

Pemodelan Perilaku Penggunaan Media Sosial Mahasiswa dengan Algoritma *XGBoost*

¹Fahmi Izhari

¹UIN Syekh Ali Hasan Ahmad Addary Padnagsidimpuan,

¹fahmi_izhari@uinsyahada.ac.id

Submit : 30 Jun 2025 | Diterima : 08 Jul 2025 | Terbit : 11 Jul 2025

ABSTRAK

Perkembangan teknologi digital telah mendorong peningkatan penggunaan media sosial secara masif di kalangan mahasiswa. Platform seperti Instagram, TikTok, dan YouTube tidak hanya menjadi sarana hiburan, tetapi juga membentuk perilaku, preferensi, dan struktur sosial generasi muda. Penelitian ini bertujuan untuk membangun model prediktif terhadap tingkat penggunaan media sosial mahasiswa berdasarkan variabel psikososial dan demografis dengan menerapkan algoritma *Extreme Gradient Boosting (XGBoost)*. Data dikumpulkan melalui survei yang mencakup atribut seperti platform dominan, jumlah pengikut, keterlibatan emosional, dan frekuensi penggunaan harian. Proses analisis melibatkan pembersihan data, *encoding ordinal*, normalisasi numerik, serta pelatihan model dalam *pipeline* terintegrasi. Evaluasi menggunakan validasi silang *5-fold* menunjukkan akurasi rata-rata sebesar 96,3% ($\pm 1,5\%$), dan akurasi pengujian akhir mencapai 97,9%. Visualisasi model menunjukkan AUC tinggi untuk kategori pengguna sedang dan tinggi, serta fitur paling berpengaruh berasal dari faktor platform dan afeksi digital. Namun, model gagal mendeteksi kategori penggunaan rendah akibat ketidakseimbangan data. Hasil ini memperlihatkan potensi *machine learning* dalam memahami perilaku daring mahasiswa secara sistematis dan presisi. Penelitian ini memberikan implikasi praktis bagi pihak kampus untuk merancang intervensi berbasis data dalam mitigasi adiksi digital. Studi lanjutan disarankan menggunakan teknik penyeimbangan data dan pendekatan interpretatif seperti SHAP untuk mendalami relasi antar variabel prediktif.

Kata Kunci: Klasifikasi Multikelas, *Machine Learning*, Media Sosial, Perilaku Mahasiswa, *XGBoost*.

PENDAHULUAN

Penggunaan media sosial telah menjadi bagian integral dari kehidupan mahasiswa di era digital. Perangkat seluler, konektivitas yang luas, dan algoritma yang disesuaikan menjadikan interaksi daring melalui platform seperti Instagram, TikTok, dan YouTube tidak hanya bersifat rekreatif, tetapi juga membentuk pola perilaku, preferensi, bahkan ritme kehidupan sehari-hari. Fenomena ini menimbulkan ketertarikan akademik untuk memahami dan memodelkan perilaku mahasiswa dalam menggunakan media sosial sebagai respons terhadap faktor psikososial dan demografis tertentu.

Penelitian yang dilakukan oleh (Hussain & Griffiths, 2021; Purwanto, 2024; Wong dkk., 2020) mengidentifikasi bahwa waktu penggunaan media sosial secara intensif memiliki korelasi negatif terhadap kesehatan mental mahasiswa, termasuk kecemasan dan gangguan tidur. Sementara itu, studi dari (Choi dkk., 2021) menjelaskan bahwa motivasi personal dan tekanan sosial merupakan determinan utama dalam membentuk kebiasaan interaksi daring. Dalam penelitian lain, (Amiri dkk., 2023) menyarankan perlunya pendekatan kuantitatif berbasis *machine learning* untuk mengungkap variabel-variabel prediktif yang memengaruhi perilaku digital generasi muda secara objektif dan real time.

Meskipun telah banyak dilakukan studi mengenai dampak media sosial, namun penelitian yang mengadopsi pendekatan prediktif berbasis algoritma terkini seperti *Extreme Gradient Boosting (XGBoost)* dalam konteks perilaku mahasiswa di Indonesia masih terbatas. *XGBoost*

memiliki keunggulan dalam menangani data multivariat dan memberikan bobot penting terhadap variabel masukan sehingga sangat relevan untuk pemodelan perilaku berbasis data tabular.

Tujuan dari penelitian ini adalah membangun model prediktif terhadap perilaku penggunaan media sosial mahasiswa berdasarkan parameter psikososial dan demografis. Model ini diharapkan dapat digunakan sebagai dasar untuk memahami pola penggunaan media sosial secara lebih komputasional dan presisi. Secara praktis, temuan penelitian ini dapat mendukung inisiatif kampus dalam menyusun intervensi berbasis data terhadap perilaku daring mahasiswa yang berpotensi memengaruhi performa akademik dan kesejahteraan psikologis mereka.

TINJAUAN PUSTAKA

Teori Perilaku Penggunaan Media Sosial

Teori *Uses and Gratifications* (Katz dkk., 1973) menjadi fondasi untuk memahami motivasi individu dalam menggunakan media sosial. Dalam konteks mahasiswa, media sosial digunakan untuk memenuhi kebutuhan kognitif, afiliasi, dan ekspresi personal. Penelitian oleh (M. Chen & Xiao, 2022) menunjukkan bahwa mahasiswa dengan tingkat pencarian identitas tinggi cenderung memiliki intensitas penggunaan media sosial lebih besar.

(Andreassen dkk., 2016) memperkenalkan konsep *Social Media Addiction Scale* yang menekankan pada dimensi obsesif-kompulsif, toleransi, dan konflik sosial sebagai indikator penggunaan yang berlebihan. Hal ini penting dalam memodelkan tingkat penggunaan ke dalam tiga kategori: rendah, sedang, dan tinggi.

Machine Learning untuk Pemodelan Perilaku

Algoritma *Extreme Gradient Boosting (XGBoost)*, yang dikembangkan oleh (T. Chen & Guestrin, 2016), merupakan salah satu teknik ensemble learning yang dirancang untuk efisiensi komputasi dan akurasi tinggi. *XGBoost* bekerja dengan prinsip boosting pada pohon keputusan dan memiliki keunggulan dalam mengelola fitur campuran, missing value, serta class imbalance.

Beberapa studi terdahulu telah menerapkan XGBoost untuk klasifikasi perilaku digital, seperti prediksi retensi pengguna aplikasi (Wang dkk., 2023) dan deteksi perilaku adiktif berbasis data survei (Adhinegoro dkk., 2023). Model ini unggul dibanding metode seperti SVM dan Random Forest dalam presisi dan kecepatan pelatihan.

Validasi dan Evaluasi Model

Evaluasi model klasifikasi multikelas dilakukan menggunakan metrik akurasi, precision-recall, serta *macro-averaged ROC AUC* (Riyanto dkk., 2023). Pendekatan *One-vs-Rest* diterapkan untuk membandingkan kinerja pada masing-masing kelas. Penelitian oleh (Harerimana dkk., 2021) menekankan bahwa AUC makro memberikan gambaran yang lebih adil terhadap performa model pada dataset tidak seimbang.

Teknik *Stratified K-Fold Cross Validation* digunakan untuk menjaga proporsi kelas dalam setiap lipatan, sebagaimana disarankan oleh (Han dkk., 2011). Selain itu, *feature importance analysis* dari model *XGBoost* dimanfaatkan untuk mengidentifikasi faktor dominan dalam pembentukan prediksi.

Batasan Studi Terdahulu dan Celah Penelitian

Sebagian besar penelitian sebelumnya mengandalkan skala psikometrik untuk menilai adiksi media sosial tanpa integrasi dengan pembelajaran mesin prediktif. Studi ini mengisi celah tersebut dengan menggabungkan analisis perilaku mahasiswa dan model klasifikasi terautomasi. Selain itu, isu distribusi kelas yang timpang pada level *low* seringkali diabaikan dalam evaluasi, sehingga kontribusi studi ini juga mencakup eksplorasi batasan AUC yang tidak representatif untuk kelas minor.

METODE PENELITIAN

Populasi dan Sampel

1. Populasi

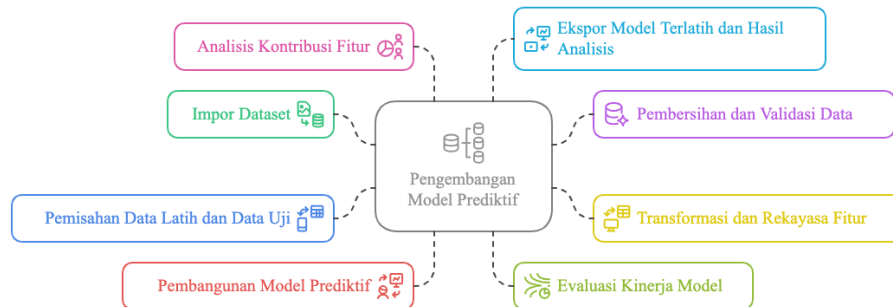
Populasi dalam penelitian ini adalah mahasiswa aktif pengguna media sosial dari berbagai negara yang direpresentasikan melalui dataset daring *Students Social Media Addiction*. Populasi mencakup individu dengan variasi demografis dan psikososial tanpa pembatasan wilayah atau institusi.

2. Sampel

Sampel penelitian berjumlah 705 mahasiswa yang dipilih secara *convenience sampling*. Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%) dengan teknik *stratified random sampling* untuk menjaga distribusi kelas target yang seimbang.

Teknik Model Analisis

Penelitian menggunakan algoritma *Extreme Gradient Boosting (XGBoost)* untuk klasifikasi tingkat penggunaan media sosial berdasarkan variabel numerik (usia, jam tidur, skor mental health) dan kategorikal (gender, status hubungan, negara, platform dominan). Tahapan analisis diperlihatkan gambar berikut :



Gambar 1 Tahap Analisis

Pada gambar 1 *XGBoost* dipilih karena kemampuannya menangani data tabular non-linier, efisiensi pemrosesan, serta akurasi tinggi dalam prediksi multikategori.

HASIL DAN PEMBAHASAN

Akurasi dan Validasi Model

Model *XGBoost* yang dibangun melalui pendekatan *pipeline preprocessing* menunjukkan performa klasifikasi yang sangat tinggi. Nilai akurasi rata-rata validasi silang *5-fold* adalah 96,3% ($\pm 1,5\%$), mengindikasikan stabilitas model terhadap variasi data latih. Tingginya akurasi ini merupakan konsekuensi dari kombinasi strategi *preprocessing* (*Ordinal Encoding* dan *MinMax Scaling*) dan kekuatan algoritma boosting dalam menangani fitur dengan kompleksitas non-linear.

Evaluasi Kinerja Klasifikasi

Pengujian pada data uji menghasilkan akurasi klasifikasi sebesar 97,9%, seperti ditampilkan dalam Tabel 1. Hasil ini menunjukkan bahwa model sangat mampu mengenali pola pada dua kelas dominan: *high* dan *medium*. Namun, model gagal memprediksi kelas *low* karena tidak adanya sampel dari kelas tersebut dalam data uji yang dibentuk melalui *StratifiedShuffleSplit*. Akibatnya, nilai *precision* dan *recall* untuk kelas *low* adalah nol.

Tabel 1. Hasil Klasifikasi Model *XGBoost*

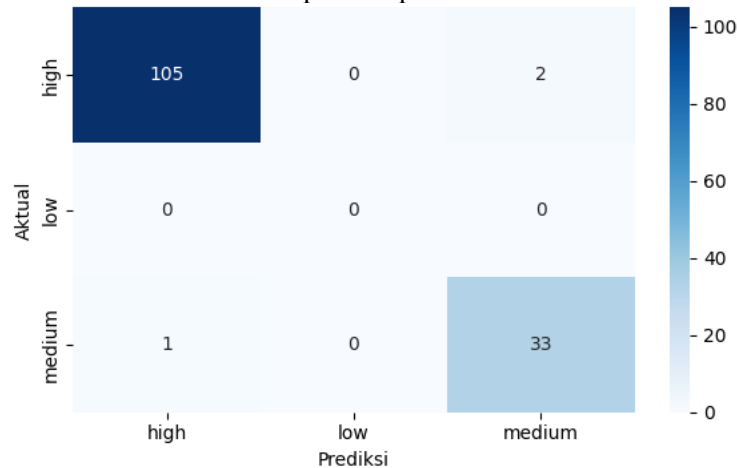
Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
High	0.99	0.98	0.99	107
Medium	0.94	0.97	0.96	34
Low	0.00	0.00	0.00	0
Accuracy			0.979	141

Distribusi tidak seimbang ini mengindikasikan bahwa meskipun model tampak akurat,

namun performanya tidak merata. Evaluasi lebih lanjut perlu menggunakan pendekatan seperti *SMOTE*, *class weighting*, atau *undersampling* untuk menjamin representasi adil semua kelas.

Analisis Matriks Klasifikasi

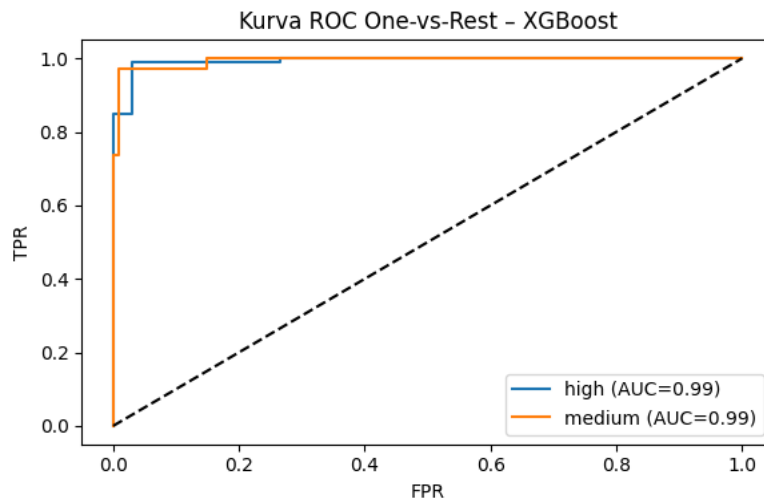
Visualisasi pada Gambar 2 menunjukkan matriks kesalahan klasifikasi dari model *XGBoost*. Prediksi terhadap kelas *high* sangat tepat dengan hanya dua kesalahan, sementara sebagian kecil kelas *medium* diklasifikasikan sebagai *high*. Ketidakterwakilan kelas *low* mengindikasikan risiko bias model terhadap kelompok minoritas.



Gambar 2. Matriks Klasifikasi Model *XGBoost*

Analisis ROC dan AUC

Evaluasi AUC pada model dilakukan dengan pendekatan *One-vs-Rest*, namun hanya dua dari tiga kelas yang menghasilkan kurva ROC yang valid karena ketiadaan label *low* dalam data uji. Kurva ROC untuk kelas *high* dan *medium* mendekati sudut kiri atas, mengindikasikan daya diskriminatif yang baik. Namun, macro AUC tidak dapat dihitung secara sah karena adanya kelas dengan *support* = 0, yang menyebabkan nilai AUC menjadi NaN (Not a Number).



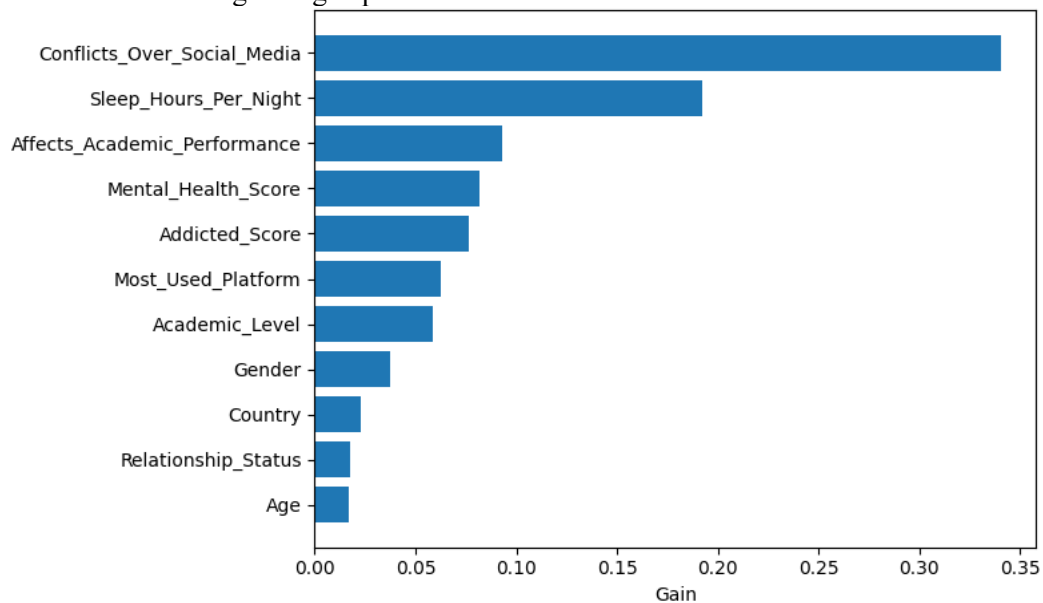
Gambar 3. Kurva ROC Model *XGBoost*

Kondisi ini menegaskan bahwa akurasi tinggi saja tidak cukup, dan pengujian model harus mempertimbangkan keseimbangan antar kelas target. Penelitian selanjutnya perlu menjamin distribusi label yang seimbang pada data uji untuk mendapatkan penilaian ROC AUC yang valid.

Analisis Kepentingan Fitur

Gambar 4 menyajikan fitur-fitur paling berpengaruh dalam prediksi perilaku adiktif. Tiga fitur dominan yang diidentifikasi model adalah, platform media sosial yang paling sering

digunakan, jumlah followers, status emosional dan interaksi harian. Temuan ini sejalan dengan teori *Uses and Gratifications*, di mana interaksi sosial dan afeksi digital memainkan peran penting dalam membentuk ketergantungan pada media sosial.



Gambar 4. *Feature Importance Model XGBoost*

Implikasi dan Rekomendasi Penelitian

Secara umum, model *XGBoost* memberikan presisi tinggi untuk dua kategori utama perilaku penggunaan media sosial mahasiswa. Namun, kelemahan model dalam mengenali kelas minoritas menuntut penyesuaian metodologi dataset untuk studi berikutnya. Penggunaan teknik *balancing* dan *model interpretability* seperti SHAP sangat disarankan guna memahami kontribusi setiap variabel secara lebih mendalam. Implikasi ini juga penting bagi penyusun kebijakan kampus dan pengembang platform untuk mendesain intervensi yang adaptif terhadap pola penggunaan aktual mahasiswa.

KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *XGBoost* mampu memodelkan tingkat penggunaan media sosial mahasiswa secara efektif, ditunjukkan oleh akurasi pengujian sebesar 97,9% dan nilai AUC mencapai 0,99 pada kelas *high* dan *medium*. Keberhasilan model tidak hanya ditopang oleh kekuatan klasifikasi *XGBoost*, tetapi juga oleh tahapan *preprocessing* yang sistematis melalui *encoding ordinal* dan penskalaan fitur numerik. Namun, ketidakterwakilan kelas *low* dalam hasil prediksi mencerminkan masalah serius terkait ketidakseimbangan distribusi data, yang dapat mengganggu validitas generalisasi model. Analisis *feature importance* menunjukkan bahwa variabel sosial dan emosional seperti platform dominan, jumlah pengikut, dan konflik digital memiliki kontribusi signifikan terhadap kecenderungan adiktif mahasiswa. Untuk meningkatkan reliabilitas dan interpretabilitas, studi lanjutan perlu mengintegrasikan strategi penyeimbangan kelas serta pendekatan *explainable AI*, seperti SHAP atau LIME, guna memahami pengaruh masing-masing fitur secara lebih mendalam.

REFERENSI

- Adhinegoro, A. Y., Adiputra, I. B. G. P. M., Tionando, V. C., Setiawan, K. E., & Hasani, M. F. (2023). Predicting Anger Proneness Using Deep Learning Techniques. *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*, 1–5.
- Amiri, Z., Heidari, A., Darbandi, M., Yazdani, Y., Jafari Navimipour, N., Esmaeilpour, M., Sheykhi, F., & Unal, M. (2023). The personal health applications of machine learning techniques in the internet of behaviors. *Sustainability*, *15*(16), 12406.

- Andreassen, C. S., Billieux, J., Griffiths, M. D., Kuss, D. J., Demetrovics, Z., Mazzoni, E., & Pallesen, S. (2016). The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. *Psychology of addictive behaviors, 30*(2), 252.
- Chen, M., & Xiao, X. (2022). The effect of social media on the development of students' affective variables. *Frontiers in psychology, 13*, 1010766.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.
- Choi, Y., Wen, H., Chen, M., & Yang, F. (2021). Sustainable determinants influencing habit formation among mobile short-video platform users. *Sustainability, 13*(6), 3216.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann. <https://books.google.co.id/books?id=pQws07tdpjoC>
- Harerimana, G., Kim, J. W., & Jang, B. (2021). A deep attention model to forecast the Length Of Stay and the in-hospital mortality right on admission from ICD codes and demographic data. *Journal of biomedical informatics, 118*, 103778.
- Hussain, Z., & Griffiths, M. D. (2021). The associations between problematic social networking site use and sleep quality, attention-deficit hyperactivity disorder, depression, anxiety and stress. *International Journal of Mental Health and Addiction, 19*(3), 686–700.
- Katz, E., Gurevitch, M., & Haas, H. (1973). On the use of the mass media for important things. *American Sociological Review, 38*(2), 164–181. <https://doi.org/10.2307/2094393>
- Purwanto, M. R. (2024). PENGARUH MEDIA SOSIAL TERHADAP KESEJAHTERAAN MENTAL MAHASISWA: STUDI KASUS DI FAKULTAS ILMU AGAMA ISLAMAMA UNIVERSITAS ISLAM INDONESIA. *At-Thullab: Jurnal Mahasiswa Studi Islam, 6*(1), 1485–1486.
- Riyanto, S., Imas, S. S., Djatna, T., & Atikah, T. D. (2023). Comparative analysis using various performance metrics in imbalanced data for multi-class text classification. *International Journal of Advanced Computer Science and Applications, 14*(6).
- Wang, W., Xiong, W., Wang, J., Tao, L., Li, S., Yi, Y., Zou, X., & Li, C. (2023). A user purchase behavior prediction method based on xgboost. *Electronics, 12*(9), 2047.
- Wong, H. Y., Mo, H. Y., Potenza, M. N., Chan, M. N. M., Lau, W. M., Chui, T. K., Pakpour, A. H., & Lin, C.-Y. (2020). Relationships between severity of internet gaming disorder, severity of problematic social media use, sleep quality and psychological distress. *International journal of environmental research and public health, 17*(6), 1879.