

---

# Prediksi Tingkat Stres dan Mood Berdasarkan Kebiasaan Digital Menggunakan Linear Regression, XGBoost, dan MLPRegressor

Fitran Alfian Nizar<sup>1</sup>, Marvi Yoga Pratama<sup>2</sup>, Irsyad Nur Hidayatulloh<sup>3</sup>

<sup>1</sup>Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung

<sup>2</sup>Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung

<sup>3</sup>Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung

---

## Article Info

### Article history:

Received month dd, yyyy

Revised month dd, yyyy

Accepted month dd, yyyy

### Keywords:

Mental health

Digital habits

Linear Regression

XGBoost

MLP Regressor

---

## ABSTRACT

Kesehatan mental merupakan isu penting dalam era digital saat ini. Berbagai penelitian menunjukkan bahwa kebiasaan digital seperti durasi *screen time* dan penggunaan media sosial memiliki korelasi terhadap tingkat stres dan suasana hati seseorang [1]. Dalam studi ini, dilakukan analisis terhadap dataset Digital Habits vs Mental Health yang berisi data kebiasaan digital dan kondisi psikologis responden. Tiga algoritma *machine learning* regresi digunakan untuk membangun model prediktif, yaitu Linear Regression, XGBoost Regressor, dan MLPRegressor. Model diuji pada dua target yaitu *stress\_level* dan *mood\_score*. Berdasarkan evaluasi menggunakan metrik  $R^2$  Score, MAE, MSE, dan RMSE, model MLPRegressor menunjukkan performa terbaik dalam memprediksi kedua target. Hasil ini menunjukkan bahwa model *Neural Network* lebih unggul dalam menangani data non-linear yang kompleks dibandingkan model linier maupun *boosting*.

This is an open access article under the [CC BY-SA](#) license.



---

## Corresponding Author:

Fitran Alfian Nizar

Jurusan Teknik Informatika, UIN Sunan Gunung Djati Bandung

Email: fitran.nizar@gmail.com

---

## 1. PENDAHULUAN

Dalam era digital yang semakin maju, penggunaan perangkat digital seperti smartphone, media sosial, dan berbagai platform hiburan telah menjadi bagian integral dari kehidupan sehari-hari [2]. Aktivitas digital ini, meskipun membawa berbagai manfaat, juga menimbulkan kekhawatiran terhadap kesehatan mental penggunaannya [3]. Sejumlah penelitian menunjukkan bahwa *screen time* yang berlebihan, penggunaan media sosial yang intens, serta pola tidur yang terganggu berpotensi mempengaruhi tingkat stres dan suasana hati seseorang [4]. Oleh karena itu, penting untuk memahami sejauh mana kebiasaan digital berkontribusi terhadap kondisi psikologis individu [5].

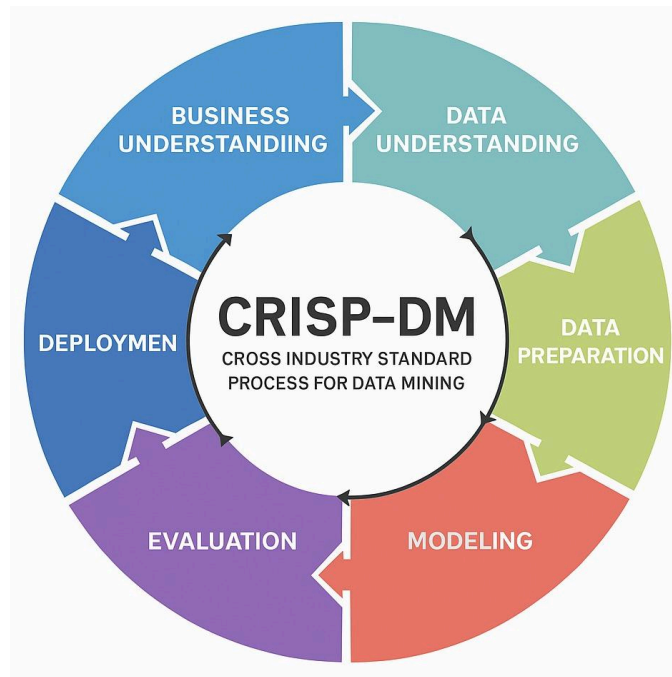
Penelitian ini bertujuan untuk mengeksplorasi hubungan antara kebiasaan digital dengan kesehatan mental, khususnya dalam memprediksi tingkat stres (*stress level*) dan skor suasana hati (*mood score*) [6]. Dengan menggunakan dataset “Screen Time Impact on Mental Health” yang tersedia di Kaggle dengan url <https://www.kaggle.com/datasets/abhishek9/digital-habits-vs-mental-health-dataset>, studi ini membangun dan membandingkan model prediksi dengan tiga algoritma *machine learning* regresi, yaitu Linear Regression, XGBoost Regressor, dan MLPRegressor. Ketiga algoritma ini dipilih karena mewakili

---

pendekatan yang berbeda dalam pemrosesan data: model linier klasik, model *boosting* berbasis pohon keputusan, dan model *neural network* sederhana. Tujuan utama dari penelitian ini adalah untuk mengetahui algoritma mana yang memiliki performa terbaik dalam memprediksi kesehatan mental berdasarkan data kebiasaan digital. Dengan hasil tersebut, diharapkan penelitian ini dapat menjadi pijakan awal dalam pengembangan sistem monitoring kesehatan mental berbasis data digital secara pasif dan *real-time*, terutama bagi kalangan muda yang sangat lekat dengan penggunaan teknologi [7].

## 2. METODE

Penelitian ini menggunakan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka metodologi utama [8]. CRISP-DM terdiri dari enam tahapan, yaitu: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [9]. Setiap tahapan dalam penelitian ini dirancang untuk mendukung proses pembangunan model prediksi stres dan mood berdasarkan kebiasaan digital seseorang menggunakan algoritma machine learning.



Gambar 1. CRISP-DM

### 2.1. Business Understanding

Permasalahan utama dalam penelitian ini adalah meningkatnya risiko gangguan kesehatan mental akibat pola hidup digital yang intens, seperti *screen time* berlebih dan penggunaan media sosial [10]. Oleh karena itu, tujuan dari penelitian ini adalah membangun model *machine learning* yang mampu memprediksi tingkat stres (*stress level*) dan suasana hati (*mood score*) seseorang berdasarkan kebiasaan digital mereka. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan sistem pemantauan kesehatan mental berbasis teknologi [11].

### 2.2. Data Understanding

Dataset yang digunakan diperoleh dari platform Kaggle dengan judul Digital Habits vs Mental Health. Dataset ini berisi data hasil survei dari responden mengenai kebiasaan digital harian dan kondisi psikologis mereka [12]. Beberapa fitur yang digunakan dalam penelitian ini antara lain: *screen\_time\_hours*, *social\_media\_platforms\_used*, *hours\_on\_TikTok*, dan *sleep\_hours*. Sedangkan target variabel yang diprediksi meliputi *stress\_level* dan *mood\_score*. Eksplorasi awal dilakukan untuk memastikan bahwa data bersih dan sesuai untuk proses pemodelan, termasuk pemeriksaan tipe data dan ketersediaan nilai [13].

### 2.3. Data Preparation

Tahapan ini meliputi seleksi fitur, pembagian data, dan normalisasi. Fitur-fitur yang dianggap relevan dipilih untuk dijadikan variabel input (X), dan target (y) dipisahkan berdasarkan kasus yang dianalisis, yakni prediksi *stress\_level* dan *mood\_score*. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi *train\_test\_split* [14]. Untuk

meningkatkan performa model, dilakukan proses normalisasi menggunakan StandardScaler pada fitur numerik, khusus untuk model Linear Regression dan MLP Regressor. Proses ini bertujuan untuk menghindari dominasi fitur tertentu akibat perbedaan skala [15].

#### 2.4. Modeling

Tahapan *modeling* bertujuan untuk membangun model prediktif berdasarkan data yang telah dipersiapkan sebelumnya. Pada tahap ini, algoritma regresi diterapkan untuk mempelajari hubungan antara kebiasaan digital pengguna dengan kondisi psikologis yang diukur melalui skor stres dan mood [16]. Tiga algoritma regresi dipilih untuk membangun model prediktif, yaitu:

1. Linear Regression: algoritma klasik yang digunakan untuk mengidentifikasi hubungan linier antar variabel.
2. XGBoost Regressor: algoritma berbasis *gradient boosting* yang unggul dalam menangani data non-linier dan interaksi antar fitur [17].
3. MLP Regressor (Multi-layer Perceptron): algoritma berbasis jaringan saraf tiruan dengan satu lapisan tersembunyi yang dirancang untuk menangkap pola kompleks dalam data [18].

Setiap model dilatih dua kali, masing-masing untuk prediksi *stress\_level* dan *mood\_score*. Proses pelatihan disertai pencatatan waktu (*training time*) dan waktu prediksi (*prediction time*) untuk mengukur efisiensi komputasi [19].

#### 2.5. Evaluation

Evaluasi dilakukan dengan menggunakan metrik regresi, yaitu  $R^2$  Score, *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) [20]. Selain itu, waktu pelatihan dan prediksi juga dicatat untuk mengukur efisiensi setiap model. Hasil evaluasi disajikan dalam bentuk tabel dan digunakan untuk membandingkan performa ketiga model dalam masing-masing kasus. Model dengan  $R^2$  paling tinggi dan *error* paling rendah dianggap sebagai model terbaik. Berdasarkan hasil evaluasi, model XGBoost Regressor menunjukkan performa paling unggul dalam kedua skenario prediksi.

#### 2.6. Deployment

Tahap *deployment* belum dilakukan dalam penelitian ini karena fokus utama adalah pada proses eksplorasi dan evaluasi model. Namun, berdasarkan hasil evaluasi, model XGBoost Regressor memiliki potensi tinggi untuk diterapkan dalam sistem prediksi waktu nyata (*real-time monitoring system*) untuk mendeteksi kondisi psikologis pengguna berdasarkan aktivitas digital mereka [21].

### 3. HASIL PENELITIAN

Setelah dilakukan pelatihan model terhadap data kebiasaan digital, dilakukan pengujian pada dua target variabel yang berbeda, yaitu *stress\_level* dan *mood\_score*. Masing-masing model—Linear Regression, XGBoost Regressor, dan MLP Regressor—dilatih dan dievaluasi berdasarkan metrik performa seperti  $R^2$  Score, MAE, MSE, RMSE, serta efisiensi waktu pelatihan dan prediksi.

#### 1. Prediksi Tingkat Stres (*stress\_level*)

Pada prediksi terhadap variabel *stress\_level*, model MLP Regressor menunjukkan performa paling unggul dengan nilai  $R^2$  tertinggi dibandingkan dua model lainnya. Model ini juga menghasilkan nilai MAE dan RMSE yang lebih rendah, yang menandakan akurasi prediksi yang lebih baik. XGBoost memiliki performa cukup baik namun cenderung underfitting terhadap kompleksitas data. Sementara itu, Linear Regression memberikan hasil yang fluktuatif, kemungkinan karena belum optimal dalam hal parameterisasi.

Tabel 1 hasil evaluasi model terhadap prediksi *stress\_level*

Model	$R^2$ Score	MAE	MSE	RMSE	Train Time (s)	Predict Time (s)
MLP Regressor	0.751	0.815	1.046	1.023	7.9191	0.0150
Linear Regression	0.748	0.823	1.056	1.028	0.0352	0.0005
XGBoost	0.746	0.823	1.069	1.034	0.4289	0.0401

## 2. Prediksi Skor Mood (mood\_score)

Pada skenario kedua, yaitu prediksi terhadap mood\_score, hasil yang diperoleh juga menunjukkan bahwa model MLP Regressor konsisten memberikan performa terbaik dengan nilai  $R^2$  tertinggi dan error terendah. Model ini menunjukkan kestabilan baik dalam mempelajari hubungan antar variabel, terutama ketika terdapat non-linearitas dalam data. XGBoost menunjukkan peningkatan performa dibandingkan skenario pertama, namun masih belum melampaui MLP Regressor. Linear Regression menunjukkan performa paling rendah, menandakan bahwa hubungan antar fitur dan mood\_score tidak sepenuhnya linier.

Tabel 2 hasil evaluasi model terhadap prediksi mood\_score

Model	$R^2$ Score	MAE	MSE	RMSE	Train Time (s)	Predict Time (s)
MLP Regressor	0.662	0.518	0.549	0.741	7.4827	0.0127
XGBoost	0.653	0.511	0.563	0.750	0.3452	0.0424
Linear Regression	0.567	0.660	0.702	0.838	0.0118	0.0005

## 4. CONCLUSION

Penelitian ini menunjukkan bahwa kebiasaan digital seperti durasi penggunaan layar, jumlah platform media sosial yang digunakan, waktu yang dihabiskan di TikTok, dan jumlah jam tidur memiliki pengaruh terhadap tingkat stres dan suasana hati seseorang. Tiga algoritma regresi telah diterapkan, yaitu Linear Regression, XGBoost Regressor, dan MLPRegressor, untuk membandingkan performa dalam memprediksi dua variabel psikologis yaitu *stress\_level* dan *mood\_score*. Hasil evaluasi menunjukkan bahwa MLP Regressor memiliki performa terbaik secara konsisten untuk kedua target prediksi, ditinjau dari nilai  $R^2$  Score tertinggi dan nilai error (MAE, MSE, RMSE) yang paling rendah. Hal ini menunjukkan bahwa arsitektur jaringan saraf memiliki kemampuan lebih baik dalam menangkap pola kompleks dan non-linier dari data kebiasaan digital dibandingkan model lainnya. Sementara itu, XGBoost Regressor menunjukkan performa yang kompetitif, namun sedikit di bawah MLP dalam hal akurasi prediksi, meskipun unggul dari sisi efisiensi waktu pelatihan dan prediksi. Linear Regression menghasilkan performa paling rendah, terutama karena keterbatasannya dalam memodelkan hubungan non-linier yang kemungkinan besar ada dalam data.

Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi arsitektur MLP yang lebih dalam dan teknik *tuning* hiperparameter untuk lebih mengoptimalkan performa. Selain itu, penggunaan dataset yang lebih besar dan beragam, serta penambahan fitur seperti jenis kelamin, usia, pekerjaan, dan pola interaksi digital lainnya, dapat meningkatkan akurasi prediksi. Pendekatan *ensemble learning* atau model *deep learning* yang lebih kompleks juga layak dipertimbangkan untuk meningkatkan generalisasi dalam memprediksi kondisi psikologis berdasarkan jejak digital seseorang.

## DAFTAR PUSTAKA

- [1] M. T. T. Nguyen, M. D. Tran, and H. V. Nguyen, "Mental health prediction using machine learning algorithms," *ICT Express*, vol. 8, no. 3, pp. 374–379, 2022.
- [2] A. Mohr, M. Zhang, dan S. Schueller, "Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning," *Annu. Rev. Clin. Psychol.*, 2017–2024.
- [3] M. L. Birnbaum, H. Wen, *et al.*, "Identifying emerging mental illness utilizing search engine activity: a feasibility study," *PloS One*, 2020. doi: 10.1371/journal.pone.0240820.
- [4] A. Campbell *et al.*, "StudentLife: Assessing Behavioral Trends, Mental Well-being and Academic Performance using Smartphones," *UbiComp*, 2014; revisited for COVID-19 in 2020.
- [5] "Stress can be detected during emotion-evoking smartphone use: a pilot study using machine learning," *Frontiers in Digital Health*, 2025.
- [6] "Multimodal Privacy-preserving Mood Prediction from Mobile Data: A Preliminary Study," *arXiv*, 2020.
- [7] M. Razavi *et al.*, "Machine learning, deep learning and data preprocessing techniques for detection, prediction, and monitoring of stress and stress-related mental disorders: a scoping review," *arXiv*, 2023.
- [8] G. Xu, Y. Shi, X. Sun, and W. Shen, "Internet of things in marine environment monitoring: A review," *Sensors (Switzerland)*, vol.

- 
- 19, no. 7, pp. 1–21, 2019, doi: 10.3390/s19071711.
- [9] S. Leonelli and N. Tempini, *Data Journeys in the Sciences*. 2020.
  - [10] H. A. Rahman *et al.*, “Machine Learning-Based Prediction of Mental Well-Being Using Health Behavior Data from University Students,” *Diagnostics*, vol. 10, no. 5, 2024.
  - [11] N. Stylos and J. Zwiegelaar, *Big Data as a Game Changer: How Does It Shape Business Intelligence Within a Tourism and Hospitality Industry Context?*, 2019.
  - [12] N. Suryawanshi, “Predicting Mental Health Outcomes Using Wearable Device Data and Machine Learning,” *Int. J. Innov. Sci. Res. Technol.*, vol. 6, no. 3, pp. 1334–1341, Mar. 2021. doi: 10.38124/ijisrt/IJISRT21MAR587.
  - [13] J. Nan *et al.*, “Depression Prediction by Using Ecological Momentary Assessment, Actiwatch Data, and Machine Learning,” *JMIR Mhealth Uhealth*, vol. 7, no. 10, 2019-2025 reprint. doi: 10.2196/14149.
  - [14] S. Mahalingam *et al.*, “How do machine learning models perform in the detection of depression, anxiety, and stress among undergraduate students? A systematic review,” *PMC*, 2023.
  - [15] Q. Song, H. Ge, J. Caverlee, and X. Hu, “Tensor completion algorithms in big data analytics,” *arXiv*, vol. 13, no. 1, 2017.
  - [16] G. S. *et al.*, “EmoSens: Emotion Recognition based on Sensor data analysis using LightGBM,” *arXiv*, 2022.
  - [17] E. Turska, S. Jurga, dan J. Piskorski, “Mood Disorder Detection in Adolescents by Classification Trees, Random Forests and XGBoost in Presence of Missing Data,” *Entropy*, vol. 23, no. 9, p. 1210, 2021.
  - [18] N. Izzah, A. P. Sutarto, dan M. Hariyadi, “Machine Learning models for the Cognitive Stress Detection Using Heart Rate Variability Signals,” *J. Teknik Industri*, vol. 24, no. 2, pp. 83–94, 2023. doi: 10.9744/jti.24.2.83-94.
  - [19] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, *Smarter traffic prediction using big data, in-memory computing, deep learning and gpus*, vol. 19, no. 9. 2019.
  - [20] K. D. K. Wardani, T. Wijaya, P. Madona, J. N. Sari, dan Y. Yuliska, “Mobile-based Stress Level Detection using Tree-Based Machine Learning Algorithms,” in *Proc. ABEC*, Bengkalis, Riau, Indonesia, Sep. 2023, pp. –, 2024.
  - [21] M. Asif, S. Mishra, *et al.*, “Proactive Emotion Tracker: AI-Driven Continuous Mood and Emotion Monitoring,” *arXiv*, 2024.
-