²Sistem Informasi, STIKOM Tunas Bangsa, Indonesia

E-Mail : ¹ivananaomi27123@gmail.com, ²abdirahimdamanik@stikomtunasbangsa.ac.id

Article Info

Article history:

Received
Revised
Accepted

Kata Kunci:

Stres Remaja
Jam Tidur
Media Sosial
Klastering
K-Means

Keywords:

Teen Stress
Sleep Duration
Social Media
Clustering
K-Means

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan tingkat stres remaja berdasarkan durasi tidur dan aktivitas media sosial menggunakan metode K-Means Clustering. Data diperoleh dari dataset "Mental Health Analysis among Teenagers" yang tersedia secara publik di Kaggle. Sebelum proses klusterisasi, data melalui tahap pra-pemrosesan berupa pembersihan data dan normalisasi menggunakan metode Min-Max Scaling. Pemilihan jumlah klaster optimal dilakukan menggunakan metode Elbow, yang menunjukkan bahwa tiga klaster merupakan jumlah yang paling representatif. Klastering dilakukan berdasarkan dua variabel, yaitu jam tidur dan durasi penggunaan media sosial, dan hasilnya divisualisasikan dalam bentuk scatter plot dua dimensi. Implementasi dilakukan menggunakan Python dan evaluasi model dilakukan melalui visualisasi dan analisis karakteristik masing-masing klaster. Hasil menunjukkan bahwa remaja dengan jam tidur cukup dan aktivitas media sosial rendah cenderung memiliki tingkat stres yang lebih rendah, sedangkan penggunaan media sosial yang tinggi atau jam tidur yang pendek berkorelasi dengan stres sedang hingga tinggi. Penelitian ini membuktikan bahwa kombinasi perilaku harian seperti tidur dan penggunaan media digital dapat digunakan untuk memetakan risiko stres remaja secara efektif melalui pendekatan unsupervised learning.

ABSTRACT

This study aimed to cluster the stress levels of teenagers based on their sleep duration and social media activity using the K-Means Clustering method. The data were obtained from the publicly available dataset "Mental Health Analysis among Teenagers" on Kaggle. Prior to clustering, the data underwent preprocessing steps including data cleaning and normalization using Min-Max Scaling. The optimal number of clusters was determined using the Elbow Method, which indicated that three clusters were the most representative. Clustering was based on two variables: sleep hours and social media usage, and the results were visualized using a two-dimensional scatter plot. The implementation was carried out using Python, and the model was evaluated through visualization and analysis of the characteristics of each cluster. The results showed that teenagers with sufficient sleep and low social media activity tended to experience lower stress levels, while those with high social media use or insufficient sleep correlated with moderate to high stress levels. This study demonstrated that daily behavior patterns such as sleep and digital activity can effectively be used to map adolescent stress risk using an unsupervised learning approach.

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ivana Naomi Simanjuntak,
Teknik Informatika, STIKOM Tunas Bangsa
Jln. Jendral Sudirman Blok A No. 1, 2&3 Pematangsiantar, 21127, Sumut, Indonesia.
Email: ivananaomi27123@gmail.com

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah memberikan dampak besar bagi kehidupan remaja, terutama dalam kebiasaan penggunaan media sosial dan pola tidur sehari-hari. Di satu sisi, media sosial telah menjadi sarana komunikasi yang penting bagi remaja dalam menjalin hubungan sosial, memperoleh informasi, dan mengekspresikan diri. Namun di sisi lain, penggunaan media sosial yang berlebihan juga dapat menimbulkan dampak negatif terhadap kesehatan mental, seperti stres, kecemasan, hingga depresi ringan (Kales et al., 2020). Selain itu, pola tidur remaja saat ini cenderung tidak teratur dan berkorelasi dengan meningkatnya gangguan psikologis, terutama jika dikombinasikan dengan penggunaan *gadget* hingga larut malam (Scott et al., 2022).

Berdasarkan laporan *World Health Organization* (2021), sekitar 13% remaja di duynia mengalami gangguan mental, dan sebagian besar tidak mendapatkan deteksi atau penanganan secara dini. Salah satu penyebab yang sering tidak disadari adalah kebiasaan perilaku harian seperti kurang tidur dan terlalu banyak terpapar media digital. Kondisi ini menunjukkan bahwa faktor-faktor perilaku harian berpotensi menjadi indikator penting dalam memetakan tingkat risiko kesehatan mental pada remaja.

Penelitian sebelumnya telah mengkaji hubungan antara durasi tidur, intensitas penggunaan media sosial, dan kondisi mental remaja secara terpisah. Sebagian besar pendekatan yang digunakan bersifat korelasional atau regresi linier, sehingga belum mampu memberika gambaran klasifikasi atau pengelompokan perilaku secara menyeluruh (Salamah et al., 2021). Padahal, dengan mengelompokkan individu berdasarkan kombinasi dari faktor-faktor tersebut, intervensi dan tindakan pencegahan dapat dilakukan dengan lebih terarah.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan pengelompokan (*clustering*) terhadap data kesehatan mental remaja berdasarkan aktivitas media sosial dan kebiasaan tidur menggunakan algoritma *K-Means Clustering*. Algoritma ini dipilih karena lebih efisien dalam menemukan struktur data dan mengidentifikasi klaster berdasarkan kemiripan perilaku. Data yang digunakan dalam penelitian ini mencakup variabel-variabel seperti durasi tidur, frekuensi penggunaan media sosial, dan indikator tingkat stres.

Kebaruan pada penelitian ini terletak pada integrasi dua variabel perilaku harian-media sosial dan pola tidur-dalam analisis berbasis *unsupervised learning* untuk memetakan potensi risiko kesehatan mental. Penelitian ini diharapkan dapat memberikan kontribusi dalam bentuk pendekatan berbasis data untuk mendeteksi kelompok remaja yang memiliki risiko stres tinggi, sehingga strategi intervensi psikososial dapat dilakukan secara lebih dini dan tepat sasaran.

2. METODE PENELITIAN

2.1. Teknik Pengumpulan Data

Sumber data pada penelitian ini diperoleh dari *website* Kaggle, dengan dataset yang bernama “*Mental Health Analysis among Teenagers*”. Dataset terdiri dari 5000 entri dengan 11 atribut, dan dari jumlah entri tersebut, hanya digunakan sebanyak 375 entri data. Data kemudia diklasterisasi ke dalam 3 klaster, di mana klaster menjelaskan seberapa tinggi tingkat stres yang dialami oleh remaja.

Tabel 1. Mental Health Analysis among Teenagers

Social_Media_Hours	Sleep_Hours	Survey_Stress_Score
9.654486346	5.198925522	3
9.158143482	8.866096662	5
5.028755201	4.94309483	2
7.951102825	5.262773303	5
1357458531	6.196080351	5
9.774381842	9.282089175	5
9.64353234	4.790720012	4
1.637289579	4.733405157	4
2.9942314248	6.96384993	5
1.001549975	5.658463247	2

2.2. Algoritma K-Means

Berikut langkah – langkah melakukan *clustering* dengan menggunakan algoritma *K- Means*:

- Tentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk.
- Membangkitkan k centroid (titik pusat *cluster*) awal secara random/acak, kemudian untuk menghitung *centroid cluster* ke-i berikutnya, digunakan rumus sebagai berikut:

$$v = \sqrt{\frac{\sum_{i=1}^n x_i}{n}}$$

Dimana:

i : 1,2,3,...n

v : centroid pada *cluster*

x_i : objek ke- i

n : banyaknya jumlah objek yang menjadi anggota *cluster*

- c. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

d_{ij} : Jarak objek antara objek x dan y

n : Jumlah Atribut

x_i : Objek Data

y_i : Data *Cluster*

- d. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya (C).
e. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

2.3. Preprocessing

Sebelum dilakukan proses analisis lebih lanjut, data terlebih dahulu melalui tahap pra-pemrosesan guna memastikan kualitas dan konsistensi data yang digunakan. Langkah pertama dalam tahap ini adalah pembersihan data, yang mencakup identifikasi serta penanganan nilai-nilai kosong (*missing values*) yang dapat mengganggu validitas hasil analisis. Nilai-nilai kosong tersebut dihapus atau diimputasi tergantung pada konteks dan distribusi data. Selanjutnya, untuk menyamakan skala antar fitur dan menghindari dominasi atribut dengan rentang nilai yang besar terhadap algoritma pembelajaran mesin yang sensitif terhadap skala, dilakukan proses normalisasi data.

Teknik normalisasi yang digunakan adalah *Min-Max Scaling*, di mana setiap nilai atribut dipetakan ke dalam rentang antara 0 hingga 1. Proses normalisasi ini bertujuan untuk meningkatkan kinerja algoritma serta mempercepat konvergensi dalam proses pelatihan model. Seluruh proses pra-pemrosesan ini dilaksanakan menggunakan *software RapidMiner Studio*, yang menyediakan antarmuka visual serta modul-modul yang mendukung untuk pembersihan data dan transformasi numerik secara efisien dan sistematis.

2.4. Pelatihan Model Menggunakan K-Means

Penelitian ini mengimplementasikan metode *K-Means* melalui bahasa pemrograman *Python*. Proses pelatihan data dilakukan dengan menyusun skrip program menggunakan pustaka *Scikit-Learn*, yang sudah banyak digunakan dalam pengembangan system yang berbasis pembelajaran mesin. Implementasi kode *Python* dapat dilihat sebagai berikut:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
df = pd.read_csv(r'C:\Users\HP\Documents\mental_health_analysis_dataset.csv')
print("Column names before cleaning:", df.columns)
features = df[['Sleep_Hours', 'Social_Media_Hours']]
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)
kmeans = KMeans(n_clusters=3, random_state=42)
df['Cluster'] = kmeans.fit_predict(scaled_features)
plt.figure(figsize=(14,7))
sns.scatterplot(data=df, x='Sleep_Hours', y='Social_Media_Hours', hue='Cluster', palette='Set1',
s=10)
plt.title('KMeans Clustering: Sleep vs Social Media Hours')
plt.xlabel('Sleep Hours')
plt.ylabel('Social Media Hours')
plt.grid(True)
plt.legend(title='Cluster')
plt.show()
```

Dengan menggunakan pendekatan algoritma ini, *Python* memberi kemudahan dalam mengelola data dengan efisien dalam proses komputasi. Penggunaan kode menjadi basis dalam membentuk model yang mampu mengidentifikasi dan menghasilkan pola klasterisasi secara otomatis.

2.5. Evaluasi Hasil Model

Setelah data siap diolah, dilakukan penentuan jumlah klaster optimal menggunakan metode *Elbow*. Metode ini bekerja dengan menghitung nilai *inertia* untuk berbagai nilai *k* (jumlah klaster) dan memvisualisasikannya dalam bentuk grafik. Titik tekuk pada grafik tersebut dipilih sebagai jumlah klaster yang optimal, di mana pada penelitian ini diperoleh nilai *k* = 3. Dengan jumlah klaster yang telah ditentukan, algoritma *K-Means* kemudian diimplementasikan secara iteratif untuk mengelompokkan data ke dalam klaster-klaster berdasarkan jarak terdekat terhadap pusat klaster (*centroid*), dengan tujuan meminimalkan total jarak antar data dan *centroid*-nya.

Evaluasi terhadap hasil *clustering* dilakukan menggunakan dua pendekatan. Pertama, melalui perhitungan nilai *Silhouette Coefficient* untuk mengukur kualitas pemisahan antar klaster. Kedua, dengan visualisasi hasil klaster dalam bentuk *scatter plot* dua dimensi, di mana sumbu X merepresentasikan durasi tidur dan sumbu Y menggambarkan durasi penggunaan media sosial. Tahap akhir dari metode ini adalah interpretasi klaster, yang dilakukan dengan menganalisis karakteristik dominan dari masing-masing klaster, khususnya berdasarkan distribusi rata-rata waktu tidur dan penggunaan media sosial. Hasil interpretasi ini kemudian dikaitkan dengan tingkat stres yang mungkin dialami oleh remaja dalam setiap klaster, sehingga dapat memberikan gambaran awal terhadap kelompok yang memerlukan perhatian lebih dalam aspek kesehatan mental.

Algoritma ini menggunakan rumus jarak *Euclidean* sebagai ukuran kedekatan antara data dan *centroid*. Rumus tersebut dituliskan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

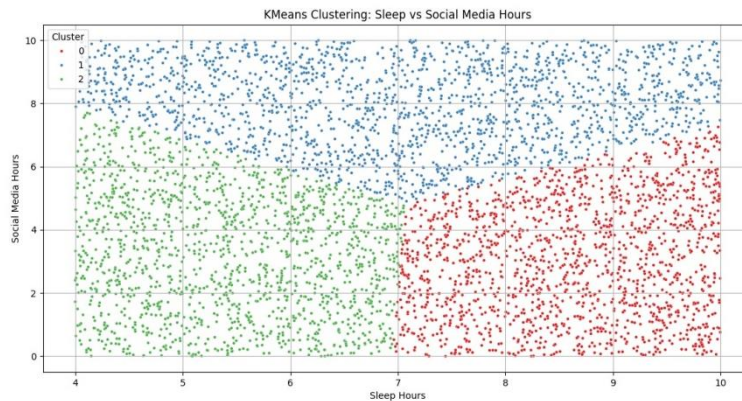
2.6. Interpretasi Klaster

Setelah proses klasterisasi selesai dilakukan, langkah selanjutnya adalah interpretasi terhadap masing-masing klaster yang terbentuk. Tujuan dari tahap ini adalah untuk memahami karakteristik utama yang membedakan setiap klaster, sehingga dapat memberikan makna yang relevan terhadap hasil analisis. Analisis dilakukan dengan mengevaluasi ciri khas dari masing-masing klaster berdasarkan tiga variabel utama, yaitu durasi tidur, waktu penggunaan media sosial, dan tingkat stres (diukur berdasarkan skor stres). Ketiga variabel ini dipilih karena dianggap memiliki hubungan yang signifikan dan saling memengaruhi dalam konteks kesejahteraan individu.

Interpretasi dilakukan dengan cara membandingkan nilai rata-rata (*mean*), distribusi, serta pola dominan dari ketiga variabel tersebut dalam setiap klaster. Dengan demikian, dapat diidentifikasi apakah suatu klaster memiliki kecenderungan terhadap durasi tidur yang pendek atau panjang, intensitas penggunaan media sosial yang tinggi atau rendah, serta tingkat stres yang rendah, sedang, atau tinggi. Hasil interpretasi ini tidak hanya memberikan pemahaman mendalam terhadap karakteristik anggota klaster, tetapi juga menjadi dasar untuk menyusun rekomendasi atau strategi intervensi yang lebih tepat sasaran sesuai dengan profil masing-masing klaster.

3. HASIL DAN PEMBAHASAN

Pengelompokan data dilakukan dengan menggunakan algoritma *K-Means* setelah melalui proses normalisasi dan penentuan jumlah klaster optimal menggunakan metode *Elbow*. Berdasarkan grafik *Elbow* yang telah dianalisa sebelumnya, diperoleh bahwa jumlah klaster optimal adalah tiga (*K*=3). Klastering dilakukan menggunakan dua atribut utam, yaitu *Sleep_Hours* yang merepresentasikan durasi tidur harian, dan *Social_Media_Hours* yang mencerminkan waktu yang dihabiskan untuk menggunakan sosial media setiap harinya.



Gambar 1. Hasil Visualisasi Klaster

Hasil visualisasi klaster yang ditampilkan pada Gambar 1 menunjukkan bahwa data terbagi dengan baik ke dalam tiga kelompok berdasarkan sebaran dua variabel tersebut. Warna yang berbeda dalam visualisasi tersebut menandakan masing-masing klaster, yaitu Klaster 0 (merah), Klaster 1 (biru), dan klaster 2 (hijau). Distribusi titik-titik menunjukkan pemisahan yang cukup jelas berdasarkan kombinasi jam tidur dan aktivitas media sosial.

Hasil ini menguatkan sejumlah studi terdahulu yang menyatakan bahwa kualitas dan kuantitas tidur memiliki korelasi negative terhadap tingkat stress. Studi oleh Fryer dan Dinsmore (2020) menyimpulkan bahwa durasi tidur yang kurang dapat meningkatkan kerentanan terhadap tekanan psikologis, terutama pada kelompok usia remaja. Selain itu, penelitian oleh Stewart (2021) menunjukkan bahwa penggunaan media sosial yang berlebihan berkaitan dengan gejala stress, kecemasan dan kelelahan digital (*digital fatigue*). Maka dari itu, kombinasi antara waktu tidur yang cukup dan manajemen waktu penggunaan media sosial yang baik memiliki potensi menjadi pendekatan preventif dalam menjaga kesehatan mental remaja.

Dengan demikian, hasil klastering ini memberikan gambaran pola perilaku harian remaja yang berkaitan erat dengan tingkat stres yang dialami. Analisis berbasis *unsupervised learning* ini dapat digunakan sebagai dasar dalam pengembangan system peringatan dini berbasis data atau strategi intervensi preventif yang lebih terarah bagi kalangan pelajar dan remaja.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengelompokkan tingkat stress pada remaja berdasarkan durasi tidur dan aktivitas penggunaan media sosial dengan menggunakan metode *K-Means Clustering*. Data yang digunakan diperoleh dari dataset di platform *Kaggle*, kemudian melalui tahapan pre-pemrosesan, penentuan jumlah klaster optimal menggunakan metode *Elbow*, serta implementasi dan evaluasi klaster menggunakan *Python* dan *RapidMiner*.

Hasil klastering menunjukkan bahwa tiga klaster utama dapat terbentuk secara jelas: (1) klaster dengan jam tidur tinggi dan aktivitas media sosial yang rendah diasosiasikan dengan tingkat stres yang rendah, (2) klaster dengan aktivitas media sosial yang tinggi berkorelasi dengan tingkat stres yang sedang hingga tinggi, serta (3) klaster dengan durasi tidur rendah yang mencerminkan tingkat stres sedang. Temuan ini mengindikasikan adanya hubungan yang signifikan antara kebiasaan tidur dan penggunaan media sosial terhadap kondisi psikologis remaja, khususnya tingkat stres.

Dengan demikian, *unsupervised learning* berbasis *K-Means* terbukti efektif dalam mengidentifikasi pola stres pada remaja dari data kebiasaan harian. Penelitian ini juga membuka peluang dalam penerapan system berbasis data untuk mendukung program pemantauan dan pencegahan stres pada remaja. Studi serupa kedepannya juga dapat dikembangkan dengan menambahkan variabel lain yang lebih kompleks, guna meningkatkan akurasi model dan ketepatan intervensi.

ACKNOWLEDGEMENTS

Terimakasih kepada platform yang sudah membuka akses dan menyediakan dataset secara gratis sebagai dasar dari penelitian ini. Ucapan terimakasih juga disampaikan kepada dosen pembimbing dan seluruh pihak yang sudah membantu dalam proses implementasi dan evaluasi hasil menggunakan *RapidMiner* dan *Python*.

REFERENCES

- Al Yasin, R., Roro Kirani Annisa Anjani, R., Salsabil, S., Rahmayanti, T., & Amalia, R. (2022). *PENGARUH SOSIAL MEDIA TERHADAP KESEHATAN MENTAL DAN FISIK REMAJA: A SYSTEMATIC REVIEW*. 3(2).
- Fadhilla, M., Wandri, R., Hanafiah, A., Setiawan, P. R., Arta, Y., & Daulay, S. (2025). Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa. *Journal of Informatics Management and Information Technology*, 5(1), 40–47. <https://doi.org/10.47065/jimat.v5i1.473>
- Faridi. Ahmad, Vidyarini, A., & Syifa, N. (2024). Hubungan antara Penggunaan Media Sosial, Aktivitas Fisik dan Tingkat Stres Terhadap Status Gizi Usia 18–25 Tahun Di RW. 03 Kelurahan Kebayoran Lama Utara. *Jurnal Kesehatan Indonesia*.
- Hayran, C., & Anik, L. (2021). Well-being and fear of missing out (Fomo) on digital content in the time of covid-19: A correlational analysis among university students. *International Journal of Environmental Research and Public Health*, 18(4), 1–13. <https://doi.org/10.3390/ijerph18041974>
- Kholidin¹, N., Suarna², N., & Prihartono³, W. (n.d.). *Segmentasi Gaya Hidup Mahasiswa Menggunakan Algoritma K-Means Clustering*.
- Maulida, E., Fitri Hidayati, A., Silalahi, F. Y., & Prasetyo, Y. D. (2024). Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup. In *Journal Information Engineering and Educational Technology* (Vol. 08).
- Muhafilah, I., & Suwarningsih, S. (2023). Durasi Penggunaan Media Sosial dan Tingkat Stres dengan Kualitas Tidur pada Remaja. *Jurnal Ilmu Kesehatan Masyarakat*, 12(05), 346–351. <https://doi.org/10.33221/jikm.v12i05.2076>
- Nisa, I. C. (2024). PENERAPAN ALGORITMA K-MEANS DALAM PENGKATEGORIAN INSOMNIA. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 8(1), 117–122. <https://doi.org/10.46880/jmika.Vol8No1.pp117-122>
- Rezki, A. N., Naryoso, A., & Pradekso, T. (n.d.). *HUBUNGAN INTENSITAS PENGGUNAAN MEDIA SOSIAL DENGAN TINGKAT STRES*. <https://fisip.undip.ac.id>
- Septiana, N. Z. (2021). Dampak Penggunaan Media Sosial Terhadap Kesehatan Mental Dan Kesejahteraan Sosial Remaja Dimasa Pandemi Covid-19. *Nusantara of Research : Jurnal Hasil-Hasil Penelitian Universitas Nusantara PGRI Kediri*, 8(1), 1–13. <https://doi.org/10.29407/nor.v8i1.15632>