

Analisis Kecenderungan Depresi Remaja dalam Penggunaan Media Sosial dengan Algoritma *Random Forest*

DOI: <http://dx.doi.org/10.35889/progresif.v22i3.3257>

Creative Commons License 4.0 (CC BY –NC)



I Made Adi Bhaskara^{1*}, I Made Surya Kumara², Ni Wayan Diana Ekayani³

¹² Teknik Komputer, Universitas Warmadewa, Bali, Indonesia

³ Kedokteran, Universitas Warmadewa, Bali, Indonesia

*e-mail *Corresponding Author*: adibhaskara@warmadewa.ac.id

Abstract

Depression is a major factor in Bali's high suicide rates, particularly among social media-active adolescents. This study analyzes the effects of social media usage intensity, addiction, and sleep quality on adolescents' mental health using the Random Forest algorithm. Quantitative research involved data exploration, preprocessing, modeling, hyperparameter tuning, and evaluation with R^2 and MSE. From 36 valid respondents, the Addicted Score was the strongest predictor of sleep quality and mental health (feature importance ± 0.7), exceeding average daily usage (0.2–0.3). Negative correlations linked addiction to mental health ($r = -0.6257$) and sleep quality ($r = -0.4588$). Models achieved R^2 of 0.19 (sleep quality) and 0.38 (mental health). Findings show addictive behavior impacts more than usage duration, so prevention should focus on controlling social media addiction.

Keywords: *Random Forest; Regression; Depression*

Abstrak

Depresi merupakan salah satu faktor utama yang berkontribusi terhadap tingginya angka bunuh diri di Bali, terutama pada remaja yang aktif menggunakan media sosial. Penelitian ini bertujuan menganalisis pengaruh intensitas penggunaan media sosial, kecanduan media sosial, dan kualitas tidur terhadap kesehatan mental remaja menggunakan algoritma Random Forest. Penelitian kuantitatif dilakukan melalui tahapan eksplorasi data, preprocessing, pemodelan, hyperparameter tuning, dan evaluasi menggunakan R^2 serta MSE. Berdasarkan 36 data responden valid, Addicted Score menjadi faktor paling dominan dalam memprediksi kualitas tidur dan kesehatan mental (feature importance $\pm 0,7$), dibandingkan rata-rata penggunaan media sosial harian (0,2–0,3). Korelasi menunjukkan hubungan negatif antara kecanduan media sosial dengan kesehatan mental ($r = -0,6257$) dan kualitas tidur ($r = -0,4588$). Model menghasilkan R^2 0,19 untuk kualitas tidur dan 0,38 untuk kesehatan mental. Temuan menunjukkan bahwa perilaku adiktif lebih berpengaruh daripada durasi penggunaan, sehingga pencegahan perlu menekankan pengendalian kecanduan media sosial.

Kata kunci: *Random Forest; Regresi; Depresi*

1. Pendahuluan

Berdasarkan penelitian Onie dkk. (2024), Bali menempati posisi sebagai provinsi dengan angka bunuh diri tertinggi di Indonesia berdasarkan analisis berbagai sumber data nasional periode 2016–2021 [1]. Chan dkk. (2025) melaporkan bahwa individu dengan depresi memiliki risiko kematian akibat bunuh diri hampir 10 kali lebih tinggi dibandingkan individu tanpa depresi [2]. Depresi adalah masalah kesehatan mental yang semakin mendapatkan perhatian di kalangan remaja. Berdasarkan data yang diungkapkan oleh *World Health Organization* (WHO) pada tahun 2021, depresi adalah penyebab utama gangguan kesehatan mental di seluruh dunia, termasuk di Indonesia [3]. Remaja rentan mengalami gangguan mental akibat perubahan fisik dan

psikologis yang dialami, serta tekanan akademik, sosial, dan keluarga [4]. Di tengah perkembangan teknologi, penggunaan media sosial, terutama mobilitas dalam berinteraksi secara maya melalui *handphone* semakin mendominasi kehidupan sehari-hari siswa. Di Indonesia, fenomena ini semakin relevan dengan semakin tingginya penetrasi internet dan penggunaan ponsel pintar di kalangan pelajar dimana interaksi *online* dapat meningkatkan rasa kesepian dan memperburuk kesehatan mental dibanding tatap muka [5]. Berdasarkan data Asosiasi Penyelenggara Jasa Internet Indonesia (APJII, 2023), lebih dari 90% pelajar SMA memiliki akses internet yang sering digunakan untuk mengakses media sosial seperti Instagram, TikTok, dan Facebook [6]. Penggunaan media sosial yang tidak terkendali pada remaja diharapkan dapat berpengaruh pada kesejahteraan psikologis mereka.

Objek kajian pada penelitian yang dilakukan berfokus pada remaja sebagai kelompok yang rentan terhadap dampak negatif media sosial, di mana intensitas penggunaan yang tinggi sering kali memicu gangguan psikologis seperti depresi. Masalah utama yang muncul dari judul tersebut adalah hubungan antara pola penggunaan media sosial dengan kecenderungan depresi, yang dapat diperburuk oleh faktor *mediating* seperti *social media disorder*. Hal ini sejalan dengan berbagai temuan empiris sebelumnya. Mustaqina dan Mulyeni menemukan bahwa penggunaan media sosial berdampak pada gangguan psikologis mahasiswa, menunjukkan kerentanan kelompok muda terhadap masalah mental akibat paparan digital yang intens [7]. Sementara itu, Kaligis et al. mengungkapkan bahwa *social media disorder* berperan sebagai mediator antara intensitas penggunaan media sosial dan perilaku *self-harm* pada remaja Indonesia [8]. Selain itu, penelitian dalam Triwikrama: Jurnal Ilmu Sosial mengeksplorasi dampak penggunaan media sosial terhadap kesehatan mental remaja, yang semakin memperkuat urgensi analisis prediktif seperti Random Forest untuk mendeteksi kecenderungan depresi dini pada kelompok tersebut [9].

Beragam penelitian menunjukkan adanya hubungan negatif antara intensitas penggunaan media sosial, yang diukur melalui frekuensi dan durasi, dengan kondisi psikologis seperti harga diri dan kesehatan mental, khususnya pada remaja [10]. mengungkapkan bahwa harga diri yang rendah berkaitan erat dengan ketergantungan pada media sosial, sementara Han et al. (2018) menambahkan bahwa penggunaan berlebihan dapat memengaruhi pola kebiasaan yang berdampak pada stabilitas emosional [11]. Temuan ini selaras dengan Teori Regulasi Diri Sosial Daring yang dikemukakan Brandenberg et al. (2018), yang menyatakan bahwa jejaring sosial digunakan untuk regulasi diri, sehingga meningkatkan ketergantungan dan memicu dampak psikologis seperti kecemasan atau depresi [12]. Selain itu, Ophir et al. (2020) menemukan bahwa tingkat penggunaan teknologi digital yang tinggi berkorelasi dengan gejala depresi dan penurunan harga diri, dengan ketergantungan sebagai faktor mediasi, yang juga diperkuat oleh meningkatnya intensitas penggunaan *smartphone* [13]. Di sisi lain, perkembangan teknologi dan penggunaan media sosial yang melonjak, terutama selama pandemi, turut memperumit dinamika kesehatan mental. Saha et al. (2020) mencatat bahwa peningkatan frekuensi dan durasi penggunaan media sosial berkorelasi dengan ekspresi simptomatik seperti depresi, dengan ketergantungan sebagai mediator, sehingga pendekatan analisis data besar seperti *Random Forest* menjadi relevan untuk memetakan pola ini [14]. Hal ini didukung oleh Eichstaedt et al. (2018), yang menunjukkan bahwa data penggunaan media sosial selama enam bulan sebelum diagnosis depresi dapat memprediksi kecenderungan depresi dengan akurasi tinggi (AUC = 0.72) melalui metode prediktif serupa [15]. Sementara itu, Qi (2019) menyoroti bahwa menjamurnya teknologi *mobile*, meskipun mendukung pendidikan, juga menimbulkan tantangan seperti stress di kalangan pengguna di pendidikan tinggi, menegaskan kompleksitas hubungan antara intensitas penggunaan dan dampak psikologis [16]. Penggunaan media sosial yang berlebihan, terutama pada malam hari, dapat mengganggu kualitas tidur remaja melalui paparan cahaya biru yang menekan produksi melatonin serta gangguan ritme sirkadian [17]. Penurunan kualitas tidur ini diketahui meningkatkan risiko gangguan emosional dan depresi karena memengaruhi regulasi emosi, fungsi kognitif, serta keseimbangan hormonal [18]. Oleh karena itu, kualitas tidur diperkirakan menjadi variabel mediasi penting yang menjembatani hubungan antara ketergantungan media sosial dan depresi pada remaja [19] [20].

Random Forest merupakan algoritma pembelajaran mesin berbasis *ensemble* yang menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang akurat dan stabil. Pada penelitian Algoritma *Random Forest* dengan regresi logistik pada *dataset* kecil dan menunjukkan bahwa Random Forest sering kali lebih unggul karena kemampuannya menangani interaksi kompleks dan non-linieritas dalam data [21]. *Random Forest* adalah metode *supervised*

learning yang berfungsi dalam menyelesaikan masalah klasifikasi dan regresi [22]. Regresi *Random Forest* unggul dalam menangani hubungan tidak linier antara variabel *input* dan *output* karena mampu memodelkan pola kompleks tanpa asumsi distribusi tertentu. Berbeda dengan klasifikasi *Random Forest* yang fokus pada pemisahan kelas diskrit, regresi *Random Forest* memprediksi nilai kontinu dengan akurasi tinggi melalui *ensemble decision tree*. Algoritma ini juga tahan terhadap *overfitting* dan efektif menangani data dengan interaksi variabel yang rumit.

Berbagai studi empiris dalam 5 tahun terakhir menunjukkan bahwa *Random Forest Regression* secara konsisten mengungguli *Multiple Linear Regression (MLR)* dalam memprediksi gejala depresi karena kemampuannya mendeteksi efek interaksi *nonlinear* dan *high-order* tanpa memerlukan spesifikasi model eksplisit, sementara MLR rentan terhadap pelanggaran asumsi linearitas, multikolinearitas, serta sensitivitas terhadap *outlier* pada dataset berukuran kecil [23] [24]. Selain itu, analisis mediasi tradisional memiliki keterbatasan dalam menangani *confounding* kompleks, measurement error pada mediator, dan ketidakpastian estimasi efek tidak langsung, sedangkan *Random Forest Regression* melalui pendekatan *ensemble* serta teknik interpretabilitas seperti SHAP memberikan hasil yang lebih *robust* dan mengurangi risiko *overfitting* pada sampel terbatas [25][26].

Berdasarkan data survei yang dikumpulkan dari remaja di Bali, diharapkan mampu memberikan hasil yang lebih relevan karena Bali merupakan salah satu wilayah dengan angka bunuh diri yang relatif tinggi, di mana depresi menjadi salah satu faktor penyebab utamanya. Depresi pada remaja dipengaruhi oleh berbagai faktor, salah satunya adalah penggunaan media sosial yang telah menjadi bagian dari aktivitas sehari-hari. Penelitian terdahulu berjudul "*Social Media Dependency, Depression and Self-Esteem for Cymraeg and English-Speaking Adolescents*" menggunakan fitur frekuensi dan durasi penggunaan media sosial untuk menganalisis keterkaitannya dengan depresi dan harga diri. Berbeda dengan penelitian tersebut, studi yang dilakukan menerapkan algoritma *Random Forest* sebagai metode pembelajaran mesin untuk menganalisis hubungan antara intensitas penggunaan media sosial harian, skor kecanduan media sosial, kualitas tidur, dan kondisi kesehatan mental. Selain itu, riset ini menambahkan fitur kualitas tidur sebagai variabel prediktor, mengingat remaja sering menggunakan media sosial hingga larut malam sehingga durasi dan kualitas tidurnya berkurang juga dapat mempengaruhi kesehatan mental.

3. Metodologi

Metodologi penelitian kuantitatif ini menggunakan teknik data mining dengan pendekatan *Random Forest* untuk analisis regresi. Langkah-langkahnya mencakup: Identifikasi Masalah, Pengumpulan Data, Pengolahan Data melalui Teknik *Data Mining*, dan Evaluasi Model Regresi.

3.1 Identifikasi masalah

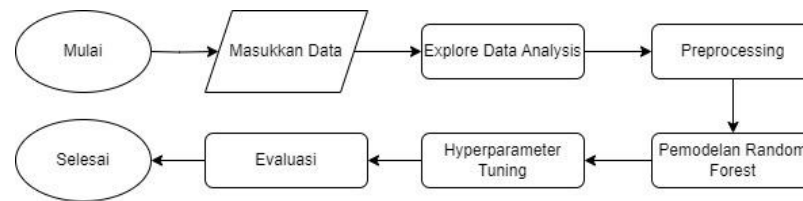
Penelitian yang dilakukan berfokus pada pengukuran dampak media sosial terhadap tingkat depresi menggunakan regresi pohon berbasis algoritma *Random Forest* dan fitur kesehatan mental lainnya, serta membandingkan hasil prediksi dengan nilai aktual kesehatan mental dan variabel pendukung dari metode tersebut.

3.2. Pengumpulan Data

Proses pengumpulan data merupakan tahapan yang sangat penting dalam penelitian ini. Data diperoleh melalui pendekatan kuantitatif dengan menggunakan kuesioner yang dibuat melalui *Google Form*. Responden kuesioner adalah siswa sekolah menengah atas (SMA) yang termasuk dalam kategori remaja berusia 15–19 tahun. Kuesioner dirancang semenarik mungkin agar mudah dipahami dan diisi oleh responden. Selain itu, beberapa pertanyaan validasi disertakan dalam kuesioner untuk memastikan bahwa responden mengisi dengan serius dan tidak asal-asalan, sehingga data yang diperoleh dapat dipercaya dan berkualitas. Setelah itu data dikategorikan berdasarkan jenis pertanyaannya yakni terkait rata-rata waktu penggunaan media sosial, seberapa besar gangguan tidur setelah menggunakan media sosial, skor kesehatan mental terkait tingkat depresi dan tingkat kepercayaan diri responden dan seberapa penting dan seberapa besar pengaruhnya media sosial dalam penggunaannya sehari-hari.

3.3. Pengolahan Data melalui Teknik Data Mining

Proses pengolahan data menggunakan teknik *Data Mining* merupakan tahapan yang tepat dalam penerapan algoritma *Random Forest*.



Gambar 1. Pemrosesan Data Mining

Pada *flowchart Random Forest* dijabarkan sebagai berikut:

- 1) *Input data* : dimulai dari input data melalui python dan menggunakan berbagai macam library panda
- 2) *Explore data analysis* : berbagai macam data yang dimiliki dicarikan keterkaitan sehingga mendapatkan hubungan dari data yang ada
- 3) *Preprocessing* : terdapat pembersihan data, penanganan nilai yang hilang, *encoding* fitur kategori (jika ada), dan penambahan normalisasi (jika diperlukan & data sudah bersih)
- 4) *Pemodelan Random Forest* : pembentukan model random forest untuk regresi menggunakan data pelatihan. Model algoritma ini memprediksi fitur-fitur yang ada. Hasil akhir didapatkan dari hasil vote mayoritas dari semua pohon dalam gabungan beberapa pohon keputusan.
 - a. Menentukan data awal : Setiap pengambil keputusan memberikan skor berdasarkan preferensinya terhadap pentingnya masing-masing kriteria.
 - b. Normalisasi data awal : Kurangi nilai setiap kriteria dengan nilai terbaik (nilai ideal maksimum). Hasil pengurangan ini disebut k_{ij}' = nilai awal setelah dinormalisasi (skor kriteria ke- j oleh responden ke- i setelah dikurangi dengan nilai ideal)
 - c. Menentukan elemen matriks normalisasi a_{ij}

$$a_{ij} = \frac{k_{ij}}{\sum_{i=1}^m \sum_{j=1}^n k_{ij}} \quad (1)$$

Keterangan variabel :

a_{ij} : hasil perhitungan matriks data kriteria

k_{ij} : nilai awal untuk kriteria ke- j dari responden ke- i

i : indeks responden/pengambil keputusan (1, 2, ..., m)

j : indeks kriteria (1, 2, ..., n)

m : jumlah total pengambil Keputusan

n : jumlah total kriteria

- d. Menghitung nilai entropi untuk tiap kriteria

$$E_j = \frac{-1}{\ln m} \sum_{i=1}^m (a_{ij} \cdot \ln a_{ij}) \quad (2)$$

Keterangan variabel:

E_j : nilai bobot entropi untuk kriteria ke- j

$\ln$: logaritma natural (basis e)

m : jumlah total pengambil Keputusan

a_{ij} : nilai normalisasi dari responden ke- i terhadap kriteria ke- j

- e. Menghitung nilai dispersi (tingkat variasi)

$$D_j = 1 - E_j \quad (3)$$

Keterangan variabel:

D_j : nilai dispersi untuk kriteria ke- j , semakin tinggi nilainya, semakin besar tingkat variasi/ketidakpastian data

- f. Menormalisasi nilai dispersi menjadi bobot

$$W_j = \frac{D_j}{\sum_{j=1}^n D_j} \quad (4)$$

Keterangan variabel:

W_j : bobot akhir (bobot prioritas) untuk kriteria ke- j , diperoleh dari proporsi nilai dispersi terhadap total dispersi semua kriteria

- 5) *Hyperparameter tuning*: tahapan untuk pengaturan *hyperparameter* (*hyperparameter tuning*) dimana disesuaikan *hyperparameter* seperti jumlah pohon ($n_estimators$), kedalaman maksimum, minimum sampel untuk membagi *internal node*, minimum sampel untuk *node* daun, dan bibit untuk kemampuan reproduksi
- 6) Evaluasi : Evaluasi kinerja model menggunakan metrik regresi seperti *Mean Squared Error* (MSE), *R-squared*, atau *Mean Absolute Error* (MAE).

- a) Formula yang digunakan untuk perhitungan *Mean Squared Error* (MSE)

- Menentukan data awal: Setiap pengamatan keputusan nilai sesuai preferensi yang diberikan.
- Normalisasi data awal: Kurangkan tiap nilai dengan rata-rata, lalu kuadratkan hasilnya.
- Menentukan nilai MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

Keterangan :

MSE : hasil perhitungan *Mean Squared Error*

y_i : nilai sebenarnya ke- i

$\hat{y}_i$: nilai prediksi ke- i

n : jumlah pengamatan

- b) Rumus berikut digunakan untuk perhitungan *R-squared*:

- Menentukan data awal: Setiap pengamatan keputusan nilai sesuai preferensi yang diberikan.
- Normalisasi data awal: Kurangkan tiap nilai dengan rata-rata, lalu hitung variansi total.
- Menentukan nilai *R-squared*:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

R^2 : nilai koefisien determinasi

y_i : nilai sebenarnya ke- i

$\hat{y}_i$: nilai prediksi ke- i

$\bar{y}$: rata-rata nilai sebenarnya

n : jumlah pengamatan

- c) Formula berikut digunakan untuk menghitung *Mean Absolute Error* (MAE) :

- Menentukan data awal: Setiap pengamatan keputusan nilai sesuai preferensi yang diberikan.
- Normalisasi data awal: Kurangkan tiap nilai dengan nilai prediksi, lalu ambil nilai absolut.
- Menentukan nilai MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

MAE : hasil perhitungan *Mean Absolute Error*

y_i : nilai sebenarnya ke- i

$\hat{y}_i$: nilai prediksi ke- i

n : jumlah pengamatan

4. Hasil dan Pembahasan

Dataset yang diperoleh dari kuesioner *Google Form* yang ditujukan kepada remaja SMA berhasil mengumpulkan 52 *respons* dari target 100 responden. Kuesioner berisi 29 pertanyaan yang merupakan pilihan ganda dan 2 slot untuk *upload* berkas yakni untuk *screenshot* rata-rata waktu penggunaan Instagram dan lagi satu untuk *screenshot* rata-rata waktu penggunaan Tiktok.

Dari total 29 pertanyaan pada kuesioner terdapat 1 pertanyaan sebagai indikator rentang rata-rata waktu penggunaan media sosial, 16 pertanyaan sebagai indikator skor tingkat stress, indikator skor kecanduan media sosial sebanyak 10 pertanyaan dan skor gangguan tidur akibat penggunaan media sosial sebanyak 1 pertanyaan dimana masing-masing pertanyaan terdapat 5 pilihan dengan skor 1-5. Bukan hanya itu, pada kuesioner ada 5 pertanyaan validasi untuk memastikan kualitas data, sehingga setelah dilakukan penyaringan, diperoleh hanya 36 data yang valid, terdiri dari 26 responden perempuan dan 10 responden laki-laki.

Tabel 1. Fitur dan Nilai

No	Gender	Average_Daily_Socmed	Quality_Sleep	Mental_Health_Score	Addicted_Score
1	P	3	3	43	27
2	P	2	3	42	26
3	P	3	2	28	35
4	L	3	2	42	29
5	L	2	4	42	29
6	P	3	4	42	27
7	L	2	5	43	26
8	L	4	5	32	29
9	P	4	2	25	35

Fitur yang diperoleh dari hasil kuesioner meliputi jenis kelamin, rata-rata waktu penggunaan media sosial, gangguan waktu tidur akibat penggunaan media sosial yang berkaitan dengan kualitas tidur, skor kesehatan mental, serta skor kecanduan media sosial.

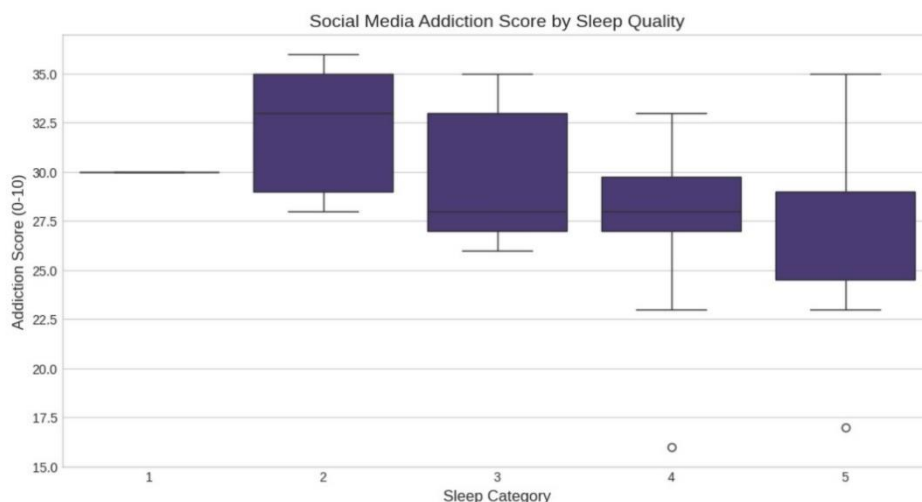
Tabel 2. Ringkasan Statistik

Statistik	No	Average_Daily_Socmed	Quality_Sleep	Mental_Health_Score	Addicted_Score
Count	36.000	36.000	36.000	36.000	36.000
Mean	18.500	2.694	3.472	35.333	29.000
Std. Dev.	10.536	0.749	1.158	6.663	4.523
Minimum	1.000	1.000	1.000	23.000	16.000
Kuartil 1 (25%)	9.750	2.000	2.000	31.000	27.000
Median (50%)	18.500	3.000	4.000	34.500	29.000
Kuartil 3 (75%)	27.250	3.000	4.000	41.250	33.000
Maksimum	36.000	4.000	5.000	54.000	36.000
75%	27.250.000	3.000.000	4.000.000	41.250.000	33.000.000

Tabel 2 menampilkan informasi dataset dalam format *Pandas DataFrame* dengan 36 entri dan 6 kolom (No, Gender, Average_Daily_Socmed, Quality_Sleep, Mental_Health_Score, Addicted_Score). Semua kolom memiliki 36 nilai *non-null*, dengan tipe data *int64* untuk kolom numerik dan *object* untuk Gender. Ringkasan statistik ini membantu dalam pra-pemrosesan data untuk *Random Forest* dengan menyediakan wawasan distribusi dan variabilitas data. Mean dan median (50%) menunjukkan nilai tengah, berguna untuk normalisasi. Standar deviasi (std) mengukur dispersi, mendukung pemilihan fitur penting. Kuartil (25%, 75%) dan rentang (min, max) memungkinkan deteksi *outlier* yang dapat memengaruhi pembagian simpul. Informasi ini

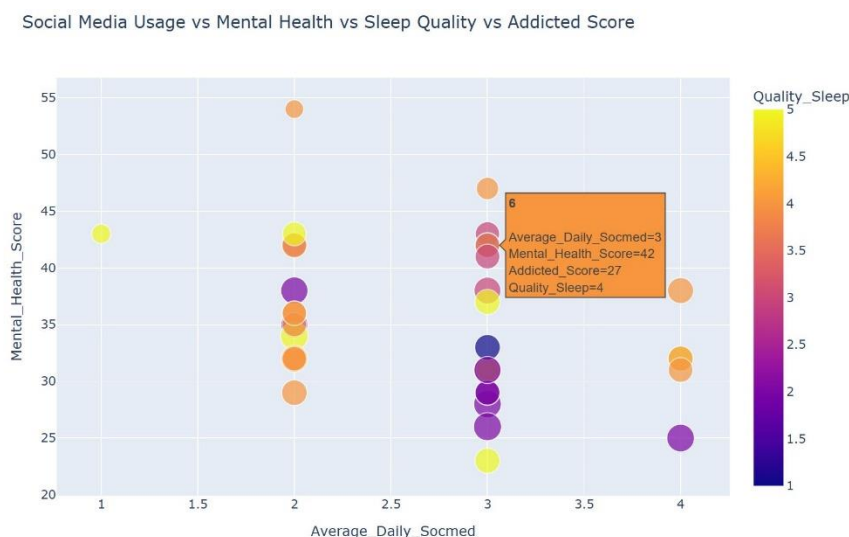
meningkatkan akurasi model dengan mengoptimalkan pemilihan fitur dan penanganan data tidak seimbang.

4.1. Explore Data Analysis (EDA)



Gambar 2. Hubungan Skor Kecanduan Media Sosial dengan Kualitas Tidur

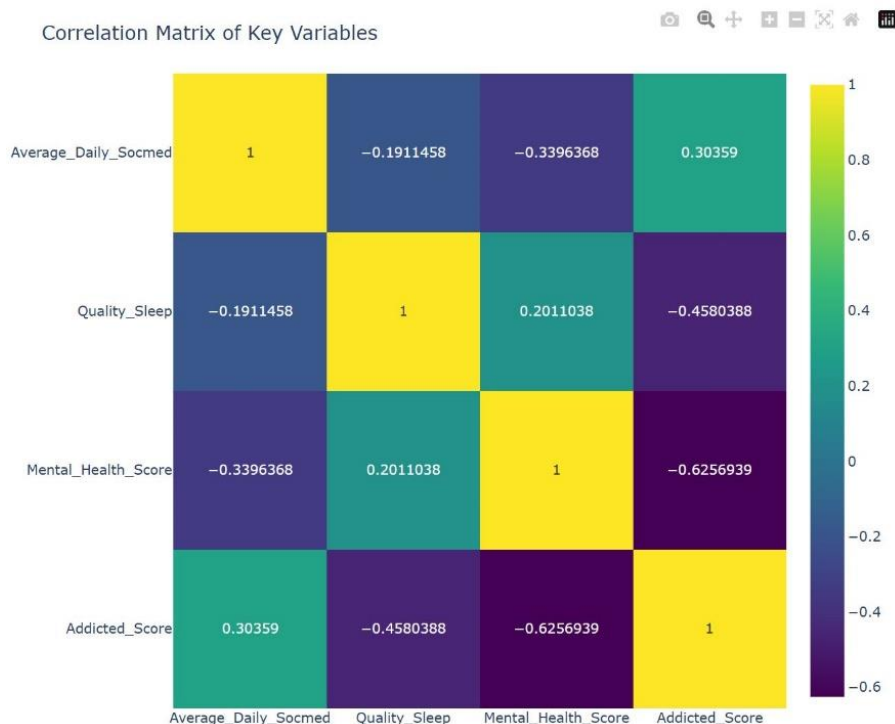
Grafik pada gambar 2 menampilkan skor kecanduan media sosial berdasarkan kategori kualitas tidur (1 hingga 5). Data disajikan dalam bentuk *box plot*, menunjukkan distribusi skor untuk setiap kategori. Kategori 2 memiliki skor median tertinggi sekitar 32,5, dengan rentang interkuartil (IQR) terluas. Kategori 1 menunjukkan *outlier* di bawah 30, sementara kategori 3, 4, dan 5 memiliki median yang lebih rendah, masing-masing sekitar 32, 27, dan 27, dengan IQR yang lebih sempit. Secara keseluruhan, kualitas tidur yang lebih rendah (kategori 1 dan 2) tampak berkorelasi dengan skor kecanduan yang lebih tinggi.



Gambar 3 Hubungan Penggunaan Media Sosial dan Kesehatan Mental dan Kualitas Tidur, dan Skor Kecanduan

Gambar 3 menampilkan hubungan antara skor kesehatan mental (sumbu Y) dan rata-rata penggunaan media sosial harian (sumbu X), dengan warna menunjukkan kualitas tidur (skala 1-5) dan ukuran titik mencerminkan skor kecanduan. Data menunjukkan bahwa skor kesehatan mental cenderung menurun seiring peningkatan penggunaan media sosial harian, dengan kualitas tidur lebih tinggi (warna kuning, 4-5) terkait skor kesehatan mental yang lebih baik (40-

50), sementara kualitas tidur rendah (warna ungu, 1-2) berkorelasi dengan skor lebih rendah (20-30) dan ukuran titik kecanduan yang bervariasi.



Gambar 4. Korelasi Matriks dari Variabel Kunci

Pada gambar 4 menjelaskan matriks korelasi antara rata-rata penggunaan media sosial harian (*Average_Daily_Socmed*), kualitas tidur (*Quality_Sleep*), skor kesehatan mental (*Mental_Health_Score*), dan skor kecanduan (*Addicted_Score*). Korelasi positif terkuat terlihat antara *Mental_Health_Score* dan *Quality_Sleep* (0.2011038), serta antara *Average_Daily_Socmed* dan *Addicted_Score* (0.30359). Korelasi negatif signifikan ditemukan antara *Addicted_Score* dan *Mental_Health_Score* (-0.6256939), serta antara *Quality_Sleep* dan *Addicted_Score* (-0.4580388), menunjukkan hubungan invers yang kuat.

Correlation Analysis:

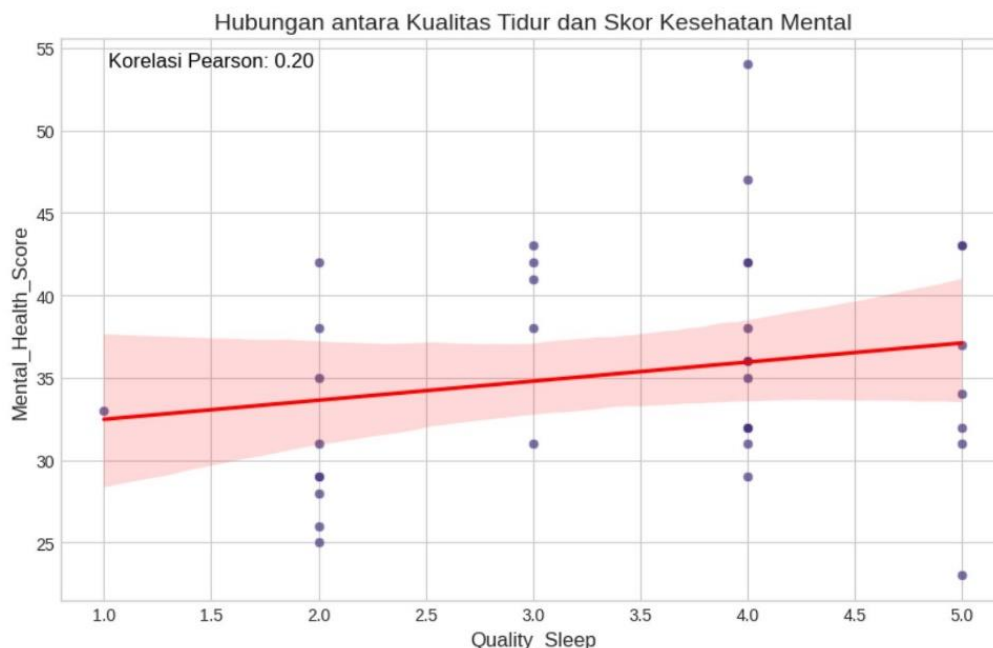
Usage Hours vs Quality Sleep: -0.1911 (Negative Weak correlation)
 Addiction Score vs Quality Sleep: -0.4580 (Negative Moderate correlation)
 Usage Hours vs Mental Health: -0.3396 (Negative Weak correlation)
 Addiction Score vs Mental Health: -0.6257 (Negative Moderate correlation)

Gambar 5. Analisis Korelasi dari Fitur yang Mempengaruhi dengan Fitur yang Dipengaruhi

Dijelaskan pada gambar 5 tentang analisis korelasi antara berbagai faktor, dengan hasil sebagai berikut:

Jam Penggunaan vs Kualitas Tidur: -0,1911 (Korelasi Lemah Negatif)
 Skor Kecanduan vs Kualitas Tidur: -0,4588 (Korelasi Sedang Negatif)
 Jam Penggunaan vs Kesehatan Mental: -0,3396 (Korelasi Lemah Negatif)
 Skor Kecanduan vs Kesehatan Mental: -0,6257 (Korelasi Sedang Negatif)

Nilai-nilai ini menunjukkan kekuatan dan arah hubungan linier antara variabel, di mana nilai negatif menunjukkan hubungan terbalik (semakin tinggi satu variabel, semakin rendah variabel lainnya). Klasifikasi menjadi "Lemas" atau "Sedang" didasarkan pada nilai absolut koefisien korelasi (biasanya, $|r| < 0,4$ adalah lemah, $0,4 \leq |r| < 0,7$ adalah sedang). Kode yang diberikan menghitung korelasi ini menggunakan Python dengan *DataFrame pandas* (df), di mana *corr()* menghitung koefisien korelasi Pearson antara pasangan variabel.



Gambar 6. Korelasi Pearson Kualitas Tidur dengan Skor Kesehatan Mental

Gambar 6 menunjukkan hubungan antara kualitas tidur (*Quality Sleep*) dan skor kesehatan mental (*Mental Health Score*) berdasarkan grafik *scatter plot*. Judul grafik adalah "Hubungan antara Kualitas Tidur dan Skor Kesehatan Mental", dengan koefisien korelasi Pearson sebesar 0,20 yang menunjukkan hubungan positif dengan kekuatan lemah. Sumbu x merepresentasikan kualitas tidur (skala 1–5), sedangkan sumbu y merepresentasikan skor kesehatan mental (skala 25–55). Garis regresi merah menunjukkan tren peningkatan yang disertai daerah kepercayaan (*confidence interval*).

Koefisien korelasi Pearson dihitung menggunakan Persamaan (8).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (8)$$

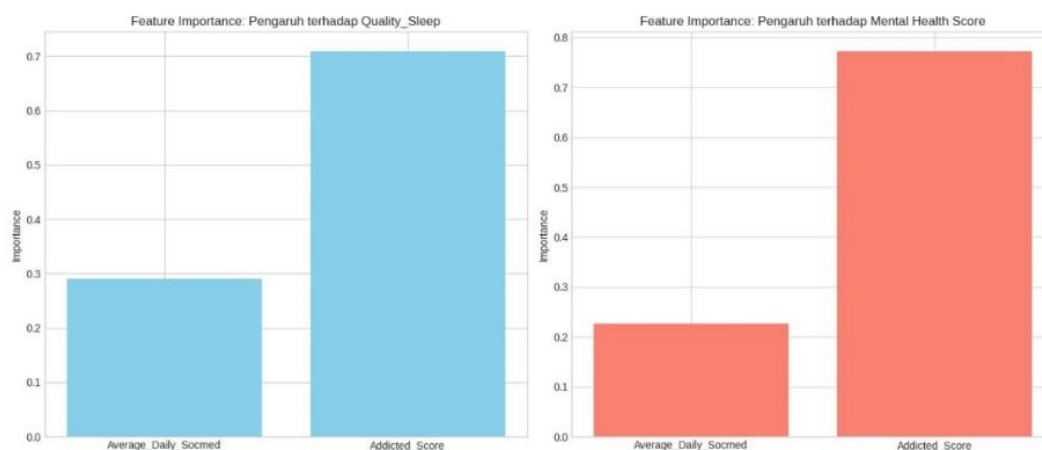
Pada Persamaan (8), x_i merupakan nilai pengamatan ke- i , sedangkan $\bar{x}$ adalah nilai rata-rata masing-masing variabel. Nilai koefisien korelasi sebesar 0,20 menunjukkan adanya hubungan linier positif dengan tingkat kekuatan yang lemah antara kualitas tidur dan skor kesehatan mental.

4.2 Penerapan Regresi *Random Forest*

Masuk ke tahapan penggunaan algoritma *Random Forest* yang merupakan metode pembelajaran mesin dimana tahapan yang dilakukan yakni :

- 1) Pendefinisian Fitur dan Variabel Target : Fitur (X) yakni rata-rata waktu penggunaan media sosial dan skor kecanduan media sosial sebagai variabel independen yang relevan serta variabel sebagai target (Y) yakni kualitas tidur dan skor kesehatan mental sebagai variabel target adalah nilai yang akan diprediksi berdasarkan *dataset*.
- 2) Pemisahan Data (*Train-Test Split*) : Data dibagi menjadi dua bagian yakni data pelatihan (*training*) untuk membangun model dan data pengujian (*testing*) untuk mengevaluasi kinerja model, dengan proporsi umum yakni 80:20 dengan *random state* 42.
- 3) *Preprocessing Data* : Tahap ini meliputi pembersihan data, penanganan nilai hilang melalui imputasi atau penghapusan, serta normalisasi data numerik untuk memastikan skala yang seragam. Dalam penelitian dilakukan normalisasi / Standardisasi data numerik agar data tidak bias
- 4) Pembentukan Model *Random Forest* untuk Regresi : Model regresi *Random Forest* dibangun menggunakan data pelatihan. Algoritma ini menggabungkan prediksi dari beberapa pohon keputusan yang dibentuk secara acak untuk menghasilkan output regresi yang lebih *robust*.

- 5) Pengaturan *Hyperparameter* (*Hyperparameter Tuning*) : *Hyperparameter* seperti jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah fitur yang dipertimbangkan pada setiap *split* (*max_features*) diatur untuk mengoptimalkan kinerja model.
- 6) Evaluasi Model : Kinerja model dievaluasi menggunakan metrik regresi seperti *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), dan *R-squared* untuk mengukur akurasi dan kecocokan model terhadap data pengujian.
- 7) Analisis Pentingnya Fitur (*Feature Importance*) : Pentingnya fitur dihitung untuk mengidentifikasi variabel yang paling berpengaruh terhadap prediksi model. Hasil ini divisualisasikan pada grafik gambar 7. menggunakan diagram seperti bar plot untuk mempermudah interpretasi.



Gambar 7. Fitur Penting yang Mempengaruhi Kualitas Tidur dan Kesehatan Mental

Gambar 7 menampilkan dua grafik batang yang membandingkan tingkat kepentingan (*importance*) antara "Average_Daily_Scored" dan "Addicted_Score" dalam dua kelompok yang berbeda.

- a) Kelompok Kiri (Warna Biru Muda):
 - Average_Daily_Scored*: Tingkat kepentingan sekitar 0,3.
 - Addicted_Score*: Tingkat kepentingan sekitar 0,7.
- b) Kelompok Kanan (Warna Merah Muda):
 - Average_Daily_Scored*: Tingkat kepentingan sekitar 0,2.
 - Addicted_Score*: Tingkat kepentingan sekitar 0,7.

Tidak ada proses perhitungan yang ditunjukkan dalam gambar; data kemungkinan besar merupakan hasil pengukuran atau analisis langsung. *Addicted_Score* konsisten menunjukkan tingkat kepentingan yang lebih tinggi (0,7) dibandingkan *Average_Daily_Scored* di kedua kelompok.

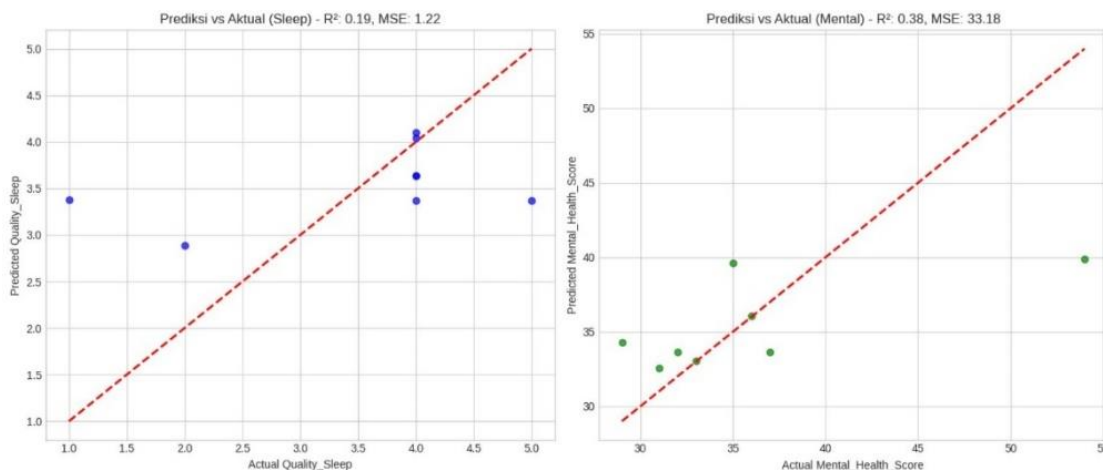
- 7) Visualisasi Prediksi vs Nilai Aktual : Hasil prediksi model dibandingkan dengan nilai aktual melalui *scatter plot* untuk menggambarkan sejauh mana model mampu memprediksi dengan akurat. Pada gambar 8 visualisasi ini membantu mengidentifikasi pola atau penyimpangan dalam prediksi.

Gambar 8 menampilkan dua grafik scatter yang membandingkan nilai prediksi dengan nilai aktual untuk dua variabel berbeda:

- a) Prediksi vs Aktual (Sleep):
 - $R^2 = 0.19$, MSE = 1.22.
 - Data titik (biru) menunjukkan hubungan lemah antara nilai prediksi dan aktual kualitas tidur, dengan garis regresi merah menunjukkan tren naik.
 - R^2 rendah menunjukkan model memiliki kekuatan penjelasan yang terbatas.
- b) Prediksi vs Aktual (Mental):
 - $R^2 = 0.38$, MSE = 33.18.

- Data titik (hijau) menunjukkan hubungan yang lebih kuat dibandingkan sleep, dengan garis regresi merah menunjukkan tren naik.
- MSE yang lebih tinggi menunjukkan kesalahan prediksi yang lebih besar dibandingkan sleep.

Proses perhitungan melibatkan regresi linier untuk menentukan garis terbaik, dengan R^2 (koefisien determinasi) mengukur proporsi variabilitas yang dijelaskan oleh model, dan MSE (*Mean Squared Error*) menghitung rata-rata kuadrat selisih antara nilai prediksi dan aktual.



Gambar 8. Grafik Prediksi dengan Aktual Kualitas Tidur dan Kesehatan Mental

4.3. Pembahasan

Pelaksanaan penelitian bertujuan untuk menganalisis pengaruh penggunaan media sosial terhadap kesehatan mental remaja dengan memanfaatkan algoritma *Random Forest* sebagai metode regresi, serta mengidentifikasi peran kualitas tidur dan tingkat kecanduan media sosial sebagai faktor yang memengaruhi kondisi kesehatan mental. Riset ini membangun dataset primer yang mengintegrasikan durasi penggunaan media sosial, kualitas tidur, tingkat kesehatan mental, dan tingkat kecanduan media sosial pada remaja SMA. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan data persepsi [19][27][28], penelitian ini melengkapi data kuesioner dengan bukti objektif berupa tangkapan layar durasi penggunaan aplikasi Instagram dan TikTok. *Dataset* tersebut kemudian dimanfaatkan untuk membangun model klasifikasi menggunakan algoritma *Random Forest* guna menganalisis hubungan antarvariabel. Berdasarkan hasil analisis terhadap 36 data responden yang lolos proses validasi, diperoleh beberapa temuan penting yang menjawab tujuan penelitian.

Hasil penelitian menunjukkan bahwa skor kecanduan media sosial (*Addicted_Score*) merupakan variabel yang memiliki pengaruh paling besar terhadap prediksi kualitas tidur maupun skor kesehatan mental dibandingkan rata-rata durasi penggunaan media sosial harian (*Average_Daily_Socmed*). Hal tersebut terlihat dari analisis *feature importance* pada model *Random Forest*, di mana *Addicted_Score* memiliki nilai kepentingan sekitar 0,7, sedangkan *Average_Daily_Socmed* hanya berkisar 0,2–0,3.

Analisis korelasi juga mendukung hasil tersebut. Skor kecanduan media sosial memiliki korelasi negatif sedang terhadap kesehatan mental ($r = -0,6257$), yang menunjukkan bahwa semakin tinggi tingkat kecanduan media sosial maka semakin rendah skor kesehatan mental responden. Selain itu, ditemukan hubungan negatif sedang antara skor kecanduan media sosial dengan kualitas tidur ($r = -0,4588$), sehingga peningkatan kecanduan media sosial juga berkaitan dengan menurunnya kualitas tidur.

Sementara itu, rata-rata penggunaan media sosial harian hanya menunjukkan korelasi negatif yang lemah terhadap kualitas tidur ($r = -0,1911$) maupun kesehatan mental ($r = -0,3396$). Temuan ini mengindikasikan bahwa bukan hanya lamanya waktu menggunakan media sosial yang memengaruhi kondisi psikologis remaja, melainkan tingkat ketergantungan atau kecanduan terhadap media sosial menjadi faktor yang lebih dominan.

Evaluasi model *Random Forest* menunjukkan bahwa model mampu menangkap pola hubungan antarvariabel meskipun tingkat akurasi masih terbatas. Pada prediksi kualitas tidur diperoleh nilai R^2 sebesar 0,19 dengan MSE sebesar 1,22, sedangkan pada prediksi kesehatan mental diperoleh nilai R^2 sebesar 0,38 dengan MSE sebesar 33,18. Nilai tersebut menunjukkan bahwa model lebih baik dalam memprediksi skor kesehatan mental dibandingkan kualitas tidur, walaupun masih terdapat variabel lain di luar penelitian yang turut memengaruhi kedua kondisi tersebut.

Hasil penelitian yang dilakukan sejalan dengan konsep psikologi yang menyatakan bahwa penggunaan media sosial secara berlebihan dapat meningkatkan risiko gangguan kesehatan mental melalui berbagai mekanisme, seperti munculnya *social comparison*, *fear of missing out (FoMO)*, *cyberbullying*, maupun berkurangnya interaksi sosial secara langsung. Selain itu, penggunaan media sosial pada malam hari diketahui dapat mengganggu ritme sirkadian sehingga menurunkan kualitas tidur, yang pada akhirnya berdampak pada meningkatnya risiko depresi.

Temuan kajian ini juga mendukung penelitian terdahulu yang berjudul *Social Media Dependency, Depression and Self-Esteem for Cymraeg and English-Speaking Adolescents*, yang menunjukkan adanya hubungan antara ketergantungan media sosial dengan kondisi psikologis remaja. Namun demikian, terdapat beberapa perbedaan mendasar. Penelitian sebelumnya lebih menitikberatkan pada analisis hubungan menggunakan frekuensi dan durasi penggunaan media sosial, sedangkan penelitian ini menggunakan pendekatan *machine learning* melalui algoritma *Random Forest* sehingga mampu memodelkan hubungan yang lebih kompleks antarvariabel.

Selain itu, studi yang dilakukan menambahkan variabel kualitas tidur sebagai prediktor, yang terbukti memiliki hubungan dengan kesehatan mental. Meskipun korelasi antara kualitas tidur dan skor kesehatan mental hanya sebesar 0,20 (positif lemah), hasil tersebut menunjukkan bahwa kualitas tidur tetap berkontribusi terhadap kondisi psikologis remaja ketika dianalisis bersama variabel lain dalam model *Random Forest*. Dengan demikian, penelitian ini memperluas konsep sebelumnya bahwa dampak media sosial terhadap kesehatan mental tidak hanya dipengaruhi oleh intensitas penggunaan, tetapi juga oleh tingkat kecanduan dan kualitas tidur.

Penelitian yang dilakukan memberikan kontribusi dalam bidang *data mining* dan kesehatan mental remaja melalui penerapan algoritma *Random Forest* untuk memprediksi kondisi kesehatan mental berdasarkan karakteristik perilaku penggunaan media sosial. Pendekatan ini menunjukkan bahwa teknik *machine learning* dapat digunakan sebagai alat bantu analisis dalam mengidentifikasi faktor-faktor yang berpengaruh terhadap kesehatan mental secara lebih komprehensif dibandingkan analisis statistik konvensional.

Kontribusi lainnya adalah ditemukannya fakta bahwa skor kecanduan media sosial merupakan indikator yang lebih penting dibandingkan hanya durasi penggunaan media sosial. Temuan ini memberikan implikasi bahwa upaya pencegahan gangguan kesehatan mental sebaiknya tidak hanya berfokus pada pembatasan waktu penggunaan media sosial, tetapi juga pada pengendalian perilaku adiktif terhadap media sosial.

Selain itu, hasil riset ini dapat menjadi referensi bagi sekolah, orang tua, maupun tenaga kesehatan dalam menyusun program edukasi mengenai penggunaan media sosial yang sehat, peningkatan kualitas tidur, serta deteksi dini risiko gangguan kesehatan mental pada remaja. Dari sisi akademik, pelaksanaan penelitian juga menunjukkan potensi penggunaan metode *Random Forest* sebagai alternatif dalam penelitian kesehatan mental berbasis data survei.

Masih terdapat beberapa keterbatasan dalam pelaksanaan penelitian. Pertama, jumlah data yang digunakan relatif kecil, yaitu hanya 36 responden yang memenuhi kriteria valid setelah proses penyaringan. Jumlah tersebut dapat memengaruhi kemampuan model dalam melakukan generalisasi terhadap populasi remaja yang lebih luas.

Kedua, seluruh data diperoleh melalui kuesioner *self-report* sehingga masih memungkinkan terjadinya bias persepsi maupun bias sosial dari responden. Meskipun telah disediakan pertanyaan validasi, kemungkinan adanya jawaban yang kurang merepresentasikan kondisi sebenarnya tetap tidak dapat dihilangkan sepenuhnya.

Ketiga, studi yang dilakukan hanya menggunakan beberapa variabel utama, yaitu rata-rata penggunaan media sosial, skor kecanduan media sosial, kualitas tidur, dan skor kesehatan mental. Padahal kesehatan mental remaja dipengaruhi pula oleh berbagai faktor lain seperti kondisi keluarga, prestasi akademik, dukungan sosial, tingkat stres, aktivitas fisik, status ekonomi, serta pengalaman perundungan.

Selain itu, nilai koefisien determinasi (R^2) yang masih relatif rendah menunjukkan bahwa masih terdapat banyak variabel lain yang belum dimasukkan ke dalam model. Oleh karena itu, penelitian selanjutnya disarankan menggunakan jumlah responden yang lebih besar dan lebih beragam, menambahkan variabel psikososial maupun faktor lingkungan, serta membandingkan performa *Random Forest* dengan algoritma *machine learning* lainnya seperti *XGBoost*, *Gradient Boosting*, *Support Vector Regression (SVR)*, atau *Artificial Neural Network (ANN)*. Penelitian lanjutan juga dapat menerapkan teknik validasi silang (*cross-validation*) dan optimasi *hyperparameter* yang lebih komprehensif agar diperoleh model prediksi dengan tingkat akurasi yang lebih tinggi.

5. Simpulan

Ketergantungan pada media sosial memiliki pengaruh signifikan terhadap kecenderungan depresi pada remaja di Bali, dengan algoritma *Random Forest* berbasis regresi efektif dalam mengidentifikasi hubungan kompleks antara intensitas penggunaan media sosial, kecanduan, dan kesehatan mental. Pada analisis menggunakan *Random Forest* didapatkan tingkat kepentingan *Addicted_Score* mencapai sekitar 0,7 di kedua kelompok, jauh lebih tinggi dibanding *Average_Daily_Scored* yang hanya sekitar 0,3 pada kualitas tidur dan 0,2 pada kesehatan mental. Selain itu, nilai R^2 model mental sebesar 0,38 hampir dua kali lipat lebih tinggi daripada R^2 model *sleep* sebesar 0,19, namun MSE-nya jauh lebih besar yaitu 33,18 dibandingkan 1,22. Dengan demikian, *Addicted_Score* secara konsisten menjadi fitur yang paling berpengaruh terhadap kedua variabel target. Meskipun model prediksi kesehatan mental memiliki hubungan yang lebih kuat, tingkat kesalahan prediksinya justru jauh lebih tinggi dibanding model kualitas tidur. Hasil analisis menggunakan *Random Forest* menunjukkan bahwa *Addicted Score* memiliki pengaruh yang lebih dominan dibandingkan penggunaan *social media* harian terhadap kualitas tidur dan kesehatan mental. Karena kualitas tidur dan kesehatan mental berkaitan erat dengan kondisi depresi, maka *Addicted Score* secara tidak langsung juga berpotensi menjadi faktor yang lebih berpengaruh terhadap depresi.

Daftar Referensi

- [1] S. Onie *et al.*, "Indonesia's first suicide statistics profile: an analysis of suicide and attempt rates, underreporting, geographic distribution, gender, method, and rurality," *The Lancet Regional Health - Southeast Asia*, vol. 22, p. 100368, 2024, doi: <https://doi.org/10.1016/j.lansea.2024.100368>.
- [2] J. K. N. Chan *et al.*, "All-cause and cause-specific mortality in people with depression: a large-scale systematic review and meta-analysis of relative risk and aggravating or attenuating factors, including antidepressant treatment," *World Psychiatry*, vol. 24, no. 3, pp. 404–421, Oct. 2025, doi: <https://doi.org/10.1002/wps.21354>.
- [3] World Health Organization, "Depressive disorder (depression)," 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/depression>
- [4] The Lancet Psychiatry, "Global Burden of Disease 2021: mental health messages," *Lancet Psychiatry*, vol. 11, no. 8, p. 573, 2024, doi: [https://doi.org/10.1016/S2215-0366\(24\)00222-0](https://doi.org/10.1016/S2215-0366(24)00222-0).
- [5] J. A. Naslund, A. Bondre, J. Torous, and K. A. Aschbrenner, "Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice," *J. Technol. Behav. Sci.*, vol. 5, no. 3, pp. 245–257, 2020, doi: [10.1007/s41347-020-00134-x](https://doi.org/10.1007/s41347-020-00134-x).
- [6] Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), "Survei Penetrasi dan Perilaku Pengguna Internet Indonesia 2023," 2023.
- [7] Reva Dena Mustaqina and Sri Mulyeni, "Dampak Penggunaan Media Sosial terhadap Gangguan Psikologis Mahasiswa," *RISOMA: Jurnal Riset Sosial Humaniora dan Pendidikan*, vol. 4, no. 2, pp. 09–20, Mar. 2026, doi: [10.62383/risoma.v4i2.1521](https://doi.org/10.62383/risoma.v4i2.1521).
- [8] F. Kaligis *et al.*, "The mediating role of social media disorder in the relationship between social media use and self-harm behavior in Indonesian adolescents," *Front. Public Health*, vol. Volume 13-2025, 2025, doi: [10.3389/fpubh.2025.1663729](https://doi.org/10.3389/fpubh.2025.1663729).
- [9] D. Safitri and Sujarwo, "Eksplorasi Dampak Penggunaan Media Sosial terhadap Kesehatan Mental Remaja," *Triwikrama: Jurnal Ilmu Sosial*, vol. 9, no. 11, pp. 201–211, 2025, doi: [10.9963/6awwk344](https://doi.org/10.9963/6awwk344).

- [10] R. Jones, I. Reppa, and P. Reed, "Social media dependency, depression and self-esteem for Cymraeg and English-speaking adolescents," *Behaviour & Information Technology*, vol. 0, no. 0, pp. 1–14, 2025, doi: 10.1080/0144929X.2025.2455395.
- [11] B. Han and C. Myers, "Perceptions of overuse, underuse, and change of use of a social media site: definition, measurement instrument, and their managerial impacts," *Behaviour & Information Technology*, vol. 37, no. 3, pp. 247–257, 2018, doi: 10.1080/0144929X.2018.1432687.
- [12] G. Brandenburg, P. Ozimek, H.-W. Bierhoff, and C. Janker, "The relation between use intensity of private and professional SNS, social comparison, self-esteem, and depressive tendencies in the light of self-regulation," *Behaviour & Information Technology*, vol. 38, no. 6, pp. 578–591, 2019, doi: 10.1080/0144929X.2018.1545049.
- [13] Y.-K. Lee, C.-T. Chang, Y. Lin, and Z.-H. Cheng, "The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress," *Comput. Human Behav.*, vol. 31, pp. 373–383, 2014, doi: <https://doi.org/10.1016/j.chb.2013.10.047>.
- [14] K. Saha, J. Torous, E. D. Caine, and M. De Choudhury, "Psychosocial Effects of the COVID-19 Pandemic: Large-scale Quasi-Experimental Study on Social Media," *J Med Internet Res*, vol. 22, no. 11, p. e22600, 2020, doi: 10.2196/22600.
- [15] J. C. Eichstaedt *et al.*, "Facebook language predicts depression in medical records," *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11203–11208, 2018, doi: 10.1073/pnas.1802331115.
- [16] C. Qi, "A double-edged sword? Exploring the impact of students' academic usage of mobile devices on technostress and academic performance," *Behaviour & Information Technology*, vol. 38, no. 12, pp. 1337–1354, 2019, doi: 10.1080/0144929X.2019.1585476.
- [17] J. M. Twenge and W. K. Campbell, "Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study," *Prev. Med. Rep.*, vol. 12, pp. 271–283, 2018, doi: <https://doi.org/10.1016/j.pmedr.2018.10.003>.
- [18] M. Walker, *Why We Sleep: Unlocking the Power of Sleep and Dreams*. New York: Scribner, 2017.
- [19] D. J. Yu, Y. K. Wing, T. M. H. Li, and N. Y. Chan, "The Impact of Social Media Use on Sleep and Mental Health in Youth: a Scoping Review," *Curr. Psychiatry Rep.*, vol. 26, no. 3, pp. 104–119, 2024, doi: 10.1007/s11920-024-01481-9.
- [20] J. Wang, N. Wang, P. Liu, and Y. Liu, "Social network site addiction, sleep quality, depression and adolescent difficulty describing feelings: a moderated mediation model," *BMC Psychol.*, vol. 13, no. 1, p. 57, 2025, doi: 10.1186/s40359-025-02372-1.
- [21] S. Han, B. Williamson, and Y. Fong, "Improving random forest predictions in small datasets from two-phase sampling designs," *BMC Med. Inform. Decis. Mak.*, vol. 21, Sep. 2021, doi: 10.1186/s12911-021-01688-3.
- [22] N. Novianti, S. Alkadri, and I. Fakhruzi, "Klasifikasi Penyakit Hipertensi Menggunakan Metode Random Forest," *Progresif: Jurnal Ilmiah Komputer*, vol. 20, no. 1, pp. 380–392, 2024, [Online]. Available: <https://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/1663>
- [23] Y. Liu, Y. Pan, F. Wang, W. Li, L. Gao, and G. Zhao, "Multivariate machine learning regression approaches to predict adolescent suicide risk," *Psychiatry Res.*, vol. 361, p. 117160, 2026, doi: <https://doi.org/10.1016/j.psychres.2026.117160>.
- [24] C. Yu, X. Kong, W. Yu, X. Ni, J. Chen, and X. Liao, "Machine learning models for predicting the risk of depressive symptoms in Chinese college students," *Front. Psychiatry*, vol. Volume 16-2025, 2025, doi: 10.3389/fpsy.2025.1648585.
- [25] D. A. Fife and J. D'Onofrio, "Common, uncommon, and novel applications of random forest in psychological research," *Behav. Res. Methods*, vol. 55, no. 5, pp. 2447–2466, 2023, doi: 10.3758/s13428-022-01901-9.
- [26] E. J. Kim, S. K. Kim, S. H. Jung, and Y. S. Ryu, "Predictive factors of adolescents' happiness: a random forest analysis of the 2023 Korea Youth Risk Behavior Survey," *Child Health Nurs Res*, vol. 31, no. 2, pp. 85–95, Apr. 2025, doi: 10.4094/chnr.2024.049.
- [27] O. Ahmed, E. I. Walsh, A. Dawel, K. Alateeq, D. A. Espinoza Oyarce, and N. Cherbuin, "Social media use, mental health and sleep: A systematic review with meta-analyses," *J. Affect. Disord.*, vol. 367, pp. 701–712, 2024, doi: <https://doi.org/10.1016/j.jad.2024.08.193>.

-
- [28] W. Chalermchutidej, B. Manaboriboon, G. Sanpawitayakul, S. Theppiban, and S. In-iw, "Sleep, social media use and mental health in female adolescents aged 12 to 18 years old during the COVID-19 pandemic," *BMC Pediatr.*, vol. 23, no. 1, p. 398, 2023, doi: 10.1186/s12887-023-04218-4.