Jurnal TRANSFORMATIKA
Vol.23, No.2, Januari 2026, pp. 164 - 173
P-ISSN: 1693-3656, E-ISSN: 2460-6731
https://journals.usm.ac.id/index.php/transformatika/

Analisa Performa Metode LightGBM untuk Prediksi Kecanduan Media Sosial

Roudhotul Jannah[1],Rastri Prathivi[2*]
[1]1Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Semarang, telp:024-6702757, e-mail: jannahr786@gmail.com
[2*]1Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, Semarang, telp:024-6702757, e-mail: vivi@usm.ac.id

ARTICLE INFO

History of the article :
Received 18 November 2025
Received in revised form 2 Desember 2025
Accepted 13 Januari 2026
Available online 31 Januari 2026

Keywords:

LightGBM; kecanduan media sosial, pembelajaran mesin, prediksi, pemilihan fitur.

*** Correspondece:**

Telepon:
+6285950251065

E-mail:
vivi@usm.ac.id

ABSTRACT

Social media has now become an integral part of daily activities, driven by the increasingly rapid development of digital technology. Excessive social media use can trigger negative impacts such as psychological disorders, sleep deprivation, and social conflict. This study assesses the effectiveness of the Light Gradient Boosting Machine (LightGBM) in predicting social media addiction using data from 705 respondents from Kaggle. The analysis stages included data cleaning, categorical variable transformation, and feature selection based on Pearson correlation. The model was trained with a 70:30 ratio and evaluated using accuracy, precision, recall, and f1-score. The results showed 98% accuracy, thus LightGBM is considered highly effective as a social media addiction prediction model.

INTRODUCTION

Media sosial diera 5.0 memiliki hubungan yang sangat erat dengan manusia, mulai dari kalangan anak-anak hingga orang dewasa [1]. Kecenderungan kecanduan media sosial adalah fenomena yang sering terjadi pada saat ini seiring dengan meningkatnya penggunaan internet serta canggihnya kemajuan teknologi [2]. Karena banyaknya hal – hal dan topik yang dapat di*posting* secara bebas di media sosial, maka media sosial menjadi sarana prertengkaran dan pemantik berbagai konflik sosial, bagi antar individu atau antar golongan .

Penggunaan media sosial yang berlebihan serta tidak bijak atau disebut kecanduan media social, dapat menimbulkan berbagai dampak negatif bagi penggunanya, seperti masalah dalam hubungan sosial, kecenderungan konsumtif, kebiasaan menunda-nunda, penurunan prestasi akademik, manajemen waktu yang tidak efektif, lemahnya kontrol diri, serta munculnya prasangka negatif [3]. Dengan kemajuan teknologi saat ini, metode Machine Learning dapat digunakan untuk menganalisis dan memprediksi kecanduan media social [4].

Penelitian ini menggunakan dataset publik yang diperoleh dari platform Kaggle dengan judul Students Social Media Addiction Dataset, yang berisi data hasil survei terhadap pelajar dan mahasiswa dari beberapa negara Asia Selatan. Pemilihan dataset luar negeri ini bukan karena ketidakmungkinan melakukan penelitian primer, melainkan karena fokus utama penelitian terletak pada penerapan dan evaluasi metode LightGBM dalam memprediksi tingkat kecanduan media sosial.

Terdapat empat mpat fitur dengan hubungan yang kuat terhadap tingkat kecanduan media sosial, yaitu penggunaan media sosial yang tidak dapat di kontrol pada kalangan remaja dapat mempengaruhi kesehatan mental [5]. Durasi penggunaan Pengguna dapat melupakan satu hal yaitu waktu atau durasi yang mereka gunakan untuk mengakses media sosial [6] . Skor kesehatan mental berdampak juga kepada berbagai aspek kehidupan manusia [7] Jam tidur berperan sebagai indikator gaya hidup yang terpengaruh oleh aktivitas daring [8].

Penelitian ini menerapkan metode Light Gradient Boosting Machine (LightGBM), sebuah algoritma gradient boosting yang dikembangkan oleh Microsoft dan memiliki berbagai keunggulan dibandingkan metode ensemble lainnya. LightGBM dikenal mampu melakukan pelatihan model dengan sangat cepat berkat penggunaan teknik Histogram-based Decision Tree Learning serta strategi pertumbuhan leaf-wise with depth limitation. Selain memiliki akurasi tinggi, metode ini hemat memori, mampu mengolah dataset berukuran besar, dan dapat memproses fitur kategori secara langsung tanpa *one-hot encoding*. Berdasarkan uraian di atas, dapat dirumuskan permasalahannya yaitu bagaimana memprediksi tingkat kecanduan media sosial menggunakan algoritma LightGBM berdasarkan perilaku penggunaan, kondisi kesehatan mental, jam tidur, dan konflik sosial yang dialami pengguna.

RESEARCH METHODS

Metodologi penelitian yang digunakan adalah sebagai berikut:

1. **Identikasi Masalah**
   Pendekatan *machine learning* seperti LightGBM dimanfaatkan untuk mengolah data secara menyeluruh, mengenali pola tersembunyi, dan menghasilkan prediksi yang lebih objektif serta mendukung proses pengambilan keputusan berbasis informasi yang terukur [9].
2. **Pengumpulan Dataset**
   Dataset yang digunakan dalam penelitian ini diambil dari situs www.kaggle.com. yang memuat data survei terkait perilaku penggunaan media sosial. Data tersebut mencakup 705 responden dengan variabel seperti ratarata durasi penggunaan harian, jam tidur, skor kesehatan mental, konflik yang ditimbulkan oleh media sosial, serta beberapa fitur kategorikal lainnya.
3. **Import Library**
   Tahap awal penelitian ini adalah mengimpor pustaka (library) yang digunakan untuk pengolahan data, pembangunan model, dan evaluasi performa. Library yang digunakan adalah pandas. numpy, matplotlib.pyplot, seaborn, matplotlib.pyplot, labelEncoder, lightgnm, serta h.classification_report,confusion_matrix [10].
4. **Load Dataset**
   Mengacu pada proses mengimpor data dari sumber eksternal ke dalam lingkungan Python agar dapat dianalisis, dimanipulasi, atau digunakan untuk tugas pembelajaran mesin. Ini merupakan langkah mendasar dalam setiap proyek yang berkaitan dengan data [11]. Dataset yang digunakan berasal dari platform Kaggle dengan total 705 responden. Data dimuat menggunakan pustaka pandas dengan fungsi pd.read_csv() untuk membaca file CSV.
5. **Preprocessing**
   Praprocesing data adalah proses mengubah data mentah menjadi format yang bersih dan terstruktur sebelum menerapkan machine learning [12]. Pre-processing data meliputi cleaning, yaitu menganti data atau menghilangkan data noise ataupun missing value, proses normalisasi untuk memodifikasi nilai dalam variabel sehingga kita dapat mengukurnya dalam skala umum atau rentang tertentu. Pada penelitian ini menggunakan MinMax Normalization lalu proses

transformasi, mengubah data asli ke bentuk data tujuan agar mudah di [13]. Praprocessing dilakukan dengan memuar data "kecanduan media.csv" menggunakan pustaka pandas.

6. **Pengecekan dan Penanganan Missing Values**
   Pengecekan nilai kosong pada setiap kolom dilakukan menggunakan fungsi isnull().sum() untuk mengetahui apakah terdapat data yang hilang. Data yang memiliki nilai kosong dihapus menggunakan fungsi dropna(). Langkah ini memastikan hanya data lengkap yang digunakan dalam pelatihan model.
7. **Prmisahan Fitur dan Label**
   Pemilihan fitur pada penelitian ini dilakukan menggunakan metode korelasi Pearson, dengan mempertimbangkan fitur-fitur yang memiliki nilai korelasi absolut terhadap variabel target lebih besar dari 0.1.
8. **Labelling Tabel & Seleksi Fitur**
   Dalam tahap pra-pemrosesan data, langkah pertama yang dilakukan adalah membentuk variabel target (Addicted_Label) melalui proses pelabelan data. Penentuan label mengacu pada skor kecanduan (Addicted_Score), di mana responden dengan skor ≥ 6 digolongkan sebagai kecanduan (label 1), sedangkan responden dengan skor < 6 dimasukkan ke dalam kategori tidak kecanduan (label 0). Ambang batas skor 6 dipilih berdasarkan hasil analisis distribusi, yang menunjukkan bahwa nilai tersebut merupakan titik yang tepat untuk memisahkan kelompok pengguna dengan perilaku penggunaan normal dan kelompok yang sudah menampilkan indikasi kecanduan.
9. **Eksplanatory Data Analysis (EDA)**
   EDA adalah suatu pendekatan untuk melihat apa yang dapat disampaikan oleh data kepada kita; itu membantu menganalisis kumpulan data dan menguraikan karakteristik statistiknya [14]. EDA akan dilakukan untuk memahami hubungan antara variabel. Ini mencakup pembuatan plot scatter untuk memvisualisasikan hubungan antara variabel, dan mungkin juga mencakup perhitungan korelasi untuk mengukur kekuatan dan arah hubungan antara variabel [15]
10. **Pembagian Dataset**
    Dataset dibagi menjadi 70% data latih dan 30% data uji menggunakan fungsi train test split dari Scikit-Learn. Pembagian ini dilakukan untuk memastikan model dapat belajar dari sebagian data, lalu diuji pada data yang belum pernah dilihat sebelumnya. Parameter random_state=42 digunakan untuk memastikan hasil yang konsisten dan dapat direproduksi. Rasio ini dipilih karena dianggap seimbang antara kebutuhan model untuk belajar pola dari data dan kebutuhan evaluasi model dengan data yang belum pernah dilihat sebelumnya.
11. **Penerapan Model LightGBM**
    Tahap ini merupakan proses pelatihan model klasifikasi menggunakan algoritma Light Gradient Boosting Machine (LightGBM), salah satu metode ensemble learning yang berbasis Gradient Boosting Decision Tree (GBDT). Prinsip kerjanya adalah membangun model secara bertahap, di mana setiap pohon keputusan baru dirancang untuk memperbaiki kesalahan prediksi dari pohon-pohon sebelumnya. Pendekatan ini membuat LightGBM mampu menghasilkan prediksi yang akurat dengan waktu komputasi yang efisien. Implementasi dilakukan menggunakan pustaka LightGBM di Python melalui kelas LGBMClassifier.

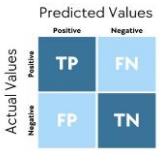
12. **Evaluasi Model**

Label dari model klasifikasi *Light Gradient Boosting Machine* (LightGBM) ditentukan sebagai berikut,

a. (Kecanduan) adalah responden yang menggunakan media sosial secara berlebihan, ditandai dengan durasi penggunaan harian tinggi, kualitas tidur rendah, skor kesehatan mental rendah, dan konflik sosial yang tinggi.

b. (Tidak Kecanduan) adalah responden yang menggunakan media sosial dalam batas wajar, memiliki kualitas tidur cukup, kesehatan mental baik, dan jarang mengalami konflik sosial akibat media sosial.

Setiap prediksi dikategorikan menjadi empat komponen seperti pada gambar di bawah ini :



Gambar 2. Evaluasi Model Confusion Ma

1. **True Positive (TP)** adalah Model memprediksi "Kecanduan" dan benar.
2. **True Negative (TN)** adalah Model memprediksi "Tidak Kecanduan" dan benar.
3. **False Positive (FP)** adalah Model memprediksi "Kecanduan" tetapi sebenarnya "Tidak Kecanduan".
4. **False Negative (FN)** adalah Model memprediksi "Tidak Kecanduan" tetapi sebenarnya "Kecanduan".

13.Equation

a. Akurasi (*Accuracy*)

Akurasi mengukur seberapa besar persentase prediksi model yang benar, baik pada kelas kecanduan maupun tidak kecanduan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Hasil pengujian menunjukkan akurasi sebesar **0.98%**, yang menandakan bahwa LightGBM mampu memprediksi status kecanduan media sosial dengan tingkat kesalahan yang sangat rendah.

b. ***Precision*** **(Presisi)**

Presisi mengukur ketepatan model dalam memprediksi kelas positif ("Kecanduan") dari seluruh data yang diprediksi sebagai positif:

$$recision = \frac{TP}{TP+FP}$$

Presisi yang tinggi menunjukkan bahwa model jarang salah mengklasifikasikan responden yang tidak kecanduan sebagai kecanduan.

c. Recall (Sensitivitas)

Recall mengukur kemampuan model untuk mendeteksi semua data yang benar-benar positif ("Kecanduan"):

$$Recall = \frac{TP}{TP + FN}$$

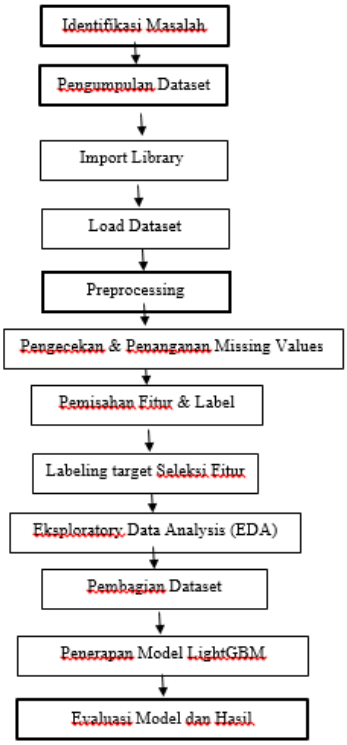
Nilai recall yang tinggi berarti model mampu mendeteksi hampir semua responden yang benar-benar kecanduan.

d. F1-Score

F1-Score adalah rata-rata harmonik dari precision dan recall:

$$F1 = \frac{2\ .\ Precision\ .\ Recall}{Precision + Recall}$$

Nilai F1-Score yang tinggi pada kedua kelas menunjukkan bahwa model tidak hanya akurat, tetapi juga seimbang antara ketepatan prediksi dan kemampuan mendeteksi semua kasus kecanduan. Metodologi penelitian yang digunakan dalam diagram alir adalah sebagai berikut:



Gambar 1. Alur Penelitian

RESULT AND DISCUSSION

1. Import Library

Penelitian ini dijalankan sepenuhnya di Google Colab, platform berbasis cloud dengan memanfaatkan pustaka pandas. Proses visualisasi dilakukan dengan matplotlib.pyplot dan seaborn guna menghasilkan grafik dan heatmap yang mempermudah analisis keterkaitan antar variabel,

2. Load Dataset

Data yang digunakan berjumlah 705 data responden dengan 8 kolom numerik, statistik data numerik dapat dilihat pada tabel 1 berikut :

Tabel 1. Data Primer

Id	Usia	Rata-rata Durasi penggunaan harian	Platform	Jam Tidur Per Malam	Pengaruh jam tidur per malam	Skor Kesehatan Mental	Konflik Media Sosial	Skor Kecanduan
1	19	5,2	Instagram	6,5	Yes	6	3	8
2	22	2,1	Twitter	7,5	No	8	0	3

3	20	6	TikTok	5	Yes	5	4	9
4	18	3	YouTube	7	No	7	1	4
5	21	4,5	Facebook	6	Yes	6	2	7

3. Pengolahan Dataset (Prepocessing)

Pengolahan data set (*Preprocessing*) meliputi penanganan dan pengecekan missing values Hasil pengecekan menunjukkan bahwa terdapat beberapa kolom numerik yang memiliki nilai kosong, yaitu: 1. Rata-rata Jam Penggunaan Harian 3 nilai kosong 2. Jam Tidur per Malam 2 nilai kosong 3. Skor Kesehatan Mental 1 nilai kosong Untuk mengatasi menghapus baris yang memiliki nilai kosong (dropna()), penelitian ini menggunakan metode imputasi median untuk mengisi kekosongan data.

4. Pemisahan Fitur dan Label

Penelitian ini menggunakan empat variabel utama terhadap 705 responden. DIbagi menjadi 2 jenis label yaitu tidak kecanduan (0) dan kecanduan (1).

5. Pelabelan Target dan Seleksi Fitur

Labeling dan target seleksi fitur ditampilkan pada Tabel 2

Tabel 2 Label Target

Kategori	Jumlah Responden
Tidak Kecanduan (0)	236
Kecanduan (1)	469

Heatmap korelasi dalam bentuk tabel ditampilkan sebagai berikut

Tabel 3. Seleksi Fitur

	Fitur	Nilai Korelasi
1	Konflik Media Sosial	0.802294
2	Rata-rata Jam Penggunaan Harian	0.653111
3	Jam Tidur Per Malam	-0.612959
4	Skor Kesehatan Mental	-0.787843

Hasil seleksi fitur menunjukkan bahwa empat variabel memiliki korelasi signifikan dengan kecanduan media sosial. Konflik Media Sosial dan Rata-rata Jam Penggunaan Harian memiliki korelasi positif, menandakan semakin tinggi nilainya, semakin besar risiko kecanduan. Sementara itu, Jam Tidur Per Malam dan Skor Kesehatan Mental berkorelasi negatif, artinya semakin rendah nilainya, semakin tinggi kecenderungan kecanduan. Keempat fitur ini dipilih karena korelasinya cukup kuat terhadap label kecanduan. Terlihat pada table 3.

DOI : 10.26623/transformatika.v23i2.13165

6. Eksploratori Data Analisis

Dari analisis statistik deskriptif, diperoleh bahwa dataset terdiri dari 705 responden. Usia rata-rata responden adalah 21 tahun, dengan rentang antara 18 hingga 24 tahun. Durasi penggunaan media sosial per hari rata-rata sebesar 5 jam, dengan penggunaan terendah 2 jam dan tertinggi 8 jam. Sementara itu, jam tidur rata-rata adalah 7 jam per malam, dengan kisaran 4 hingga 10 jam.

Skor kesehatan mental memiliki rata-rata 6, dengan nilai terendah 4 dan tertinggi 9, sedangkan konflik akibat media sosial berada pada rata-rata 3, dengan maksimum 5. Skor kecanduan media sosial menunjukkan rata-rata 6, dengan nilai minimum 2 dan maksimum 9.

Variabel target (Addicted_Label) memperlihatkan bahwa mayoritas responden masuk kategori kecanduan (label 1). Temuan ini memberikan gambaran awal bahwa tingginya intensitas penggunaan media sosial di kalangan responden berpotensi memengaruhi jam tidur, kondisi kesehatan mental, dan tingkat konflik sosial, sehingga mendukung relevansi variabelvariabel tersebut dalam pemodelan prediksi kecanduan. Terlihat pada tabel 4.

Tabel 4. Tabel Fitur

Konflik Media Sosial	Rata-rata Durasi penggunaan harian	Pengaruh jam tidur per malam	Skor Kesehatan Mental
3	5,2	6,5	6
0	2,1	7,5	8
4	6	5	5
1	3	7	7
2	4,5	6	6

7. Pembagian Dataset

Hasil dari proses split data menunjukkan pembagian dataset menjadi 70% data latih dan 30% data uji. Pembagian ini memastikan evaluasi dilakukan pada data yang belum pernah digunakan saat pelatihan, sehingga hasilnya lebih objektif.

8. Penerapan Model LightGBM

Berdasarkan confusion matrix yang dihasilkan, model LightGBM menunjukkan performa klasifikasi yang sangat baik dalam membedakan antara responden yang kecanduan dan yang tidak kecanduan media sosial. Dalam hal ini, label Actual: 0 mewakili responden yang tidak mengalami kecanduan, sedangkan Actual: 1 menunjukkan responden yang mengalami kecanduan. Dari total 212 data uji, sebanyak 142 responden yang kecanduan berhasil diklasifikasikan dengan benar, dan 67 responden yang tidak kecanduan juga diprediksi secara akurat.

9. Evaluasi Model dan Hasil

a. Hasil Akurasi

Berdasarkan confusion matrix, dari total 67 data pada kelas 0 (Tidak Kecanduan), sebanyak 66 data berhasil diprediksi dengan benar dan hanya 1 data yang keliru. Sementara itu, pada kelas 1 (Kecanduan), model memprediksi benar 142 data dari total 145 data, dengan 3 data mengalami kesalahan klasifikasi. Laporan klasifikasi (classification report) memperlihatkan bahwa untuk kelas 0, nilai precision mencapai 0,96, recall sebesar 0,99, dan f1-score sebesar 0,97. Pada kelas 1, diperoleh precision sebesar 0,99, recall sebesar 0,98, dan f1-score sebesar 0,99. Nilai yang tinggi pada semua metrik ini menunjukkan bahwa model memiliki kinerja yang seimbang dan konsisten pada kedua kelas.

b. Hasil Data Training dan Data Testing

Tabel 5. Data Training dan Testing

Dataset	Akurasi
Data Latih	97%
Data Uji	98%

Tingginya capaian ini memperlihatkan bahwa model mampu mempelajari pola keterkaitan antara fitur-fitur, seperti usia, rata-rata durasi penggunaan media sosial harian, jam tidur per malam, skor kesehatan mental, serta konflik akibat media sosial, dengan kategori kecanduan secara efektif. Hasil akurasi terlihat pada tabel 5.

c. Analisis Feature Importance

Tabel 6. Fitur yang Penting

No	Fitur	Hasil
1	Jam Tidur Per Malam	515
2	Rata-rata Durasi Penggunaan Harian	374
3	Konflik Akibat Media Sosial	145
4	Skor Kesehatan Mental	72

Berdasarkan hasil analisis *feature importance* pada table 6, dari model LightGBM, variabel Jam Tidur per Malam memiliki pengaruh paling besar terhadap prediksi kecanduan media sosial, dengan tingkat kepentingan tertinggi dibandingkan variabel lainnya

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, model Light Gradient Boosting Machine (LightGBM) mampu memprediksi tingkat kecanduan media sosial dengan tingkat akurasi sangat tinggi, yaitu 0.98%. Pada kategori Tidak Kecanduan, model memperoleh precision sebesar 0,92, recall sebesar 0.97, dan skor F1 sebesar 0,96, sedangkan pada kategori Kecanduan, model mencapai precision sempurna sebesar 1,00, recall sebesar 0,97, dan skor F1 sebesar 0,99. Nilai rata-rata makro untuk precision, recall, dan skor F1 masing-masing adalah 0,96, 0,99, dan 0,97, dengan rata-rata tertimbang yang konsisten sebesar 0,98 untuk ketiga metrik tersebut. Hasil ini menunjukkan bahwa

DOI : 10.26623/transformatika.v23i2.13165

model dapat mengenali kedua kategori dengan tingkat kesalahan yang sangat rendah. Saran yang dapat diajukan untuk penelitian selanjutnya adalah penelitian selanjutnya dapat mempertimbangkan penggunaan algoritma pembelajaran mesin lain seperti XGBoost , CatBoost, atau Random Forest, untuk dibandingkan dengan metode dalam penelitian ini. Dapat digunakan dataset yang lebih bervariasi, dapat ditambahkan fitur lain, serta dapat dibangun sistem web atau mobile dengan sistem prediksi yang telah dibangun.

REFERENCES

[1] Syifa, S. F., Nur Istirohmah, A., Lestari, P., & Nur Azizah, M. (2023). Dampak Penggunaan Media Sosial terhadap Prestasi Belajar Peserta Didik. *Jurnal BELAINDIKA (Pembelajaran Dan Inovasi Pendidikan)*, 5(1), 21–27.

[2] Muna, R. F., & Astuti, T. P. (2014). Hubungan Antara Kontrol Diri Dengan Kecenderungan Kecanduan Media Sosial Pada Remaja Akhir. *Jurnal EMPATI*, 3(4), 481–491. https://doi.org/10.14710/empati.2014.7610

[3] Awalia, R., Fikrie, F., & Rifandi, A. (2022). Peranan Regulasi Diri Terhadap Kecenderungan Kecanduan Media Sosial Pada Mahasiswa. *Jurnal Psikologi Mandala*, 6(2), 85–100. https://doi.org/10.36002/jpm.v6i2.2006

[4] Rahma, M., Fikry, M., & Afrillia, Y. (2025). Prediksi Kesehatan Mental Remaja Berdasarkan Faktor Lingkungan Sekolah Menggunakan Machine Learning. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(2), 382–390. https://doi.org/10.30591/jpit.v10i2.8556

[5] Arsini, Y., Azzahra, H., Tarigan, K. S., & Azhari, I. (2023). Pengaruh Media Sosial Terhadap Kesehatan Mental Remaja. MUDABBIR Journal Reserch and Education Studies, 3(2), 50–54. https://doi.org/10.56832/mudabbir.v3i2.370

[6] Sinaga, M. N., & Aritonang, N. N. G. (2023). Hubungan Antara Durasi Penggunaan Media Sosial Dengan Kestabilan Emosi pada Pengguna Media Sosial Usia Dewasa Awal di Kota Medan. *Journal Of Social Science Research*, 3(3), 3870– 3883.

[7] Aisyah Fitriah, Dzaky Juliansyah, Umi Salamah, M Anugrah Utama, Opie Karunia Falah, & Aseh Miati. (2023). Pengaruh Media Sosial Terhadap Kesehatan Mental Pada Remaja Educate : *Journal Of Education and Learning*, 1(1), 32– 38. https://doi.org/10.61994/educate.v1i1.114

[8] SaThierbach, K., Petrovic, S., Schilbach, S., Mayo, D. J., Perriches, T., Rundlet, E. J. E. J. E. J., … Hoelz, A. (2015). No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title. *Proceedings of the National Academy of Sciences* (Vol. 3). Retrieved from http://dx.doi.org/10.1016/j.bpj.2015.06.056%0Ahttps://academic.oup.com/bioinformatics/article-abstract/34/13/2201/4852827%0Ainternalpdf://semisupervised-3254828305/semisupervised.ppt%0Ahttp://dx.doi.org/10.1016/j.str.2013.02.005%0Ahttp://dx.doi.org/10.10

[9] Umilizah, Nia [Editor]. (2025). *Machine Learning dalam Dunia Pendidikan*. Penamuda Media, Yogyakarta.

[10] Rossum, van Guido. (2018). *Phyton Tutorial 3.7.0.* Phyton Software Foundation

[11] Navlani, A., Fandango, A., & Idris, I. (2021*). Phyton Data Analysis [3rd Ed.]: Perform Data Collection, Data Processing, Wrangling, Visualization, And Model Building Using Python.* Packt, Birmingham.

[12] Gori, T., Sunyoto, A., Al Fatta, H. (2024). Preprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi dan Ilm Komputer (JTIIK),* 11(1), 215-224.

[13] Fadillah, D. H., Octavian, M. R., Marwadin, A., Dhany, M. R., Sari, D. K., Muthiah, E. K., … Primawati, A. (2025). Implementasi Lightgbm dan LLM Gemini pada Website Psychobot untuk Analisis Emosi Saat Bersosial Media. *Jurnal Riset Dan Aplikasi Mahasiswa Informatika Riset Dan Aplikasi Mahasiswa Informatika (JRAMI),* 6(01), 224–233. https://doi.org/10.30998/jrami.v6i01.13500

[14] Ramadhani, R., Ramadhanu, R., & Hidayat, T. (2024). Exploratory Data Analysis (EDA) untuk Mengetahui Distribusi Data Kualitas Susu Sapi. J*urnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 23(1), 68. https://doi.org/10.53513/jis.v23i1.9500

[15] Mayasari, R., Nugraha, B., Juwita, A. R., & Heryana, N. (2023). Analisis Produktifitas Padi di Pulau Sumatera menggunakan Exploratory Data Analysis (EDA). *Jurnal Elektronik Sistem Informasi Unsika*, 1(1), 17–24. (JRAMI), 6(01), 224–233. https://doi.org/10.30998/jrami.v6i01.13500