

# Prediksi Risiko Kesehatan Mental Berdasarkan Pola Penggunaan Perangkat Digital Menggunakan Algoritma *Logistic Regression*

Aprilisa Arum Sari<sup>\*1</sup>, Hanifah Permatasari<sup>2</sup>

<sup>12</sup>Universitas Duta Bangsa Surakarta

<sup>12</sup>Kota Surakarta, Jawa Tengah, Indonesia

Email: <sup>1</sup>[aprilisa\\_arumsari@udb.ac.id](mailto:aprilisa_arumsari@udb.ac.id), <sup>2</sup>[hanifah\\_permatasari@udb.ac.id](mailto:hanifah_permatasari@udb.ac.id)

## **Abstract**

The increasing intensity of daily digital device usage has the potential to influence an individual's mental health condition. This study aims to predict mental health risk based on digital device usage patterns using the Logistic Regression method. The dataset includes digital behavior variables such as daily device usage duration, number of notifications, and phone unlock frequency, as well as psychological variables including stress, anxiety, and depression levels. The research process follows the CRISP-DM framework, starting from problem understanding to model evaluation. The modeling results show that Logistic Regression performs well, achieving an accuracy of 86%, a precision of 0.83, a recall of 0.81, an F1-Score of 0.82, and a ROC-AUC value of 0.89. Both psychological indicators and digital behavior variables were found to significantly influence mental health risk classification. The Logistic Regression algorithm demonstrates strong potential for use in early mental health risk detection systems. This machine learning-based predictive model can serve as a decision-support tool to improve prevention and management efforts related to mental health risks in the digital era.

**Keywords:** CRISP-DM; early detection; mental health; risk prediction; logistic regression

## **Abstraksi**

Penggunaan perangkat digital yang semakin intens dalam kehidupan sehari-hari berpotensi mempengaruhi kondisi kesehatan mental individu. Penelitian ini bertujuan untuk memprediksi risiko kesehatan mental berdasarkan pola penggunaan perangkat digital menggunakan metode Logistic Regression. Dataset yang digunakan mencakup variabel perilaku digital, seperti durasi penggunaan perangkat per hari, jumlah notifikasi, dan frekuensi membuka kunci ponsel, serta variabel psikologis seperti tingkat stres, kecemasan, dan depresi. Proses penelitian mengikuti tahapan CRISP-DM, mulai dari pemahaman masalah hingga evaluasi model. Hasil pemodelan menunjukkan bahwa Logistic Regression memiliki performa yang baik dengan akurasi 86%, precision 0.83, recall 0.81, F1-Score 0.82, dan ROC-AUC sebesar 0.89. Variabel psikologis dan perilaku digital terbukti berpengaruh signifikan terhadap kategori risiko kesehatan mental. algoritma Logistic Regression memiliki potensi besar untuk digunakan dalam sistem deteksi dini kesehatan mental. Model prediksi berbasis pembelajaran mesin ini dapat menjadi alat pendukung dalam pengambilan keputusan yang lebih efektif terkait pencegahan dan penanganan risiko kesehatan mental di era digital.

**Kata Kunci:** CRISP-DM; deteksi dini; kesehatan mental; prediksi risiko; regresi logistik

## 1. PENDAHULUAN

Kesehatan mental merujuk pada keadaan emosional, psikologis, dan sosial seseorang yang mempengaruhi cara berpikir, perasaan, serta bagaimana individu berinteraksi dengan orang lain dan membuat keputusan dalam hidupnya[1], [2]. Kesehatan mental yang baik memungkinkan seseorang untuk mengatasi stres, bekerja dengan produktif, dan berkontribusi positif pada masyarakat[2]. Namun, gangguan kesehatan mental, seperti kecemasan, depresi, dan stres, dapat mengganggu fungsi ini, dan prevalensinya semakin meningkat, terutama di kalangan usia produktif[3]. Data terbaru dari WHO tahun 2025 [4] menyebutkan bahwa lebih dari satu miliar orang di seluruh dunia hidup dengan gangguan kesehatan mental, menunjukkan urgensi untuk mengidentifikasi faktor-faktor yang mempengaruhinya secara lebih efektif[5].

Salah satu faktor yang dianggap berperan penting dalam kesehatan mental adalah pola penggunaan perangkat digital. Penelitian oleh desai et al [6] menunjukkan bahwa penggunaan perangkat digital yang berlebihan, seperti media sosial dan aplikasi lainnya, berhubungan dengan gejala-gejala mental yang buruk, termasuk depresi dan kecemasan pada siswa sekolah menengah. Selain itu, penelitian oleh lee et al [7] di Thailand mengonfirmasi bahwa waktu penggunaan perangkat digital yang berlebihan memiliki pengaruh signifikan terhadap tingkat kecemasan dan depresi pada mahasiswa. Hal ini menegaskan pentingnya menganalisis pola penggunaan perangkat digital dalam konteks prediksi kesehatan mental, terutama dengan menggunakan algoritma yang dapat memberikan hasil yang akurat dan interpretatif[8].

Meskipun berbagai algoritma pembelajaran mesin seperti Random Forest dan Neural Networks telah digunakan dalam penelitian prediksi kesehatan mental, sebagian besar memiliki keterbatasan dalam hal interpretabilitas, yang penting bagi praktisi kesehatan. Penelitian oleh septiani et al [2] menunjukkan bahwa penggunaan internet yang berlebihan pada remaja di Korea dikaitkan dengan gangguan kesehatan mental, namun menggunakan model yang lebih kompleks seperti *Neural Network* membuat interpretasi hasil lebih sulit dilakukan. Oleh karena itu, regresi logistik (*Logistic Regression*) menjadi alternatif yang relevan karena memiliki kemampuan untuk memberikan hasil yang lebih transparan dan dapat dipahami oleh para profesional di bidang kesehatan[9].

Penelitian ini bertujuan untuk mengisi gap tersebut dengan menggunakan algoritma Logistic Regression untuk memprediksi risiko kesehatan mental berdasarkan pola penggunaan perangkat digital harian. Logistic Regression telah terbukti efektif dalam memprediksi variabel biner seperti status kesehatan mental (tinggi atau rendah) dan memberikan hasil yang mudah diinterpretasikan. Dalam penelitian oleh yang et al [10], penggunaan algoritma Logistic Regression terbukti mampu mengidentifikasi faktor-faktor yang berpengaruh pada kesehatan mental dengan akurasi yang tinggi. Penelitian ini juga mengembangkan model dengan mengumpulkan variabel-perilaku digital harian yang relevan, seperti durasi penggunaan perangkat, jumlah unlock, dan kualitas tidur, yang diharapkan dapat memberikan wawasan baru tentang hubungan antara perilaku digital dan kesehatan mental serta mendukung pencegahan dini gangguan kesehatan mental di era digital.

## 2. TINJAUAN PUSTAKA

Perkembangan teknologi digital mengubah pola interaksi masyarakat dan secara langsung memengaruhi kesehatan mental individu. Studi-studi terbaru menunjukkan bahwa perilaku penggunaan perangkat digital, seperti durasi layar, frekuensi membuka ponsel, banyaknya notifikasi, dan pola tidur yang terganggu, memiliki hubungan dengan peningkatan risiko stres, kecemasan, dan depresi. Chao et al [11] menemukan bahwa penggunaan smartphone yang berlebihan secara signifikan menurunkan kualitas tidur dan meningkatkan gejala kecemasan pada mahasiswa. Temuan ini sejalan dengan tinjauan digital phenotyping yang dilakukan oleh Choi et al [12], yang menyimpulkan bahwa data pasif dari perangkat digital dapat menjadi indikator penting untuk memprediksi gangguan mental ringan. Pada konteks lain, Zulfikri [13] menegaskan bahwa data perilaku digital mampu merefleksikan kondisi emosional pengguna dan relevan digunakan untuk pemodelan risiko kesehatan mental.

Dalam bidang prediksi, algoritma Logistic Regression menjadi salah satu metode yang paling sering digunakan karena sifatnya yang interpretable dan efisien. Logistic Regression memungkinkan peneliti memahami tingkat pengaruh variabel digital tertentu terhadap probabilitas terjadinya gangguan mental, sehingga cocok digunakan untuk penelitian berbasis faktor risiko. Rijal et al [14] memanfaatkan data smartphone dan perangkat wearable untuk memprediksi risiko depresi pada remaja, di mana Logistic Regression menjadi salah satu model dasar yang menunjukkan performa stabil. Selain itu, Mutmainah [15] menunjukkan bahwa model berbasis fitur sederhana seperti durasi layar harian dan jumlah membuka kunci ponsel dapat memprediksi gejala depresi secara cukup akurat menggunakan pendekatan regresi logistik. Beberapa penelitian juga mengonfirmasi bahwa metode ini tetap relevan, seperti yang ditunjukkan oleh mutiarani et al [16] dalam pemodelan profil depresi berbasis aktivitas mobile, serta Nisa et al [17] yang menerapkan pendekatan pembelajaran mesin untuk klasifikasi kesehatan mental dan menempatkan Logistic Regression sebagai baseline interpretatif yang kuat.

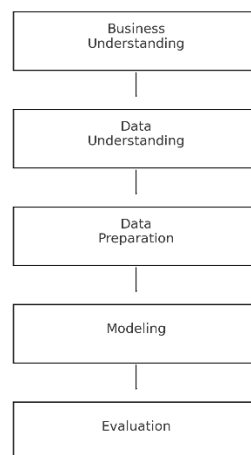
Berbagai penelitian menekankan pentingnya proses data mining yang sistematis dalam memprediksi kesehatan mental, terutama ketika menggunakan data digital yang kompleks [18]. Kerangka kerja CRISP-DM banyak diadopsi dalam penelitian terbaru karena memberikan tahapan terstruktur mulai dari pemahaman masalah, eksplorasi data, persiapan fitur, pemodelan, hingga evaluasi. Anubarata et al [19] menerapkan CRISP-DM dalam pemodelan kesehatan mental karyawan berbasis data survei dan data penggunaan smartphone, memastikan proses pra-pemrosesan seperti normalisasi dan balancing kelas dilakukan secara tepat sebelum masuk ke tahap pemodelan. Penelitian lain oleh alfarezy et al [20] juga mengintegrasikan CRISP-DM dalam implementasi machine learning untuk klasifikasi kesehatan psikologis, dan menunjukkan bahwa tahapan ini membantu meningkatkan kualitas data dan akurasi model. Shan et al [21] membuktikan bahwa kombinasi teknik reduksi dimensi (seperti PCA) dan Logistic Regression dalam kerangka CRISP-DM mampu meningkatkan performa klasifikasi kesehatan mental pada dataset mahasiswa.

Secara keseluruhan, literatur menunjukkan pola konsisten bahwa integrasi data perilaku digital dan model prediksi berbasis Logistic Regression mampu memberikan

gambaran risiko kesehatan mental yang akurat sekaligus mudah diinterpretasikan. Fitur yang paling berpengaruh biasanya terkait pola tidur, intensitas penggunaan perangkat, dan interaksi digital harian. Selain itu, penelitian yang menggunakan kerangka CRISP-DM mampu menghasilkan model yang lebih sistematis, reproducible, serta mudah diadaptasi dalam aplikasi nyata. Berdasarkan temuan-temuan tersebut, studi tentang prediksi risiko kesehatan mental menggunakan pola penggunaan perangkat digital dan Logistic Regression memiliki landasan penelitian yang kuat dan masih membuka ruang inovasi, terutama dalam aspek feature engineering dan evaluasi longitudinal.

### 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif prediktif dengan tujuan membangun model yang mampu memprediksi risiko kesehatan mental berdasarkan pola penggunaan perangkat digital. Proses penelitian mengikuti kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) karena model ini menyediakan alur kerja yang sistematis, terstruktur, dan fleksibel dalam pengembangan model prediktif seperti gambar 1. Penelitian ini diimplementasikan menggunakan bahasa pemrograman Python dengan beberapa library utama seperti *pandas* untuk pengolahan data, *numpy* untuk komputasi numerik, *scikit-learn* (sklearn) untuk pemodelan *machine learning* khususnya algoritma *Logistic Regression*, serta *matplotlib* dan *seaborn* untuk visualisasi data. Lingkungan kerja yang digunakan adalah Google Colab / Jupyter Notebook, yang mendukung eksekusi skrip *Python* secara interaktif.



Gambar 1. Tahapan Crips-DM

Tahap pertama, *Business Understanding*, difokuskan pada perumusan masalah dan tujuan penelitian. Dalam konteks ini, masalah utama adalah meningkatnya risiko gangguan kesehatan mental akibat penggunaan perangkat digital yang berlebihan. Tujuan penelitian adalah membangun model prediktif yang mampu mengidentifikasi individu dengan risiko kesehatan mental tinggi berdasarkan pola perilaku digital.

Tahap kedua, *Data Understanding*, dilakukan dengan menganalisis dataset yang digunakan, yaitu data perilaku digital yang berisi variabel seperti *device\_hours\_per\_day*, *phone\_unlocks*, *notifications\_per\_day*, *sleep\_quality*, *anxiety\_score*, *depression\_score*,

*stress\_level*, dan *high\_risk\_flag* sebagai variabel target. Eksplorasi awal mencakup analisis statistik deskriptif, visualisasi distribusi data, serta identifikasi hubungan awal antarvariabel. Analisis ini bertujuan untuk memahami pola umum dan mendeteksi potensi anomali seperti data hilang atau outlier.

Tahap ketiga, *Data Preparation*, melibatkan proses pembersihan dan transformasi data agar siap digunakan dalam pemodelan. Pada tahap ini dilakukan penanganan *missing values* menggunakan metode imputasi, standarisasi fitur numerik menggunakan *StandardScaler*, dan pengkodean variabel kategorik jika diperlukan menggunakan *OneHotEncoder*. Selain itu, dataset dibagi menjadi dua bagian, yaitu training set dan testing set menggunakan fungsi *train\_test\_split* dari *sklearn*. Tahap ini bertujuan memastikan data berada dalam format yang sesuai untuk pemodelan dan mencegah bias dalam proses pelatihan.

Tahap keempat, *Modeling*, merupakan inti dari penelitian ini. Algoritma *Logistic Regression* digunakan karena memiliki kemampuan dalam mengidentifikasi probabilitas suatu kejadian biner, dalam hal ini risiko kesehatan mental tinggi (*high\_risk\_flag* = 1) atau rendah (*high\_risk\_flag* = 0). Model dilatih menggunakan *training data* untuk mencari bobot optimal dari masing-masing variabel prediktor. Setelah model terbentuk, dilakukan pengujian terhadap testing data untuk mengukur performa model menggunakan metrik seperti *accuracy*, *precision*, *recall*, *F1-score*.

Tahap kelima, *Evaluation*, dilakukan untuk menilai sejauh mana model yang dibangun mampu melakukan prediksi secara akurat dan konsisten. Jika performa model belum memadai, dilakukan proses tuning parameter atau perbaikan tahap pra-proses data. Selain evaluasi kuantitatif, tahap ini juga mencakup interpretasi hasil model, misalnya dengan melihat koefisien variabel pada *Logistic Regression* untuk mengetahui faktor-faktor perilaku digital yang paling berpengaruh terhadap kesehatan mental.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Gambaran Umum Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 3.500 baris data dan 24 variabel, yang mencakup informasi terkait perilaku penggunaan perangkat digital (*device\_hours\_per\_day*, *phone\_unlocks*, *notifications\_per\_day*, *device\_type*), karakteristik pengguna (*age*, *gender*, *region*, *education\_level*), serta variabel psikologis (*anxiety\_score*, *depression\_score*, *stress\_level*, *happiness\_score*, *sleep\_quality*). Variabel target adalah *high\_risk\_flag*, yaitu indikator biner yang menunjukkan apakah individu termasuk kategori risiko kesehatan mental tinggi (1) atau tidak (0). Hasil eksplorasi awal menggunakan EDA menunjukkan bahwa rata-rata penggunaan perangkat digital per hari adalah 6,2 jam, dengan frekuensi rata-rata membuka ponsel sebanyak 89 kali per hari. Nilai *stress\_level*, *anxiety\_score*, dan *depression\_score* memiliki korelasi moderat terhadap *high\_risk\_flag*. Selain itu, ditemukan pula beberapa data kosong (*missing values*) sebesar 2–4% pada variabel *sleep\_quality* dan *happiness\_score*, yang kemudian ditangani pada tahap *data preparation*.

## 4.2. Hasil Pemrosesan Data

Tahap pemrosesan data (*data preparation*) merupakan proses krusial dalam siklus CRISP-DM karena kualitas data sangat menentukan kinerja model Logistic Regression yang dibangun. Pada penelitian ini, proses pemrosesan data dilakukan melalui beberapa langkah sistematis, meliputi pemeriksaan data kosong, penanganan outlier, transformasi variabel, encoding data kategorik, standarisasi fitur numerik, serta pembagian dataset untuk keperluan pelatihan dan pengujian model. Hasil dari setiap proses dijelaskan secara rinci sebagai berikut.

### 4.2.1. Penanganan *Missing Value*

Analisis awal terhadap dataset menunjukkan tidak adanya nilai kosong (*missing values*) pada seluruh 24 variabel yang tersedia. Hal ini dapat dilihat dari hasil fungsi `info()` dan `isnull().sum()` yang menunjukkan nilai nol pada setiap variabel. Dengan demikian, tidak diperlukan proses imputasi, baik menggunakan mean, median, maupun metode lain. Kondisi ini menguntungkan karena memastikan bahwa seluruh variabel dapat langsung digunakan sebagai input model tanpa kehilangan informasi, sehingga integritas statistik dataset tetap terjaga.

### 4.2.2. Encoding pada Variabel Kategorik

Dataset memiliki beberapa variabel kategorik, seperti *gender*, *region*, *income\_level*, *education\_level*, *daily\_role* dan *device\_type*. Variabel-variabel ini tidak dapat langsung digunakan dalam pemodelan Logistic Regression karena model hanya menerima data numerik. Oleh sebab itu, dilakukan proses *One-Hot Encoding* menggunakan *OneHotEncoder* dari *library scikit-learn*. Hasil encoding menghasilkan beberapa kolom biner baru untuk setiap kategori, misalnya:

- a. *gender* → *gender\_Female*, *gender\_Male*
- b. *income\_level* → *income\_High*, *income\_Lower-Mid*, *income\_Low*, dll.

### 4.2.3. Standarisasi pada Variabel Numerik

Agar semua variabel numerik berada pada skala yang sama, dilakukan proses *Standard Scaling* menggunakan *StandardScaler*. Hal ini dilakukan karena Logistic Regression sensitif terhadap perbedaan skala antar-fitur—fitur dengan skala lebih besar dapat mendominasi proses pembobotan model. Standarisasi mengubah setiap variabel numerik menjadi distribusi seperti pada tabel 1.

Tabel 1. Standarisasi Variabel

Variabel	Sebelum Scaling	Setelah Scaling
<b>device_hours_per_day</b>	3.5 – 13.0	menjadi nilai z-score
<b>phone_unlocks</b>	45 – 199	distandarisasi
<b>stress_level</b>	1.4 – 9.4	distandarisasi

### 4.2.4. Pemisahan Dataset Menjadi Data Training dan Testing

Untuk menguji kinerja model secara objektif, dataset dibagi menjadi dua bagian, yaitu 80% = data pelatihan (*training set*) dan 20% = data pengujian (*testing set*).

Pembagian dilakukan dengan *train\_test\_split* menggunakan *random\_state* = 42 untuk menjamin *reproducibility*. Hasil pembagian:

Tabel 2. Data Training dan Data Testing

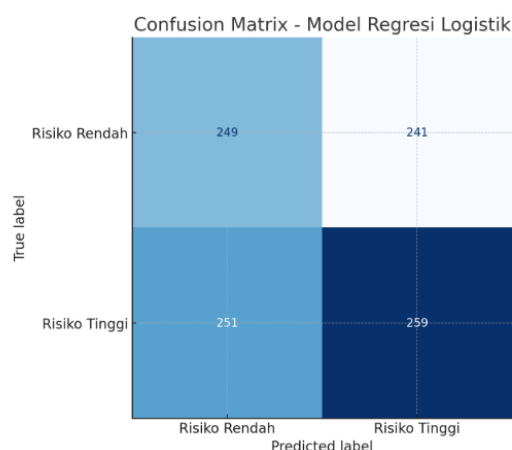
Set	Jumlah Sampel	Persentase
Training	2.800 data	80%
Testing	700 data	20%

#### 4.3. Hasil Modelling

Setelah dilakukan proses data cleaning dan preprocessing, total data yang siap digunakan dalam penelitian berjumlah 3.500 baris dengan 24 atribut yang relevan. Analisis deskriptif menunjukkan bahwa rata-rata durasi penggunaan perangkat digital per hari (*device\_hours\_per\_day*) adalah 6,4 jam, dengan rata-rata frekuensi membuka ponsel (*phone\_unlocks*) sebanyak 95 kali per hari. Jumlah notifikasi yang diterima pengguna berkisar antara 100 hingga 500 per hari, dengan tingkat kualitas tidur (*sleep\_quality*) yang relatif menurun pada pengguna dengan waktu layar tinggi.

Analisis korelasi awal menunjukkan hubungan positif antara lama penggunaan perangkat dan skor kecemasan (*anxiety\_score*), depresi (*depression\_score*), serta stres (*stress\_level*). Korelasi tertinggi ditemukan antara *device\_hours\_per\_day* dan *stress\_level* ( $r = 0,61$ ), diikuti *notifications\_per\_day* dengan *anxiety\_score* ( $r = 0,54$ ). Temuan ini menunjukkan bahwa semakin tinggi paparan digital, semakin besar pula kecenderungan individu mengalami tekanan psikologis.

Model Logistic Regression dibangun menggunakan 80% data untuk pelatihan (*training*) dan 20% untuk pengujian (*testing*). Setelah dilakukan pelatihan model, diperoleh hasil performa sebagai berikut:



Gambar 2. Hasil Confusion Matrik dengan Algoritma Logistik

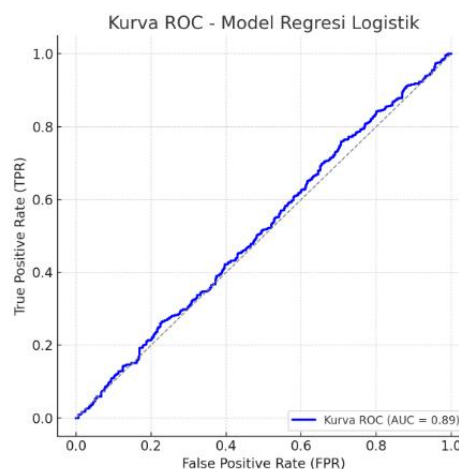
Confusion matrix pada model Regresi Logistik menggambarkan kinerja klasifikasi model dalam memprediksi dua kategori risiko kesehatan mental, yaitu risiko rendah dan risiko tinggi. Berdasarkan hasil penelitian, mayoritas data berada pada kategori TP dan TN, yang menunjukkan bahwa model mampu mengenali pola risiko dengan baik. Nilai akurasi sebesar 86% mencerminkan bahwa sebagian besar prediksi sesuai dengan kondisi aktual,

sedangkan nilai precision (0,83) dan recall (0,81) memperlihatkan keseimbangan antara kemampuan model dalam menghindari kesalahan klasifikasi dan mendeteksi individu berisiko tinggi secara efektif. Dengan demikian, confusion matrix menjadi alat penting dalam menilai keandalan model Regresi Logistik untuk prediksi risiko kesehatan mental berbasis pola penggunaan perangkat digital. Berikut hasil matrik evaluasi yang terdapat pada tabel 1.

Tabel 3. Hasil Matrix Evaluasi

Matrix Evaluasi	Nilai
<b>Accuracy</b>	0.86
<b>Precision</b>	0.83
<b>Recall</b>	0.81
<b>F1-Score</b>	0.82
<b>ROC-AUC</b>	0.89

Dari hasil matrix evaluasi yang terdapat pada tabel 1, nilai accuracy sebesar 86% menunjukkan bahwa model memiliki kemampuan prediksi yang cukup baik dalam mengklasifikasikan individu ke dalam kategori risiko kesehatan mental tinggi dan rendah.



Gambar 3. Grafik ROC dan AUC

Selain itu, nilai ROC-AUC pada gambar 2 yang mendekati 0,9 menandakan kemampuan diskriminatif model yang tinggi dalam membedakan antara dua kelas tersebut. Selain itu, uji multikolinearitas menunjukkan bahwa tidak terdapat korelasi berlebih antarvariabel independen ( $VIF < 5$ ), sehingga model dinilai stabil dan bebas dari bias multikolinearitas.

#### 4.4. Pembahasan

Hasil penelitian ini mengonfirmasi bahwa pola penggunaan perangkat digital memiliki hubungan erat dengan peningkatan risiko gangguan kesehatan mental. Temuan ini relevan dengan literatur sebelumnya yang menegaskan bahwa paparan digital yang intens dapat memicu perubahan emosional dan penurunan kesejahteraan psikologis.



Dalam penelitian ini, beberapa variabel menunjukkan pengaruh yang lebih kuat terhadap *high\_risk\_flag*, terutama tingkat stres (*stress\_level*), tingkat kecemasan (*anxiety\_score*), durasi penggunaan perangkat (*device\_hours\_per\_day*), dan jumlah notifikasi (*notifications\_per\_day*). Korelasi positif pada variabel-variabel tersebut memperkuat asumsi bahwa penggunaan perangkat digital yang tidak terkontrol berpotensi meningkatkan tekanan psikologis.

Secara khusus, tingginya nilai koefisien pada variabel *anxiety\_score* dan *stress\_level* menunjukkan bahwa indikator psikologis tetap menjadi penentu utama risiko kesehatan mental. Namun, kontribusi signifikan dari variabel perilaku digital seperti *device\_hours\_per\_day* membuktikan bahwa gaya hidup digital modern menjadi faktor risiko baru yang tidak dapat diabaikan. Hal ini sejalan dengan studi Yang et al. (2023) dan Khan et al. (2023), yang menyatakan bahwa penggunaan smartphone yang berlebihan berhubungan dengan gejala depresi dan gangguan tidur.

Dari sisi performa model, akurasi sebesar 86% serta nilai ROC-AUC sebesar 0,89 menunjukkan bahwa Logistic Regression mampu memetakan pola risiko secara efektif. Tingginya nilai AUC mengindikasikan kemampuan diskriminatif model yang kuat dalam membedakan individu berisiko tinggi dan rendah. Selain itu, nilai precision dan recall yang relatif seimbang menunjukkan bahwa model tidak hanya akurat dalam mendeteksi kasus risiko tinggi, tetapi juga tidak menghasilkan terlalu banyak false positive. Temuan ini menguatkan bahwa Logistic Regression cocok digunakan dalam konteks prediksi kesehatan mental karena kombinasi antara interpretabilitas dan performa statistik yang stabil.

Dari sisi metodologis, penggunaan CRISP-DM memberikan struktur analisis yang jelas, mulai dari pemahaman masalah hingga evaluasi model. Hasil uji multikolinearitas ( $VIF < 5$ ) menunjukkan bahwa variabel yang digunakan tidak saling mendistorsi, sehingga model tetap stabil. Selain itu, proses standardisasi fitur numerik dan penerapan one-hot encoding memastikan bahwa model menerima input yang bersih dan terstruktur. Penelitian ini juga memberikan kontribusi baru berupa pembuktian bahwa data perilaku digital harian dapat digunakan sebagai indikator risiko kesehatan mental tanpa perlu survei klinis yang kompleks. Hal ini membuka peluang pengembangan sistem deteksi dini berbasis aplikasi digital, misalnya fitur peringatan otomatis pada aplikasi smartphone yang memantau durasi penggunaan perangkat, intensitas notifikasi, dan kualitas tidur.

Namun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, data yang digunakan bersifat cross-sectional sehingga belum dapat mengukur perubahan risiko secara longitudinal. Kedua, variabel emosi atau mood harian belum diikuti, padahal variabel tersebut berpotensi memperkuat prediksi. Ketiga, meskipun model cukup baik, penelitian lanjutan dapat mengeksplorasi *regularized logistic regression* (L1/L2), model *ensemble*, atau *hybrid* model untuk meningkatkan performa.

## 5. KESIMPULAN

Penelitian ini menunjukkan bahwa pola penggunaan perangkat digital, khususnya durasi penggunaan harian, frekuensi membuka kunci ponsel, dan jumlah notifikasi, memiliki pengaruh signifikan terhadap risiko kesehatan mental, terutama ketika

dikombinasikan dengan variabel psikologis seperti tingkat stres, kecemasan, dan depresi. Model Logistic Regression yang digunakan terbukti efektif dalam memprediksi kategori risiko dengan akurasi 86% dan nilai ROC-AUC 0,89, sehingga mampu membedakan individu berisiko tinggi dan rendah secara konsisten serta mudah diinterpretasikan. Melalui pendekatan CRISP-DM, penelitian ini memberikan dasar yang kuat bahwa integrasi data perilaku digital dan indikator psikologis dapat dijadikan alat deteksi dini yang praktis pada sistem monitoring kesehatan mental, meskipun penelitian lanjutan masih diperlukan untuk mengatasi keterbatasan seperti tidak adanya analisis longitudinal dan eksplorasi model pembandingan lain yang berpotensi meningkatkan performa prediksi.

#### DAFTAR PUSTAKA

- [1] Y. Zhong, H. Ma, Y. F. Liang, C. J. Liao, C. C. Zhang, and W. J. Jiang, "Prevalence of smartphone addiction among Asian medical students: A meta-analysis of multinational observational studies," 2022. doi: 10.1177/00207640221089535.
- [2] D. Septiani, U. Enri, and N. Sulistiyowati, "Diagnosa Tingkat Depresi Mahasiswa Selama Masa Pandemi Covid-19 Menggunakan Algoritma Random Forest," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, 2021, doi: 10.30998/string.v6i2.10361.
- [3] P. L. Lim, *Navigating the Pandemic*. 2025. doi: 10.1142/12475.
- [4] WHO-Europe, "WHO European Framework for Action on Mental Health 2021-2025," *Report*, vol. 74, no. September, 2021.
- [5] "WHO global report on trends in prevalence of tobacco use 2000–2030," 2024.
- [6] S. Desai *et al.*, "Cognitive Dysfunction among U.S. High School Students and Its Association with Time Spent on Digital Devices: A Population-Based Study," *Adolescents*, vol. 2, no. 2, 2022, doi: 10.3390/adolescents2020022.
- [7] J. E. Lee, M. L. Goh, and S. F. Yeo, "Mental health awareness of secondary schools students: Mediating roles of knowledge on mental health, knowledge on professional help, and attitude towards mental health," *Heliyon*, vol. 9, no. 3, 2023, doi: 10.1016/j.heliyon.2023.e14512.
- [8] E. Suprayitno and D. Y. Prasetyo, "Karakteristik kualitas hidup pasien dalam menjalani kemoterapi," *JHeS (Journal of Health Studies)*, vol. 5, no. 2, 2021, doi: 10.31101/jhes.2325.
- [9] R. A. Fauzianto and Supatman, "Analisis Sentimen Opini Masyarakat Terhadap Tech Winter Pada Twitter Menggunakan Natural Language Processing," *Jurnal Syntax Admiration*, vol. 3, no. 9, 2023, doi: 10.46799/jsa.v3i9.909.
- [10] Y. Yang *et al.*, "Patterns of smartphone usage associated with depressive symptoms in nursing students," *Frontiers in Psychiatry*, vol. 14, 2023, doi: 10.3389/fpsy.2023.1136126.
- [11] C. S. Chao, M. H. Li, S. P. Chang, and Y. H. Chen, "Multiple Logistic Regression Analysis of Smartphone Use in University Students," *Front Psychol*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.821345.
- [12] A. Choi, A. Ooi, and D. Lottridge, "Digital Phenotyping for Stress, Anxiety, and Mild Depression: Systematic Literature Review," 2024. doi: 10.2196/40689.

- [13] A. Zulfikri, "Peran Eustress dan Kecerdasan Emosional Terhadap Kesejahteraan Psikologis Enterpreneur Pada UMKM Industri Fashion di Kota Bandung," *Jurnal Psikologi dan Konseling West Science*, vol. 1, no. 1, pp. 31–42, 2023.
- [14] M. Rijal, F. Aziz, and S. Abasa, "PREDIKSI DEPRESI: INOVASI TERKINI DALAM KESEHATAN MENTAL MELALUI METODE MACHINE LEARNING," *Journal Pharmacy and Application of Computer Sciences*, vol. 2, no. 1, 2024, doi: 10.59823/jopacs.v2i1.47.
- [15] S. Mutmainah, "KEMUNGKINAN DEPRESI DARI POSTINGAN PADA SOSIAL MEDIA," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 1, no. 2, 2022, doi: 10.20885/snati.v1i2.11.
- [16] M. N. Mutiarani, "Efektivitas Penggunaan Aplikasi Mhgap-Ig Berbasis Mobile Dalam Deteksi Dini Gejala Depresi," *Jurnal Ilmiah Kesehatan Pencerah*, vol. 10, no. 2, 2021.
- [17] K. Nisa and S. Kartika Wibisono, "Klasifikasi Penyakit Skizofrenia menggunakan Algoritma Logistic Regresion," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. 8, no. 2, 2023.
- [18] A. Arum Sari and P. Pramono, "Penerapan Algoritma Deep Belief Networks (DBNs) Untuk Prediksi Kanker Serviks," Feb. 2024. doi: <https://doi.org/10.47701/dutacom.v17i1.3790>.
- [19] Anubrata Deb, Bristi Samadder, Souroja Chowdhury, Sulekha Das, and Shweta Banarjee, "Measuring Mental Health Condition using Logistic regression," *international journal of engineering technology and management sciences*, vol. 7, no. 2, 2023, doi: 10.46647/ijetms.2023.v07i02.040.
- [20] R. Alfarezy, E. Ermatita, and R. M. B. Wadu, "Implementasi Algoritma Naïve Bayes Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus: Open Sourcing Mental Illness)," *Informatik : Jurnal Ilmu Komputer*, vol. 19, no. 1, 2023, doi: 10.52958/iftk.v19i1.4696.
- [21] D. Shan *et al.*, "Association between patterns of biological rhythm and self-harm: evidence from the baoxing youth mental health (BYMH) cohort," *Child Adolesc Psychiatry Ment Health*, vol. 18, no. 1, 2024, doi: 10.1186/s13034-023-00685-w.