

Jenis artikel	Artikel Penelitian	ID artikel	TEK-2026-0012
Tautan permanen	https://doi.org/10.xxxx/teknovia.xx.xx	Riwayat artikel	Masuk: 18-07-2026 Revisi: 09-10-2026
Keputusan & terbit	Diterima: 22-10-2026 Terbit: 25-10-2026	Lisensi	Creative Commons Attribution 4.0 International

Komparasi Random Forest Dan Xgboost Untuk Klasifikasi Overall_Impact Pada Dataset Penggunaan Media Sosial Pelajar Dan Mahasiswa

Comparative analysis of Random Forest and xgboost for Overall_Impact classification in a student social media use dataset

Diky Jonathan¹, Bambang Wijonarko², Andronias Siregar^{3*}

¹Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

ORCID: [belum dicantumkan untuk seluruh penulis]

***Penulis korespondensi: jonathandiky2805@gmail.com**

Abstract

Social media use is associated with differences in students' sleep, academic activity, and mental well-being. This study compares Random Forest and Extreme Gradient Boosting (XGBoost) for classifying the pre-existing Overall_Impact label in a public Kaggle dataset into Negative, Neutral, and Positive classes. The dataset contains 1,685 records. Preprocessing included identifier removal, categorical label encoding, exclusion of Affects_Academic_Performance and Mental_Health_Score as leakage-prone target proxies, and stratified train-test splitting. Three split scenarios (70:30, 60:40, and 80:20) were evaluated using random_state=42, and 70:30 was selected for the primary analysis because it produced the strongest overall results for both models. XGBoost achieved 95.45% accuracy, 95.61% weighted precision, 95.45% weighted recall, 95.50% weighted F1-score, and 0.9934 ROC-AUC; Random Forest achieved 85.57%, 87.95%, 85.57%, 85.92%, and 0.9791, respectively. XGBoost produced 23 misclassifications compared with 73 for Random Forest. Sleep_Hours_Per_Night and Avg_Daily_Usage_Hours were the two most influential features in both models. These findings are limited to the dataset and experimental configuration used and should not be interpreted as direct causal measurement of social media impact.

Keywords: social media; students; Random Forest; XGBoost; Overall_Impact; classification

Abstrak

Penggunaan media sosial berkaitan dengan variasi pola tidur, aktivitas akademik, dan kondisi kesejahteraan pelajar maupun mahasiswa. Penelitian ini membandingkan Random Forest dan Extreme Gradient Boosting (XGBoost) untuk mengklasifikasikan label Overall_Impact yang telah tersedia pada dataset publik Kaggle ke dalam kelas Negative, Neutral, dan Positive. Dataset terdiri atas 1.685 data. Prapemrosesan mencakup penghapusan identifier, label encoding variabel kategorikal, pengeluan Affects_Academic_Performance dan Mental_Health_Score sebagai fitur yang berpotensi menjadi proxy target, serta stratified train-test split. Tiga skenario pembagian data, yaitu 70:30, 60:40, dan 80:20, diuji dengan random_state=42. Skenario 70:30 dipilih untuk analisis utama karena memberikan hasil keseluruhan terbaik pada kedua model. XGBoost memperoleh accuracy 95,45%, weighted precision 95,61%, weighted recall 95,45%, weighted F1-score 95,50%,

dan ROC-AUC 0,9934; Random Forest memperoleh 85,57%, 87,95%, 85,57%, 85,92%, dan 0,9791. XGBoost menghasilkan 23 kesalahan klasifikasi, sedangkan Random Forest 73 kesalahan. Sleep_Hours_Per_Night dan Avg_Daily_Usage_Hours menjadi dua fitur paling berpengaruh pada kedua model. Temuan ini dibatasi pada dataset dan konfigurasi eksperimen yang digunakan dan tidak dimaknai sebagai pengukuran kausal dampak media sosial secara langsung.

Kata kunci: media sosial; pelajar; mahasiswa; Random Forest; XGBoost; Overall_Impact

1. Pendahuluan

Media sosial telah menjadi bagian penting dalam aktivitas pelajar dan mahasiswa, baik sebagai sarana komunikasi, pencarian informasi, pembelajaran, hiburan, maupun pembentukan jejaring sosial. Pemanfaatannya dapat memberikan manfaat berupa akses informasi yang cepat dan peluang pengembangan diri. Namun, penggunaan yang tidak terkendali juga dikaitkan dengan berkurangnya konsentrasi belajar, terganggunya pola tidur, menurunnya produktivitas akademik, serta tekanan terhadap kesehatan mental [1], [2]. Kondisi tersebut membuat analisis dampak media sosial memerlukan pendekatan yang tidak hanya deskriptif, tetapi juga mampu mengenali pola dari banyak atribut secara bersamaan.

Machine learning menyediakan teknik klasifikasi untuk memetakan pola perilaku penggunaan media sosial menjadi kategori dampak tertentu. Random Forest menggabungkan banyak decision tree melalui mekanisme bagging dan majority voting sehingga relatif tahan terhadap overfitting [3]. Sebaliknya, XGBoost membangun pohon secara bertahap melalui boosting, mengoreksi kesalahan model sebelumnya, dan menerapkan regularisasi untuk meningkatkan kemampuan generalisasi [4]. Kedua metode telah menunjukkan performa baik pada berbagai masalah klasifikasi, tetapi keunggulan relatifnya dipengaruhi oleh distribusi kelas, karakteristik fitur, serta kedekatan fitur terhadap target.

Penelitian terdahulu melaporkan bahwa XGBoost sering mengungguli Random Forest pada klasifikasi stunting dan analisis sentimen [5], [6]. XGBoost juga menghasilkan akurasi tinggi pada pemodelan perilaku penggunaan media sosial mahasiswa, tetapi masih menghadapi kesulitan pada kelas dengan jumlah data terbatas [7]. Studi-studi tersebut belum secara khusus membandingkan kedua algoritma untuk klasifikasi tiga kelas dampak media sosial pada gabungan kelompok pelajar dan mahasiswa dengan pengendalian data leakage dan analisis feature importance.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan membandingkan performa Random Forest dan XGBoost dalam mengklasifikasikan label Overall_Impact yang telah tersedia pada dataset ke dalam kelas Negative, Neutral, dan Positive, bukan mengukur dampak media sosial secara langsung. Kontribusi penelitian meliputi: (1) evaluasi komparatif menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan ROC-AUC; (2) pengujian tiga skenario pembagian data; serta (3) identifikasi faktor yang paling berpengaruh melalui feature importance.

2. Kajian Terkait dan Landasan Teori

2.1 Landasan Konseptual atau Teknis

Random Forest merupakan metode ensemble berbasis bagging. Sejumlah decision tree dilatih menggunakan sampel dan subset fitur yang berbeda, kemudian kelas akhir ditentukan dari nilai yang paling sering diprediksi oleh seluruh pohon. Pendekatan ini dapat menurunkan varians model tunggal dan meningkatkan ketahanan terhadap noise [3]. XGBoost merupakan metode gradient boosting yang menambahkan pohon baru secara iteratif untuk meminimalkan fungsi objektif. Regularisasi, learning rate, dan pengendalian kompleksitas pohon membantu menghasilkan model yang efisien sekaligus akurat [4]. Secara umum, klasifikasi merupakan bagian dari supervised learning yang memanfaatkan data berlabel untuk mempelajari pola prediksi [8]. Python juga banyak digunakan sebagai bahasa pemrograman untuk implementasi komputasi dan pembelajaran mesin karena sintaksnya relatif mudah dipahami [9]. Penerapan metode pembelajaran mesin pada berbagai tugas prediksi menunjukkan bahwa performa model sangat dipengaruhi oleh karakteristik data dan proses pemodelan [10].

2.2 Sintesis Penelitian Terdahulu

Tabel 1. Sintesis penelitian terdahulu

Penelitian	Objek dan metode	Temuan utama	Keterbatasan
Hadisuwarno et al. [5]	Klasifikasi stunting; XGBoost vs Random Forest	XGBoost unggul pada data seimbang	Konteks kesehatan, bukan perilaku media sosial
Izhari [7]	Perilaku media sosial mahasiswa; XGBoost	Akurasi pengujian 97,9%	Kelas rendah sulit karena ketidakseimbangan
Wati et al. [6]	Analisis sentimen; XGBoost vs Random Forest	XGBoost 68%, Random Forest 63%	Objek berupa teks sentimen
Penelitian ini	Label Overall_Impact pada dataset penggunaan media sosial pelajar dan mahasiswa; dua model ensemble	Komparasi tiga kelas, tiga skenario split, ROC-AUC, confusion matrix, dan feature importance	Dataset sekunder; aturan pembentukan label tidak didokumentasikan secara rinci

Source: Diolah oleh penulis berdasarkan [5]-[7] (2026).

Studi internasional juga melaporkan keterkaitan penggunaan media sosial dengan durasi tidur, performa dan keterlibatan akademik, kondisi psikologis, serta pencapaian akademik mahasiswa dan remaja [11]–[15]. Temuan tersebut memperkuat relevansi penggunaan atribut perilaku dan akademik dalam analisis, namun tidak digunakan untuk menyimpulkan hubungan kausal pada dataset penelitian ini.

3. Metode Penelitian

3.1 Desain, Data, dan Objek Penelitian

Penelitian menggunakan desain eksperimen komparatif dengan dataset sekunder publik berjudul “Impact of Social Media on Health” yang dipublikasikan oleh Sume Akash melalui Kaggle. Dataset terdiri atas 1.685 responden pelajar dan mahasiswa dengan variabel awal Student_ID, Age, Gender, Academic_Level, Country, Avg_Daily_Usage_Hours, Most_Used_Platform, Affects_Academic_Performance, Sleep_Hours_Per_Night, Mental_Health_Score, dan Overall_Impact. Overall_Impact merupakan label kategorikal yang sudah tersedia pada dataset sumber dengan tiga kelas, yaitu Negative, Neutral, dan Positive; penelitian ini tidak membentuk atau melakukan relabeling terhadap kelas tersebut. Dokumentasi dataset yang digunakan dalam penelitian tidak menjelaskan formula atau aturan operasional rinci pembentukan Overall_Impact, sehingga label diperlakukan sebagai ground truth bawaan dataset. Konsekuensinya, hasil penelitian diinterpretasikan sebagai kinerja klasifikasi terhadap label tersebut, bukan sebagai pengukuran langsung atau kausal atas dampak media sosial.

Tabel 2. Atribut yang digunakan dalam pemodelan

Atribut	Jenis	Keterangan
Age	Numerik	Usia responden
Gender	Kategorikal	Jenis kelamin
Academic_Level	Kategorikal	High School, Undergraduate, atau Graduate
Country	Kategorikal	Negara responden

Atribut	Jenis	Keterangan
Avg_Daily_Usage_Hours	Numerik	Rata-rata penggunaan media sosial per hari
Most_Used_Platform	Kategorikal	Platform yang paling sering digunakan
Sleep_Hours_Per_Night	Numerik	Durasi tidur rata-rata per malam
Overall_Impact	Target	Negatif, netral, atau positif

Source: Dataset publik Kaggle, diolah oleh penulis (2026).

Student_ID dihapus karena hanya berfungsi sebagai identitas. Variabel kategorikal dikonversi menjadi numerik menggunakan LabelEncoder dari scikit-learn dengan fit_transform pada masing-masing kolom kategorikal sebelum train-test split; setiap kategori dipetakan menjadi integer berdasarkan kelas yang dikenali encoder. Pada target Overall_Impact, pemetaan yang dihasilkan adalah Negative=0, Neutral=1, dan Positive=2. Mapping encoder disimpan agar transformasi tetap konsisten. Affects_Academic_Performance dan Mental_Health_Score kemudian dikeluarkan dari fitur model utama karena secara konseptual merepresentasikan aspek yang sangat dekat dengan target Overall_Impact dan berpotensi bertindak sebagai proxy target. Analisis awal juga menunjukkan Mental_Health_Score memiliki feature importance 0,501551 dan Affects_Academic_Performance 0,069278; nilai tersebut digunakan sebagai indikasi pendukung, bukan satu-satunya dasar penghapusan. Setelah penghapusan, tujuh fitur yang digunakan adalah Age, Gender, Academic_Level, Country, Avg_Daily_Usage_Hours, Most_Used_Platform, dan Sleep_Hours_Per_Night. Pemeriksaan data memastikan tidak terdapat missing value.

Tabel 3. Distribusi kelas Overall_Impact

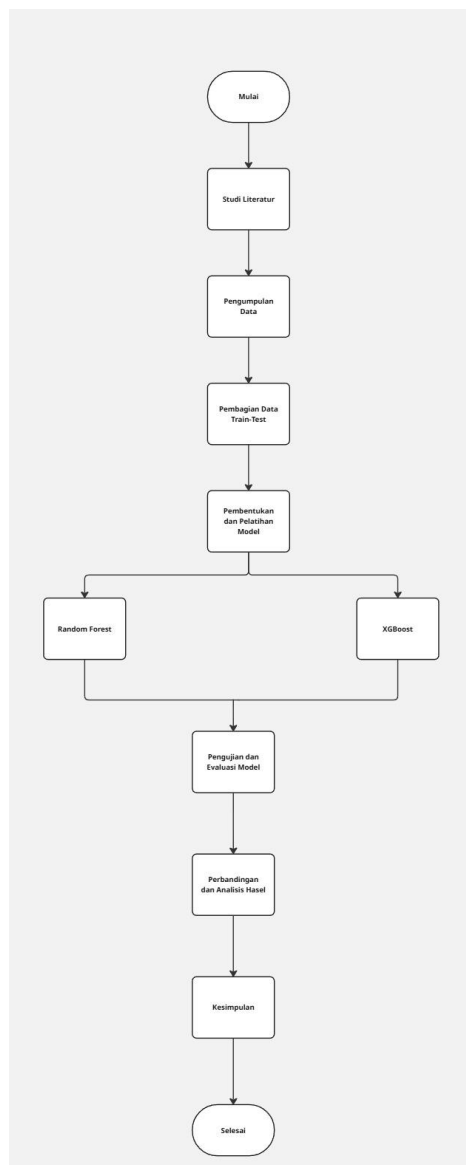
Kelas Overall_Impact	Jumlah Data	Persentase
Negative	930	55,19%
Neutral	260	15,43%
Positive	495	29,38%
Total	1.685	100,00%

Source: Hasil pengolahan data penelitian (2026).

Berdasarkan Tabel 3, kelas Negative memiliki jumlah data terbesar, yaitu 930 data (55,19%), diikuti Positive sebanyak 495 data (29,38%), sedangkan Neutral berjumlah 260 data (15,43%). Distribusi tersebut menunjukkan ketidakseimbangan kelas. Karena itu, pembagian data menggunakan stratify=y untuk mempertahankan proporsi kelas, sedangkan ketidakseimbangan pada proses pelatihan ditangani dengan class_weight=balanced pada Random Forest dan sample_weight berbasis balanced class weight pada XGBoost.

3.2 Prosedur atau Metode yang Diusulkan

Alur penelitian terdiri atas studi literatur, pengumpulan data, preprocessing data, pembagian data train-test, pembentukan dan pelatihan model Random Forest dan XGBoost, pengujian dan evaluasi model, perbandingan dan analisis hasil, serta penarikan kesimpulan. Setiap tahap dilakukan secara berurutan untuk menjaga konsistensi eksperimen dan memastikan proses penelitian dapat ditelusuri. Tiga skenario stratified train-test split diuji, yaitu 70:30, 60:40, dan 80:20 dengan random_state=42. Skenario 70:30 digunakan sebagai skenario utama karena menghasilkan performa keseluruhan tertinggi pada kedua algoritma.



Gambar 1. Alur penelitian

Source: Diolah oleh penulis (2026).

Konfigurasi model ditetapkan secara manual untuk menyelaraskan parameter utama yang memiliki fungsi konseptual serupa pada kedua algoritma. Penelitian tidak melakukan GridSearchCV, RandomizedSearchCV, atau tuning otomatis; parameter yang tidak dicantumkan menggunakan nilai default pustaka masing-masing.

Tabel 4. Konfigurasi hyperparameter Random Forest dan XGBoost

Parameter	Random Forest	XGBoost
n_estimators	200	200
max_depth	6	6
random_state	42	42
max_samples / subsample	0,9	0,9
max_features / colsample_bytree	$\sqrt{p/p} = 0,3780$	$\sqrt{p/p} = 0,3780$
Penanganan imbalance	class_weight=balanced	sample_weight=balanced
learning_rate / reg_alpha / reg_lambda	—	0,1 / 0 / 1
eval_metric	—	mlogloss

Source: Konfigurasi eksperimen penelitian (2026).

3.3 Evaluasi dan Teknik Analisis

Evaluasi menggunakan accuracy, precision, recall, dan F1-score. Nilai precision, recall, dan F1-score utama dihitung dengan weighted average karena distribusi kelas tidak seimbang, sedangkan ROC-AUC dihitung pada klasifikasi multikelas menggunakan One-vs-Rest (OvR) dengan macro average. Confusion matrix digunakan untuk menganalisis pola kesalahan antarkelas, sedangkan feature importance digunakan untuk menilai kontribusi relatif setiap fitur.

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (1)$$

4. Hasil dan Pembahasan

4.1 Hasil Utama

Tabel 5. Perbandingan performa berdasarkan skenario pembagian data

Training:Testing	Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
70:30	XGBoost	95,45%	95,61%	95,45%	95,50%	0,9934
70:30	Random Forest	85,57%	87,95%	85,57%	85,92%	0,9791
60:40	XGBoost	94,21%	94,40%	94,21%	94,28%	0,9864
60:40	Random Forest	84,27%	86,96%	84,27%	84,63%	0,9699
80:20	XGBoost	94,36%	94,57%	94,36%	94,42%	0,9926
80:20	Random Forest	84,87%	87,37%	84,87%	85,28%	0,9740

Source: Hasil pengolahan data penelitian (2026).

Tabel 5 menunjukkan bahwa XGBoost mengungguli Random Forest pada ketiga skenario. Skenario 70:30 dipilih sebagai skenario utama karena menghasilkan accuracy, F1-score, dan ROC-AUC tertinggi pada kedua algoritma, sehingga pembahasan berikutnya menggunakan 1.179 data latih dan 506 data uji.

Tabel 6. Perbandingan performa model pada data uji skenario 70:30

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	85,57%	87,95%	85,57%	85,92%	0,9791
XGBoost	95,45%	95,61%	95,45%	95,50%	0,9934

Source: Hasil pengolahan data penelitian (2026).

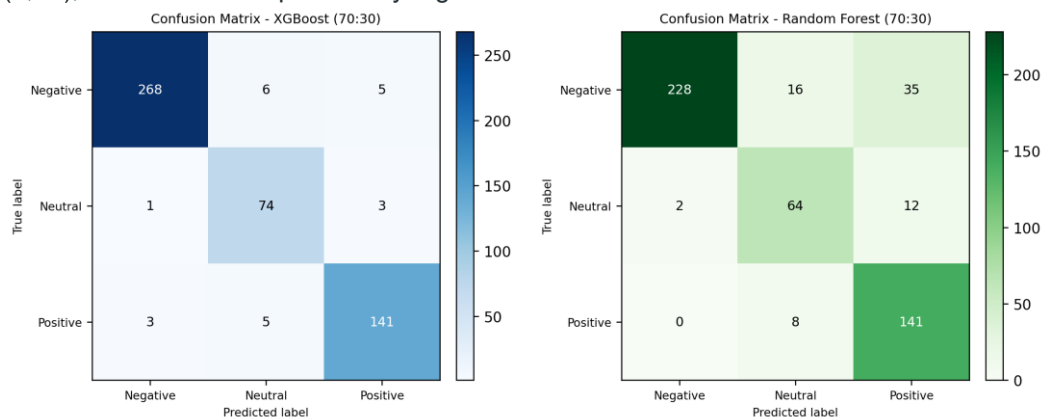
Pada dataset dan konfigurasi eksperimen penelitian ini, XGBoost menghasilkan performa lebih tinggi pada seluruh metrik utama. Dibandingkan Random Forest, selisih accuracy dan recall sebesar 9,88 poin persentase, precision 7,66 poin, dan F1-score 9,58 poin. Hasil tersebut menunjukkan keunggulan XGBoost pada eksperimen ini, tetapi tidak digunakan untuk menyimpulkan bahwa XGBoost selalu lebih baik pada dataset atau konfigurasi lain.

Tabel 7. Precision, recall, dan F1-score per kelas pada skenario 70:30

Model	Kelas	Precision	Recall	F1-score	Support
XGBoost	Negative	0,99	0,96	0,97	279
XGBoost	Neutral	0,87	0,95	0,91	78
XGBoost	Positive	0,95	0,95	0,95	149
Random Forest	Negative	0,99	0,82	0,90	279
Random Forest	Neutral	0,73	0,82	0,77	78
Random Forest	Positive	0,75	0,95	0,84	149

Source: Hasil pengolahan data penelitian (2026).

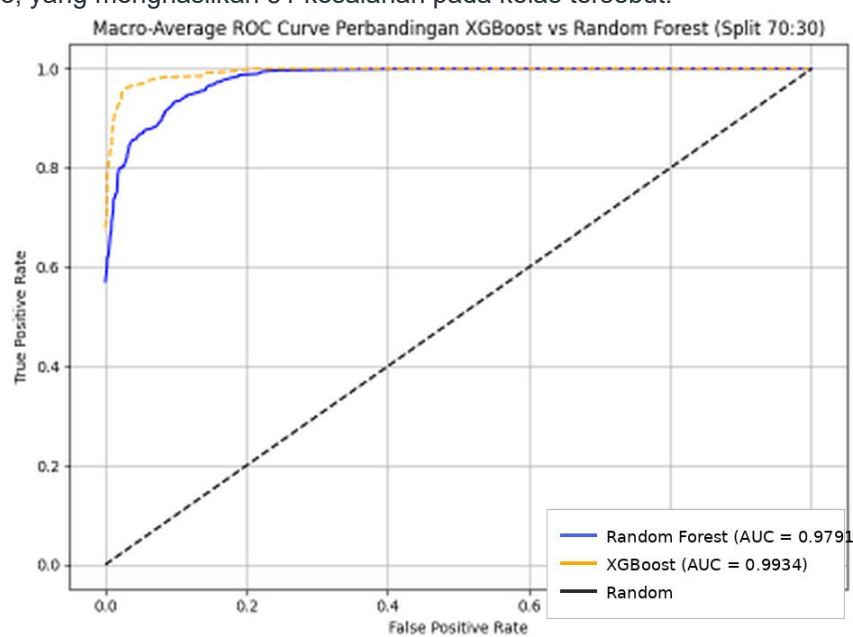
Pada XGBoost, kelas Neutral memiliki precision terendah (0,87) tetapi recall 0,95, sedangkan kelas Negative dan Positive memiliki F1-score masing-masing 0,97 dan 0,95. Pada Random Forest, penurunan paling jelas terdapat pada kelas Neutral (F1-score 0,77) dan Positive (0,84), terutama akibat precision yang lebih rendah.



Gambar 2. Confusion matrix XGBoost (kiri) dan Random Forest (kanan)

Source: Hasil pengolahan data penelitian (2026).

Confusion matrix menunjukkan bahwa XGBoost menghasilkan 483 prediksi benar dan 23 kesalahan dari 506 data uji, sedangkan Random Forest menghasilkan 433 prediksi benar dan 73 kesalahan. Pada XGBoost, prediksi benar terdiri atas 268 data Negative, 74 Neutral, dan 141 Positive. Random Forest menghasilkan 228 data Negative, 64 Neutral, dan 141 Positive yang diprediksi benar. Perbedaan terbesar terlihat pada kemampuan Random Forest mengenali kelas Negative, yang menghasilkan 51 kesalahan pada kelas tersebut.

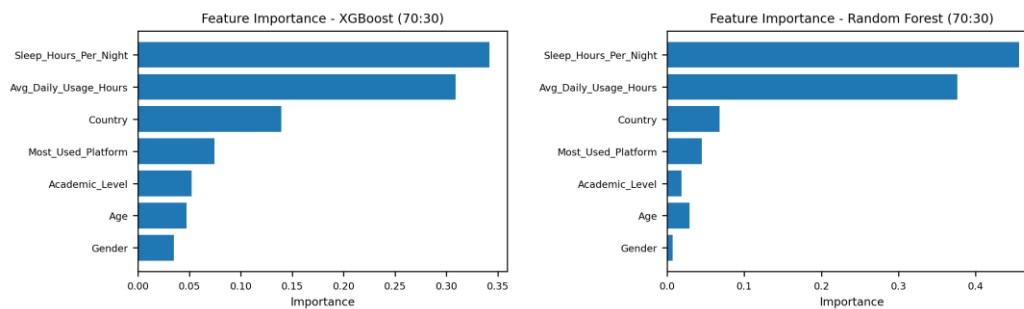


Gambar 3. Kurva ROC macro-average Random Forest dan XGBoost

Source: Hasil pengolahan data penelitian (2026).

ROC-AUC OvR macro-average pada data uji mencapai 0,9934 untuk XGBoost dan 0,9791 untuk Random Forest. Kedua nilai menunjukkan kemampuan diskriminasi antarkelas yang tinggi pada dataset ini, dengan XGBoost tetap menghasilkan nilai lebih tinggi. Kurva pada Gambar 3 berasal dari probabilitas prediksi pada skenario 70:30.

4.2 Interpretasi, Perbandingan, dan Implikasi



Gambar 4. Feature importance XGBoost (kiri) dan Random Forest (kanan)

Source: Hasil pengolahan data penelitian (2026).

Pada XGBoost, feature importance tertinggi terdapat pada `Sleep_Hours_Per_Night` (0,3420) dan `Avg_Daily_Usage_Hours` (0,3092). Random Forest menunjukkan pola yang sama dengan nilai masing-masing 0,4566 dan 0,3763. `Country`, `Most_Used_Platform`, `Academic_Level`, `Age`, dan `Gender` memiliki kontribusi relatif lebih rendah. Kesamaan peringkat pada kedua model menunjukkan bahwa pola tidur dan intensitas penggunaan media sosial merupakan sinyal prediktif yang konsisten pada dataset ini, tanpa menyiratkan hubungan kausal.

4.3 Keterbatasan dan Ancaman terhadap Validitas

Temuan bahwa XGBoost lebih unggul sejalan dengan penelitian komparatif pada stunting dan analisis sentimen [5], [6]. Mekanisme boosting memungkinkan model memperbaiki residual secara bertahap, sedangkan regularisasi membantu mengontrol kompleksitas. Namun, performa tinggi tidak serta-merta menunjukkan bahwa model dapat digunakan sebagai alat diagnosis psikologis. Model hanya merepresentasikan pola pada dataset dan lebih tepat digunakan sebagai alat bantu analisis awal untuk edukasi atau pemantauan perilaku penggunaan media sosial.

Penelitian memiliki beberapa keterbatasan. Pertama, data bersifat sekunder dan `Overall_Impact` merupakan label bawaan dataset; dokumentasi yang digunakan tidak memberikan aturan operasional rinci pembentukan label, sehingga hasil hanya merefleksikan kemampuan model memprediksi label tersebut. Kedua, kelas `Neutral` hanya mencakup 15,43% data sehingga distribusi target tidak seimbang, meskipun stratification dan pembobotan kelas telah diterapkan. Ketiga, evaluasi utama menggunakan satu split 70:30 setelah perbandingan tiga skenario dan belum menggunakan cross-validation, confidence interval, atau uji signifikansi inferensial. Keempat, `LabelEncoder` pada fitur nominal diterapkan sebelum train-test split dan dapat memperkenalkan urutan semu; penelitian lanjutan disarankan melakukan fit encoder hanya pada data latih dan mempertimbangkan one-hot encoding. Kelima, konfigurasi hyperparameter ditetapkan manual tanpa tuning otomatis. Keenam, feature importance berbasis model menunjukkan kontribusi prediktif, bukan arah pengaruh maupun hubungan kausal.

5. Kesimpulan

Random Forest dan XGBoost mampu mengklasifikasikan label `Overall_Impact` ke dalam kelas `Negative`, `Neutral`, dan `Positive` pada dataset penelitian. Pada skenario utama 70% data latih dan 30% data uji, XGBoost memperoleh accuracy 95,45%, precision 95,61%, recall 95,45%, F1-score 95,50%, dan ROC-AUC 0,9934. Random Forest memperoleh accuracy 85,57%, precision 87,95%, recall 85,57%, F1-score 85,92%, dan ROC-AUC 0,9791. XGBoost menghasilkan 23 kesalahan klasifikasi, sedangkan Random Forest 73 kesalahan.

`Sleep_Hours_Per_Night` dan `Avg_Daily_Usage_Hours` menjadi dua fitur dengan kontribusi prediktif tertinggi pada kedua model. Dengan demikian, pada dataset dan konfigurasi eksperimen penelitian ini, XGBoost menghasilkan performa klasifikasi yang lebih tinggi dibandingkan Random Forest. Kesimpulan ini tidak digeneralisasi ke seluruh kasus klasifikasi dampak media sosial. Penelitian lanjutan dapat menggunakan cross-validation, preprocessing yang di-fit hanya pada data latih, one-hot encoding untuk variabel nominal, hyperparameter tuning, SHAP atau permutation importance, serta dataset yang memiliki definisi operasional label yang terdokumentasi secara lebih rinci.

Pernyataan Tambahan

Kontribusi penulis: Diky Jonathan menyusun konsep, pengolahan data, eksperimen, analisis, dan draf artikel. Bambang Wijonarko dan Andronias Siregar memberikan supervisi, validasi metodologi, serta telaah naskah.

Konflik kepentingan: Para penulis menyatakan tidak terdapat konflik kepentingan.

Pendanaan: Penelitian ini tidak menerima pendanaan khusus dari lembaga pendanaan publik, komersial, maupun nirlaba.

Ketersediaan data: Data penelitian berasal dari dataset publik berjudul *“Impact of Social Media on Health”* yang dipublikasikan oleh Sume Akash melalui Kaggle dan dapat diakses pada <https://www.kaggle.com/datasets/sumeakash/impact-of-social-media-on-health>.

Etik penelitian: Penelitian menggunakan data sekunder publik dan tidak melakukan interaksi langsung dengan partisipan.

6. Referensi

- [1] S. Sarmini, A. Putri, C. Maria, L. Syahrias, and I. Mustika, “Penyuluhan mental health upaya untuk meningkatkan kesehatan mental remaja,” MONSU'ANI TANO Jurnal Pengabdian Masyarakat, vol. 6, no. 1, pp. 154–161, 2023, <https://doi.org/10.32529/tano.v6i1.2400>.
- [2] M. Rahman, I. Nursyabilah, P. Astuti, M. I. Syam, S. Mukramin, and W. O. I. Kurnawati, “Pemanfaatan media sosial sebagai media pembelajaran,” Journal on Education, vol. 5, no. 3, pp. 10646–10653, 2023, <https://doi.org/10.31004/joe.v5i3.1890>.
- [3] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, “Analisis sentimen ulasan aplikasi Dana dengan metode Random Forest,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 6, no. 9, pp. 4305–4313, 2022. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>.
- [4] M. D. Maulana, A. I. Hadiana, and F. R. Umbara, “Algoritma XGBoost untuk klasifikasi kualitas air minum,” JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 5, pp. 3251–3256, 2024, <https://doi.org/10.36040/jati.v7i5.7308>.
- [5] M. A. H. Hadisuwarno, K. T. Martono, and E. Adriono, “Komparasi performa model machine learning algoritma XGBoost dan Random Forest pada studi kasus mendeteksi stunting,” AITI: Jurnal Teknologi Informasi, vol. 22, no. 2, pp. 266–278, 2025, <https://doi.org/10.24246/aiti.v22i2.266-278>.
- [6] H. L. Wati, M. T. I. R. Wiradinata, N. Anggraeni, S. Kolbiah, U. Hendar, and N. Agustina, “Perbandingan algoritma Random Forest dan XGBoost untuk analisis sentimen publik terhadap Rumah Makan Payakumbuah,” NARATIF: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika, vol. 7, no. 1, pp. 64–71, 2025, <https://doi.org/10.53580/naratif.v7i1.307>.
- [7] F. Izhari, “Pemodelan perilaku penggunaan media sosial mahasiswa dengan algoritma XGBoost,” Jurnal Minfo Polgan, vol. 14, no. 1, pp. 1303–1308, 2025, <https://doi.org/10.33395/jmp.v14i1.15030>.
- [8] R. S. Nurhalizah, R. Ardianto, and P. Purwono, “Analisis supervised dan unsupervised learning pada machine learning: Systematic literature review,” Jurnal Ilmu Komputer dan Informatika, vol. 4, no. 1, pp. 61–72, 2024, <https://doi.org/10.54082/jiki.168>.
- [9] M. H. Maulana, “Python bahasa pemrograman yang ramah bagi pemula,” JISCO: Journal of Information System and Computing, vol. 2, no. 2, pp. 73–78, 2024, <https://doi.org/10.30631/jisco.v2i2.105>.

- [10] W. Hastomo, N. Aini, A. S. B. Karno, and L. M. R. Rere, "Metode pembelajaran mesin untuk memprediksi emisi manure management," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 2, pp. 131–139, 2022, <https://doi.org/10.22146/jnteti.v11i2.2586>.
- [11] M. Kolhar, R. N. A. Kazi, and A. Alameen, "Effect of social media use on learning, social interactions, and sleep duration among university students," *Saudi Journal of Biological Sciences*, vol. 28, no. 4, pp. 2216–2222, 2021, <https://doi.org/10.1016/j.sjbs.2021.01.010>.
- [12] M. Shafiq and K. Parveen, "Social media usage: Analyzing its effect on academic performance and engagement of higher education students," *International Journal of Educational Development*, vol. 98, Art. no. 102738, 2023, <https://doi.org/10.1016/j.ijedudev.2023.102738>.
- [13] M. Gull and B. R. Sravani, "Do screen time and social media use affect sleep patterns, psychological health, and academic performance among adolescents? Evidence from bibliometric analysis," *Children and Youth Services Review*, vol. 164, Art. no. 107886, 2024, <https://doi.org/10.1016/j.childyouth.2024.107886>.
- [14] D. J. E. Berdida, "Nursing students' personality traits, sleep quality, social media addiction, and academic performance: A multi-site structural equation model analysis," *Journal of Professional Nursing*, vol. 56, pp. 26–35, 2025, <https://doi.org/10.1016/j.profnurs.2024.11.004>.
- [15] N. Salari, H. Zarei, S. Rasoulpoor, H. Ghasemi, A. Hosseinian-Far, and M. Mohammadi, "The impact of social networking addiction on the academic achievement of university students globally: A meta-analysis," *Public Health in Practice*, vol. 9, Art. no. 100584, 2025, <https://doi.org/10.1016/j.puhip.2025.100584>.