Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8750 – 8760 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

OPTIMIZING THE CLASSIFICATION OF SOCIAL MEDIA'S IMPACT ON MENTAL HEALTH USING RANDOM FOREST WITH GRIDSEARCHCV HYPERPARAMETER TUNING

Oloan Sihombing[1], Jefry So[2], Geby Stevanie[3]
[1,2,3]Universitas Prima Indonesia, Medan
[1,2,3] Pusat Unggulan IPTEK (PUI-ITIK) Inovasi Teknologi Ilmu Komputer
e-mail: [1]oloansihombing@unprimdn.ac.id, [2]sojefry@gmail.com,
[3]gebste1558@gmail.com

Abstract: *The rapid growth of social media has significantly impacted various aspects of life, including individual mental health. This study aims to optimize a classification model for identifying the impact of social media usage on mental health using the Random Forest algorithm. The dataset incorporates various indicators of social media usage behavior and the psychological conditions of respondents. The research process encompasses data preprocessing, model training, and hyperparameter optimization via the GridSearchCV method to enhance model performance. The results demonstrate that hyperparameter tuning improves accuracy, precision, recall, and F1-score compared to the baseline Random Forest model. The resulting optimal model is expected to assist in the early detection of the negative impacts of social media on mental health, thereby serving as a basis for decision-making in the fields of education, healthcare, and public policy.*

Keyword: *Random Forest, GridSearchCV, Classification, Social Media, Mental Healh.*

Abstrak: Perkembangan media sosial yang pesat telah memberikan dampak signifikan terhadap berbagai aspek kehidupan, termasuk kesehatan mental individu. Penelitian ini bertujuan untuk mengoptimalkan model klasifikasi dalam mengidentifikasi dampak penggunaan media sosial terhadap kesehatan mental dengan menggunakan algoritma Random Forest. Dataset yang digunakan mencakup berbagai indikator perilaku penggunaan media sosial serta kondisi psikologis responden. Proses penelitian meliputi tahap praproses data, pelatihan model, serta optimasi hyperparameter menggunakan metode GridSearchCV untuk meningkatkan performa model. Hasil penelitian menunjukkan bahwa penerapan hyperparameter tuning mampu meningkatkan akurasi, presisi, recall, dan f1-score dibandingkan model dasar Random Forest. Model terbaik yang dihasilkan diharapkan dapat membantu dalam deteksi dini dampak negatif media sosial terhadap kesehatan mental, sehingga dapat menjadi dasar pengambilan keputusan dalam bidang pendidikan, kesehatan, maupun kebijakan publik.

Kata kunci: Random Forest, GridSearchCV, Klasifikasi, Media Sosial, Kesehatan Mental.

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong peningkatan penggunaan media sosial secara signifikan di berbagai kalangan masyarakat. Media sosial tidak hanya berfungsi sebagai sarana komunikasi dan hiburan, tetapi juga menjadi ruang utama dalam membentuk interaksi sosial, pola pikir, serta perilaku individu [1]. Namun, di balik manfaat yang ditawarkan, penggunaan media sosial yang tidak terkontrol juga menimbulkan berbagai dampak negatif, khususnya terhadap kesehatan mental. Paparan konten yang berlebihan, tekanan sosial, perbandingan diri (social comparison), hingga cyberbullying merupakan beberapa faktor yang berkontribusi terhadap munculnya gangguan psikologis seperti stres, kecemasan, dan depresi [2].

Kesehatan mental merupakan salah satu aspek fundamental dalam menentukan kualitas hidup individu serta produktivitas masyarakat secara keseluruhan. Individu dengan kondisi mental yang baik cenderung memiliki kemampuan berpikir yang lebih jernih, pengambilan keputusan yang lebih rasional, serta kinerja yang lebih optimal dalam berbagai bidang, baik pendidikan maupun pekerjaan [3]. Sebaliknya, gangguan kesehatan mental dapat menurunkan tingkat produktivitas, meningkatkan risiko kesalahan dalam pengambilan keputusan, serta berdampak pada kualitas sumber daya manusia (SDM) secara jangka Panjang [4]. Oleh karena itu, pemahaman yang komprehensif mengenai faktorfaktor yang memengaruhi kesehatan mental, termasuk peran media sosial, menjadi sangat penting dalam konteks pembangunan masyarakat modern [5].

Di Indonesia, tingginya penetrasi pengguna media sosial menunjukkan bahwa sebagian besar masyarakat terpapar secara langsung terhadap berbagai konten digital setiap harinya [6]. Hal ini menjadikan media sosial sebagai salah satu faktor yang memiliki pengaruh signifikan terhadap kondisi psikologis individu. Meskipun demikian, penelitian yang secara khusus mengkaji dampak media sosial terhadap kesehatan mental dengan pendekatan berbasis data dan teknologi masih relatif terbatas [7]. Sebagian besar studi masih bersifat deskriptif atau kualitatif, sehingga belum mampu memberikan hasil yang objektif dan terukur dalam mengidentifikasi pola hubungan antara penggunaan media sosial dan kondisi kesehatan mental [8]. Seiring dengan berkembangnya bidang data science dan kecerdasan buatan, pendekatan berbasis machine learning menjadi solusi yang potensial dalam menganalisis permasalahan kompleks secara lebih sistematis dan akurat [9]. Salah satu algoritma yang banyak digunakan dalam tugas klasifikasi adalah Random Forest, yang dikenal memiliki kemampuan dalam menangani data dengan kompleksitas tinggi serta mengurangi risiko overfitting [10]. Namun, performa model Random Forest sangat dipengaruhi oleh pemilihan hyperparameter yang tepat. Oleh karena itu, diperlukan teknik optimasi seperti GridSearchCV untuk menemukan kombinasi parameter terbaik yang dapat meningkatkan akurasi dan keandalan model [11].

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada optimasi klasifikasi dampak media sosial terhadap kesehatan mental menggunakan algoritma Random Forest dengan pendekatan hyperparameter tuning GridSearchCV [12]. Penelitian ini tidak hanya bertujuan untuk meningkatkan performa model klasifikasi, tetapi juga untuk menghasilkan pemahaman yang lebih mendalam mengenai pola hubungan antara penggunaan media sosial dan kondisi kesehatan mental individu secara data-driven. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang kecerdasan buatan, serta menjadi dasar dalam pengambilan kebijakan, penyusunan program edukasi digital, dan upaya peningkatan kualitas kesehatan mental masyarakat guna mendukung terciptanya sumber daya manusia yang produktif dan berkualitas.

METODE

Penelitian ini merupakan penelitian kuantitatif yang bertujuan untuk menganalisis dan mengklasifikasikan dampak penggunaan media sosial terhadap kesehatan mental berdasarkan data numerik yang terstruktur [13]. Pendekatan kuantitatif dipilih karena mampu memberikan hasil analisis yang objektif, terukur, dan dapat diuji secara statistik melalui penerapan metode machine learning. Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle, yang menyediakan dataset terkait penggunaan media sosial dan indikator kesehatan mental. Selanjutnya, data tersebut melalui tahapan praproses, pengolahan, serta analisis menggunakan

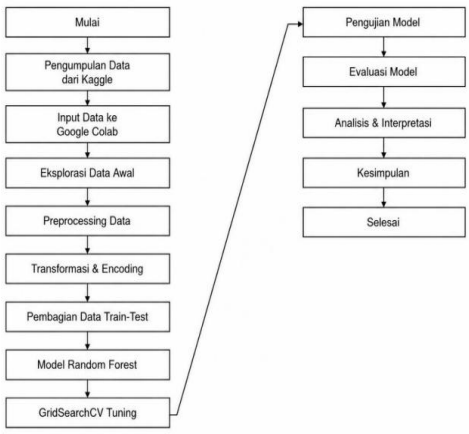
Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8750 – 8760 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

algoritma Random Forest dengan optimasi hyperparameter tuning GridSearchCV untuk menghasilkan model klasifikasi yang optimal dan akurat [14].

Data yang digunakan dalam penelitian ini adalah data individu pengguna media sosial yang memiliki informasi terkait pola penggunaan media sosial dan kondisi kesehatan mental yang diperoleh dari dataset publik. Data ini dipilih karena merepresentasikan hubungan antara aktivitas media sosial dan kondisi psikologis secara terukur.

Teknik pengambilan sampel dalam penelitian ini menggunakan Purposive Sampling, yaitu dengan memilih data yang memenuhi kriteria tertentu agar sesuai dengan kebutuhan analisis dan tujuan penelitian[[15]. Pendekatan ini digunakan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik serta relevan dalam proses klasifikasi menggunakan algoritma Random Forest.

Agar penelitian dapat berjalan dengan baik dan selesai tepat waktu maka terdapat diagram penelitian. Adapaun diagram penelitian pada penelitian ini adalah sebagai berikut:



Gambar 2 Flowchart

1. Pengumpulan dan Input Data

Data yang diperoleh dari platform Kaggle dimasukkan ke dalam Google Colab dalam
format dataset terstruktur seperti file CSV untuk memudahkan proses pengolahan.

2. Eksplorasi Data Awal

Dilakukan pemeriksaan awal untuk mengetahui struktur data, jumlah data, tipe variabel,
serta mengidentifikasi adanya missing value, duplikasi, atau outlier guna memastikan
kualitas data.

3. Preprocessing Data

Data dibersihkan melalui proses data cleaning, meliputi penanganan missing value
(imputasi atau penghapusan), penghapusan data duplikat, penanganan outlier, serta
transformasi data agar sesuai dengan kebutuhan analisis.

4. Transformasi dan Encoding Data

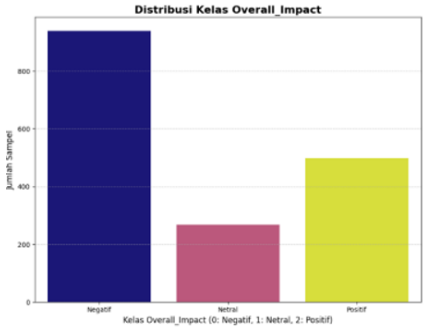
Variabel kategorikal diubah ke dalam bentuk numerik menggunakan teknik encoding
(seperti label encoding atau one-hot encoding), serta dilakukan normalisasi atau standarisasi jika diperlukan.

5. Pembagian Data Latih dan Data Uji Dataset
   dibagi menjadi data latih (training set) dan data uji (testing set) dengan proporsi tertentu untuk melatih dan menguji model.
6. Pembangunan Model Random Forest Model klasifikasi dibangun menggunakan algoritma Random Forest untuk mengidentifikasi pola hubungan antara penggunaan media sosial dan kondisi kesehatan mental.
7. Optimasi Hyperparameter (GridSearchCV) Dilakukan proses tuning menggunakan GridSearchCV untuk menemukan kombinasi parameter terbaik (seperti jumlah pohon, kedalaman maksimum, dan parameter lainnya) guna meningkatkan performa model.
8. Pengujian Model Model yang telah dilatih diuji menggunakan data uji untuk mengetahui kemampuan dalam melakukan klasifikasi terhadap data baru.
9. Evaluasi Model Kinerja model dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan

F1- score untuk mengukur tingkat keberhasilan klasifikasi.

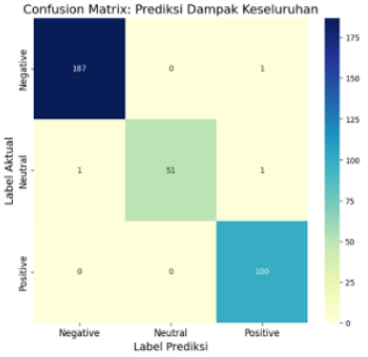
10. Analsisi dan Interpretasi Hasil Hasil klasifikasi dianalisis untuk memahami pola hubungan antara penggunaan media sosial dan dampaknya terhadap kesehatan mental, serta digunakan sebagai dasar penarikan kesimpulan penelitian.

HASIL DAN PEMBAHASAN



Gambar 1 Distribusi Kelas Overall Impact

Distribusi kelas pada variabel Overall_Impact ditunjukkan pada Gambar 3.1. Dataset yang digunakan dalam penelitian ini terdiri atas tiga kategori kelas, yaitu kelas negatif sebanyak 939 data, kelas netral sebanyak 267 data, dan kelas positif sebanyak 499 data. Berdasarkan visualisasi distribusi kelas tersebut, terlihat bahwa kelas negatif memiliki jumlah sampel paling dominan dibandingkan kelas lainnya, sedangkan kelas netral memiliki jumlah data paling sedikit. Perbedaan jumlah sampel antar kelas menunjukkan bahwa dataset mengalami ketidakseimbangan kelas (class imbalance), khususnya antara kelas negatif dan kelas netral. Kondisi ketidakseimbangan kelas dapat mempengaruhi kinerja model klasifikasi karena algoritma cenderung lebih optimal dalam mempelajari pola pada kelas mayoritas dibandingkan kelas minoritas. Akibatnya, model berpotensi menghasilkan tingkat prediksi yang lebih tinggi pada kelas negatif, namun memiliki performa yang lebih rendah dalam mengidentifikasi kelas netral maupun positif. Meskipun demikian, algoritma Random Forest memiliki kemampuan yang relatif baik dalam menangani dataset yang tidak seimbang melalui pendekatan ensemble learning berbasis pohon keputusan. Selain itu, penerapan optimasi hyperparameter menggunakan GridSearchCV diharapkan mampu meningkatkan performa model sehingga proses klasifikasi dapat dilakukan secara lebih akurat dan stabil pada seluruh kategori kelas.



Gambar 2 Heatmap Korelasi Antar Fitur

Gambar 3.2 menunjukkan heatmap korelasi antar fitur yang digunakan dalam penelitian klasifikasi dampak media sosial terhadap kesehatan mental. Visualisasi ini bertujuan untuk mengidentifikasi tingkat hubungan antar variabel, baik hubungan positif maupun negatif, sehingga dapat memberikan gambaran awal mengenai pola data sebelum dilakukan proses pemodelan. Berdasarkan hasil analisis korelasi, beberapa variabel menunjukkan hubungan yang cukup kuat terhadap variabel Overall_Impact. Variabel Mental_Health_Score memiliki korelasi positif tertinggi terhadap Overall_Impact dengan nilai korelasi sebesar 0,85. Nilai tersebut menunjukkan bahwa skor kesehatan mental memiliki keterkaitan yang kuat dalam menentukan kategori dampak media sosial terhadap kondisi mental pengguna.

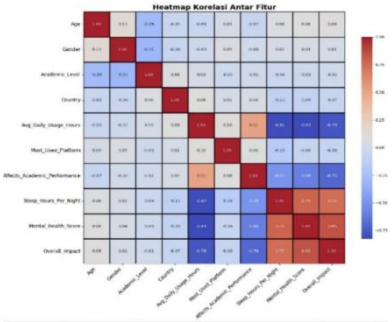
Selain itu, variabel Sleep_Hours_Per_Night juga menunjukkan korelasi positif yang cukup tinggi terhadap Overall_Impact dengan

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8750 – 8760 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

nilai sebesar 0,75. Hasil ini mengindikasikan bahwa durasi tidur memiliki hubungan dengan kondisi kesehatan mental yang dipengaruhi oleh penggunaan media sosial. Sebaliknya, variabel Avg_Daily_Usage_Hours memiliki korelasi negatif sebesar -0,70 terhadap Overall_Impact. Nilai tersebut menunjukkan bahwa peningkatan intensitas penggunaan media sosial per hari cenderung berkaitan dengan dampak kesehatan mental yang lebih negatif.

Sementara itu, beberapa variabel lain seperti Gender, Country, dan Most_Used_Platform menunjukkan nilai korelasi yang relatif rendah terhadap variabel target. Hal ini mengindikasikan bahwa kontribusi variabel-variabel tersebut terhadap proses klasifikasi cenderung lebih kecil dibandingkan fitur lainnya. Secara keseluruhan, hasil analisis korelasi menunjukkan bahwa variabel yang berkaitan dengan kondisi psikologis dan pola penggunaan media sosial memiliki hubungan yang lebih kuat terhadap klasifikasi dampak kesehatan mental. Informasi ini menjadi dasar penting dalam proses pelatihan model Random Forest, karena fitur dengan tingkat korelasi yang lebih tinggi berpotensi memberikan kontribusi yang lebih besar terhadap performa prediksi model.

Gambar 3 Confusion Matrix



Gambar 3.3 menunjukkan hasil confusion matrix dari model Random Forest dalam melakukan klasifikasi dampak media sosial terhadap kesehatan mental ke dalam tiga kategori kelas, yaitu Negative, Neutral, dan Positive. Visualisasi ini digunakan untuk mengevaluasi performa model dengan membandingkan label aktual dan label hasil prediksi. Berdasarkan hasil pengujian, model mampu mengklasifikasikan kelas Negative dengan tingkat akurasi yang sangat tinggi. Sebanyak 187 data berhasil diprediksi dengan benar sebagai kelas Negative, sedangkan hanya terdapat 1 data yang salah diklasifikasikan ke kelas Positive. Hasil tersebut menunjukkan bahwa model mampu mengenali pola pada kelas negatif dengan baik.

Pada kelas Neutral, model berhasil memprediksi 51 data secara tepat, sementara masingmasing terdapat 1 data yang salah diklasifikasikan ke kelas Negative dan Positive. Meskipun jumlah data pada kelas Neutral relatif lebih sedikit dibandingkan kelas lainnya, model tetap menunjukkan kemampuan klasifikasi yang baik pada kelas tersebut. Sementara itu, seluruh data pada kelas Positive sebanyak 100 sampel berhasil diprediksi dengan benar tanpa kesalahan klasifikasi. Hasil ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali karakteristik data pada kelas positif.

Secara keseluruhan, jumlah kesalahan klasifikasi pada confusion matrix relatif kecil dibandingkan jumlah prediksi yang benar pada seluruh kelas. Hal ini menunjukkan bahwa algoritma Random Forest yang dioptimasi menggunakan GridSearchCV mampu menghasilkan model klasifikasi dengan tingkat akurasi yang tinggi dan performa yang stabil dalam mengidentifikasi dampak media sosial terhadap kesehatan mental. Selain itu, hasil confusion matrix mengindikasikan bahwa proses optimasi hyperparameter berkontribusi terhadap peningkatan kemampuan generalisasi model pada data pengujian. Rendahnya tingkat kesalahan klasifikasi pada setiap kategori menunjukkan bahwa model mampu membedakan masing-masing kelas secara efektif.

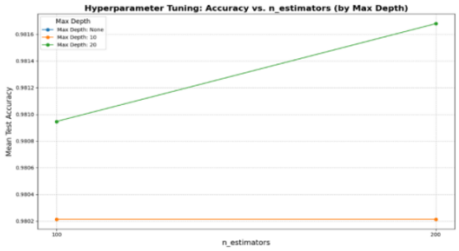
Tabel 1 Classification Report Model Random Forest

Algoritma	Label	Precision	Recall	F1-score	Support	average accuracy	Amount of data
Random Forest	0	0.99	0.99	0.99	188	0.99	341
	1	1.00	0.96	0.98	53		
	2	0.98	1.00	0.99	100		

Tabel 1 menunjukkan hasil evaluasi performa model Random Forest dalam melakukan klasifikasi dampak media sosial terhadap kesehatan mental menggunakan metrik precision, recall, F1-score, dan accuracy. Evaluasi dilakukan pada tiga kategori kelas, yaitu label 0 (Negative), label 1 (Neutral), dan label 2 (Positive). Berdasarkan hasil pengujian, model Random Forest memperoleh nilai accuracy sebesar 0,99 atau 99% dari total 341 data pengujian. Nilai tersebut menunjukkan bahwa model memiliki tingkat akurasi yang sangat tinggi dalam mengklasifikasikan dampak media sosial terhadap kesehatan mental.

Pada label 0 (Negative), model memperoleh nilai precision, recall, dan F1-score masingmasing sebesar 0,99 dengan jumlah support sebanyak 188 data. Hasil tersebut menunjukkan bahwa model mampu mengenali dan memprediksi kelas negatif dengan tingkat kesalahan klasifikasi yang sangat rendah. Untuk label 1 (Neutral), model menghasilkan nilai precision sebesar 1,00, recall sebesar 0,96, dan F1-score sebesar 0,98 dengan jumlah support sebanyak 53 data. Nilai precision yang tinggi menunjukkan bahwa prediksi pada kelas netral memiliki tingkat ketepatan yang sangat baik. Namun, nilai recall yang sedikit lebih rendah mengindikasikan bahwa masih terdapat sejumlah kecil data netral yang diklasifikasikan ke kelas lain.

Sementara itu, pada label 2 (Positive), model memperoleh nilai precision sebesar 0,98, recall sebesar 1,00, dan F1-score sebesar 0,99 dengan jumlah support sebanyak 100 data. Nilai recall sebesar 1,00 menunjukkan bahwa seluruh data positif berhasil dikenali oleh model tanpa adanya data yang tidak terdeteksi. Secara keseluruhan, hasil classification report menunjukkan bahwa algoritma Random Forest yang dioptimasi menggunakan GridSearchCV mampu menghasilkan performa klasifikasi yang sangat baik pada seluruh kategori kelas. Tingginya nilai precision, recall, dan F1-score mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik dalam membedakan kategori dampak media sosial terhadap kesehatan mental secara akurat dan konsisten.



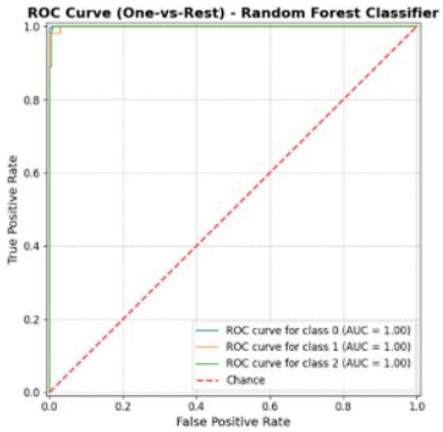
Gambar 4 Hyperparameter Tuning: Accuracy vs n_estimators

Gambar 3.4 menunjukkan hasil proses hyperparameter tuning pada algoritma Random Forest menggunakan metode GridSearchCV dengan fokus pada pengaruh parameter n_estimators terhadap nilai akurasi model. Pengujian dilakukan pada beberapa kombinasi nilai max_depth, yaitu tanpa batas (None), kedalaman 10, dan kedalaman 20. Berdasarkan visualisasi tersebut, peningkatan jumlah n_estimators dari 100 menjadi 200 memberikan pengaruh terhadap performa model, khususnya pada kombinasi parameter max_depth = 20. Pada kombinasi tersebut, nilai mean test accuracy mengalami peningkatan dan menghasilkan performa tertinggi dibandingkan kombinasi parameter
lainnya.

Sementara itu, parameter max_depth = 10 menunjukkan nilai akurasi yang relatif stabil meskipun jumlah n_estimators ditingkatkan. Hal ini mengindikasikan bahwa pembatasan kedalaman pohon yang terlalu rendah dapat membatasi kemampuan model

dalam mempelajari pola data yang lebih kompleks. Pada kondisi max_depth tanpa batas (None), nilai akurasi berada di bawah kombinasi max_depth = 20. Kondisi tersebut menunjukkan bahwa pohon keputusan yang terlalu dalam berpotensi menyebabkan overfitting, sehingga performa model pada data pengujian tidak meningkat secara optimal.

Hasil hyperparameter tuning menunjukkan bahwa kombinasi parameter tertentu mampu meningkatkan performa algoritma Random Forest dalam melakukan klasifikasi dampak media sosial terhadap kesehatan mental. Penggunaan jumlah estimator yang lebih besar memungkinkan model membangun lebih banyak pohon keputusan, sehingga hasil prediksi menjadi lebih stabil. Secara keseluruhan, proses optimasi hyperparameter menggunakan GridSearchCV terbukti efektif dalam membantu pemilihan kombinasi parameter terbaik untuk meningkatkan performa klasifikasi. Oleh karena itu, proses tuning parameter menjadi tahapan penting dalam menghasilkan model Random Forest yang optimal dan memiliki kemampuan generalisasi yang baik terhadap data pengujian.



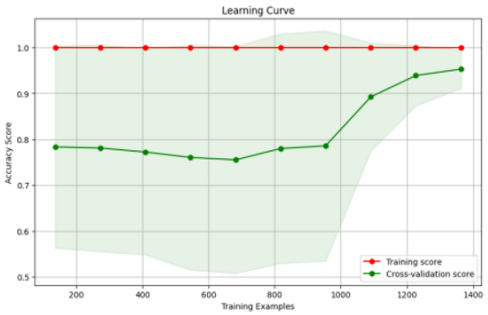
Gambar 5 ROC Curve

Gambar 3.5 menunjukkan hasil evaluasi performa model Random Forest menggunakan kurva ROC (Receiver Operating Characteristic Curve) pada proses klasifikasi dampak media sosial terhadap kesehatan mental. Kurva ROC digunakan untuk mengukur kemampuan model dalam membedakan setiap kelas berdasarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR). Pada visualisasi tersebut, garis diagonal putus-putus merepresentasikan performa klasifikasi acak (random classifier), sedangkan kurva pada masing-masing kelas menunjukkan performa model hasil pelatihan. Berdasarkan hasil pengujian, seluruh kelas memperoleh nilai Area Under Curve (AUC) sebesar 1,00. Nilai tersebut menunjukkan bahwa model Random Forest memiliki kemampuan yang sangat baik dalam membedakan setiap kategori dampak kesehatan mental. Kurva ROC yang berada mendekati sudut kiri atas mengindikasikan bahwa model memiliki tingkat sensitivitas yang tinggi serta tingkat kesalahan klasifikasi yang rendah.

Hasil ini menunjukkan bahwa model mampu memisahkan masing-masing kelas dengan sangat baik pada data pengujian. Tingginya nilai AUC juga mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik sehingga dapat menghasilkan prediksi yang konsisten pada setiap kategori kelas. Selain itu, performa ROC Curve yang tinggi menunjukkan bahwa proses optimasi hyperparameter menggunakan GridSearchCV memberikan kontribusi terhadap peningkatan performa model Random Forest. Kombinasi parameter terbaik yang diperoleh selama proses tuning membantu model dalam mempelajari pola data secara lebih efektif sehingga menghasilkan performa klasifikasi yang optimal.

Secara keseluruhan, hasil evaluasi ROC Curve menunjukkan bahwa algoritma Random Forest yang dioptimasi menggunakan GridSearchCV efektif digunakan dalam klasifikasi dampak media sosial terhadap kesehatan mental. Nilai AUC sebesar 1,00 pada seluruh kelas mengindikasikan bahwa model memiliki kemampuan klasifikasi yang sangat tinggi pada data pengujian.

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
Oct 2026, IX (5): 8750 – 8760 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

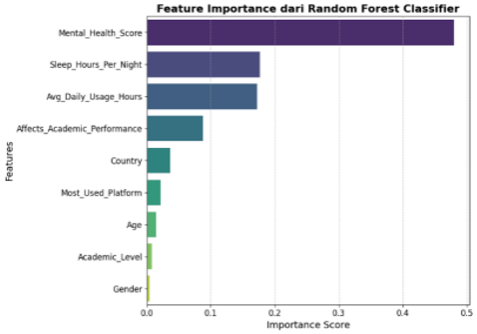


Gambar 5 Learning Curve

Gambar 3.6 menunjukkan learning curve dari model Random Forest pada proses pelatihan dan validasi dalam klasifikasi dampak media sosial terhadap kesehatan mental. Visualisasi ini digunakan untuk mengevaluasi hubungan antara jumlah data pelatihan (training examples) dan performa model berdasarkan nilai akurasi pada data pelatihan serta data validasi (crossvalidation). Pada grafik tersebut, garis merah merepresentasikan training score, sedangkan garis hijau menunjukkan cross-validation score. Berdasarkan hasil visualisasi, nilai training score cenderung stabil dan berada mendekati angka 1,00 pada hampir seluruh jumlah data pelatihan. Kondisi ini menunjukkan bahwa model Random Forest mampu mempelajari pola pada data pelatihan dengan baik.

Sementara itu, nilai cross-validation score mengalami peningkatan secara bertahap seiring bertambahnya jumlah data pelatihan. Peningkatan tersebut mengindikasikan bahwa model memperoleh kemampuan generalisasi yang lebih baik ketika dilatih menggunakan jumlah data yang lebih besar. Pada jumlah data pelatihan yang lebih sedikit, terlihat adanya selisih antara nilai akurasi pelatihan dan validasi yang menunjukkan kecenderungan overfitting, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data pelatihan. Namun, seiring meningkatnya jumlah data pelatihan, jarak antara kedua kurva semakin kecil dan nilai validasi menjadi lebih stabil. Hal ini menunjukkan bahwa model mulai mampu melakukan prediksi secara lebih konsisten terhadap data baru yang belum pernah digunakan pada proses pelatihan. Selain itu, area bayangan pada kurva validasi menggambarkan tingkat variasi performa model selama proses pelatihan. Variasi yang semakin kecil pada jumlah data yang lebih besar mengindikasikan bahwa model memiliki tingkat stabilitas yang lebih baik.

Secara keseluruhan, hasil learning curve menunjukkan bahwa algoritma Random Forest yang dioptimasi menggunakan GridSearchCV mampu menghasilkan model klasifikasi dengan performa yang stabil dan kemampuan generalisasi yang baik. Dengan demikian, proses optimasi hyperparameter berkontribusi dalam mengurangi kecenderungan overfitting sekaligus meningkatkan kualitas prediksi pada klasifikasi dampak media sosial terhadap kesehatan mental.



Gambar 7 Feature Importance

Gambar 3.7 menunjukkan hasil feature importance dari algoritma Random Forest pada proses klasifikasi dampak media sosial terhadap kesehatan mental. Visualisasi ini digunakan untuk mengidentifikasi tingkat kontribusi masing-masing variabel terhadap proses prediksi model. Semakin tinggi nilai importance score suatu fitur, maka semakin besar kontribusinya dalam menentukan hasil klasifikasi. Berdasarkan hasil visualisasi, variabel Mental_Health_Score memiliki nilai feature importance tertinggi dibandingkan variabel lainnya dengan nilai mendekati 0,50. Hasil tersebut menunjukkan bahwa skor kesehatan mental merupakan fitur yang paling dominan dalam proses klasifikasi dampak media sosial terhadap kondisi psikologis pengguna. Tingginya

kontribusi variabel ini mengindikasikan bahwa kondisi kesehatan mental memiliki keterkaitan yang kuat terhadap penentuan kategori klasifikasi pada model. Selain itu, variable Sleep_Hours_Per_Night dan Avg_Daily_Usage_Hours juga memberikan kontribusi yang cukup besar terhadap performa model. Hasil ini menunjukkan bahwa durasi tidur dan intensitas penggunaan media sosial memiliki hubungan yang erat dengan kondisi kesehatan mental individu. Sementara itu, beberapa variabel seperti Gender, Academic_Level, dan Age memiliki nilai kontribusi yang relatif lebih rendah dibandingkan fitur lainnya. Kondisi tersebut mengindikasikan bahwa variabel demografis memberikan kontribusi yang lebih kecil dalam proses klasifikasi dibandingkan variabel yang berkaitan langsung dengan perilaku penggunaan media sosial dan kondisi psikologis pengguna. Hasil feature importance juga menunjukkan bahwa algoritma Random Forest mampu mengidentifikasi fitur-fitur yang paling relevan dalam proses klasifikasi. Kemampuan ini membantu model dalam memfokuskan proses pembelajaran pada variabel yang memiliki kontribusi lebih besar sehingga dapat meningkatkan performa prediksi secara keseluruhan. Secara keseluruhan, analisis feature importance menunjukkan bahwa faktor kesehatan mental, kualitas tidur, dan intensitas penggunaan media sosial merupakan variabel utama yang 16 berkontribusi dalam klasifikasi dampak media sosial terhadap kesehatan mental. Dengan demikian, penggunaan algoritma Random Forest yang dioptimasi menggunakan GridSearchCV mampu menghasilkan model klasifikasi dengan performa yang baik serta tingkat interpretabilitas yang memadai.

SIMPULAN

Berdasarkan hasil penelitian, algoritma Random Forest yang dioptimasi menggunakan metode GridSearchCV mampu menghasilkan performa klasifikasi yang sangat baik dalam menganalisis dampak media sosial terhadap kesehatan mental. Optimasi hyperparameter membantu meningkatkan akurasi dan stabilitas model dalam mempelajari pola data. Hasil evaluasi menunjukkan nilai accuracy sebesar 99% dengan precision, recall, dan F1-score yang tinggi pada seluruh kelas. Selain itu, confusion matrix menunjukkan tingkat kesalahan klasifikasi yang rendah sehingga model mampu membedakan kategori dampak kesehatan mental secara efektif. Hasil evaluasi ROC Curve menunjukkan seluruh kelas memperoleh nilai AUC sebesar 1,00 yang menandakan kemampuan klasifikasi yang sangat baik. Learning curve juga menunjukkan bahwa model memiliki kemampuan generalisasi yang stabil setelah dilakukan optimasi parameter. Analisis feature importance menunjukkan bahwa variabel Mental_Health_Score, Sleep_Hours_Per_Night, dan Avg_Daily_Usage_Hours menjadi faktor yang paling berpengaruh dalam proses klasifikasi. Hal ini menunjukkan adanya hubungan yang kuat antara kesehatan mental, kualitas tidur, dan intensitas penggunaan media sosial terhadap dampak kesehatan mental pengguna. Secara keseluruhan, penelitian ini membuktikan bahwa algoritma Random Forest dengan optimasi GridSearchCV efektif digunakan untuk klasifikasi dampak media sosial terhadap kesehatan mental serta berpotensi diterapkan dalam pengembangan sistem prediksi kesehatan mental berbasis machine learning.

UCAPAN TERIMA KASIH

Terima kasih saya ucapkan kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini, khususnya kepada Bapak Oloan Sihombing, S.Kom., M.Kom. atas bimbingan dan masukannya yang sangat berarti. Terima kasih juga saya sampaikan kepada orang tua, keluarga, dan teman-teman atas segala dukungan, doa, dan motivasi yang diberikan sehingga penelitian ini dapat rampung dengan baik.

DAFTAR PUSTAKA

D. Wiryany, S. Natasha, R. Kurniawan, J. I. Komunikasi, and M. Bandung, "PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI TERHADAP PERUBAHAN SISTEM KOMUNIKASI INDONESIA," 2022.

N. Zuanda and R. Dini, "TREN PENELITIAN CYBERBULLYING DI INDONESIA," 2024.

R. Al Yasin, R. Roro Kirani Annisa Anjani, S. Salsabil, T. Rahmayanti, and R. Amalia, "PENGARUH SOSIAL MEDIA TERHADAP KESEHATAN MENTAL DAN FISIK REMAJA: A SYSTEMATIC REVIEW," vol. 3, no. 2, 2022.

S. Azfa and N. ' Mah, "Prosiding Seminar Nasional Bahasa."

A. Al Rivaldi, J. Soedarto, K. Tembalang, K. Semarang, and J. Tengah, "Analisis Faktor Penyebab Stres pada Mahasiswa dan Dampaknya terhadap Kesehatan Mental," Jurnal Inovasi Riset Ilmu Kesehatan, no. 4, pp. 11–18, 2024, doi: 10.55606/detector.v2i3.4378.

P. Silitonga, "PENGARUH POSITIF DAN NEGATIF MEDIA SOSIAL TERHADAP PERKEMBANGAN SOSIAL, PSIKOLOGIS, DAN PERILAKU REMAJA YANG TIDAK TERBIASA DENGAN TEKNOLOGI SOSIAL MEDIA DI INDONESIA," 2023. [Online]. Available: https://publisherqu.com/index.php/pediaqu

S. Hasna, "TINDAKAN KOLEKTIF MASYARAKAT JARINGAN DI INDONESIA: AKTIVISME SOSIAL MEDIA PADA AKSI #GEJAYANMEMANGGIL," 2022.

T. Febriyanti and R. Mesra, "Trias Febriyanti, Romi Mesra (Perubahan Kehidupan Sosial Sebagai …) Perubahan Kehidupan Sosial Sebagai Dampak Dari Penggunaan Media Sosial di Masyarakat Indonesia," 2024.

V. Marcella Angela Simalango and W. Franciska Senjaya, "Evaluasi Algoritma Decision Tree dan Random Forest serta Efektivitas Feature Selection dalam Memprediksi Kesehatan Mental," 2025.

E. Rosta Br Sebayang, Y. Herry Chrisnanto, U. Jenderal Achmad Yani Cimahi, J. Terusan Jend Sudirman, C. Selatan, and J. Barat, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," 2023. [Online]. Available: http://ijespgjournal.org

Adil Setiawan, Andri Armaginda Siregar, N. Setiawan, Jalaluddin Nasution, Naufal Dhiya Putra Dalimunthe, and Farhan Sardy Abdillah, "Optimasi Performa Model SVM dan Random Forest untuk Klasifikasi Kanker Payudara Menggunakan Penyetelan Hyperparameter," Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI), vol. 4, no. 3, pp. 2141–2149, Jan. 2026,

E. Sutanty, E. Setiyaningsih, and S. N. Yanti, "Optimasi Hyperparameter Pada Algoritma K-Nearest Neighbor untuk Analisis Sentimen terhadap Pembelajaran Jarak Jauh," Decode: Jurnal Pendidikan Teknologi Informasi, vol. 5, no. 2, pp. 348–364, Jul. 2025, doi: 10.51454/decode.v5i2.1170.

M. Irfan Syahroni, D. STIT Al-Aziziyah, J. I. TGH Umar Abdul Aziz kapek Gunung Sari Lombok Barat, and kode pos, "PROSEDUR PENELITIAN KUANTITATIF," Jurnal AlMusthafa STIT Al-Aziziyah Lombok Barat, vol. 43, no. 3, 2022.

C. Azzaria, E. Daniati, A. Ristyawan, U. Nusantara, and P. Kediri, "Peningkatan Akurasi Deteksi Liver Disease melalui Hyperparameter Tuning pada Algoritma Random Forest," 2025. [Online]. Available: https://subset.id/index.php/IJCSR

A. Fajriansi, S. Tinggi Ilmu Kesehatan Nani Hasanuddin, J. Perintis

Kemerdekaan VIII, and K. Makassar, "Hubungan Antara Pengetahuan Dan Sikap Dengan Kontrol Diet Rendah Garam Pada Penderita Hipertensi."

I. Ameer, M. Arif, G. Sidorov, H. Gòmez-Adorno, and A. Gelbukh, "Mental Illness Classification on Social Media Texts using Deep Learning and Transfer Learning," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.01012

A. Priyono, M. Shodiq, D. P. Alvinsyah, and S. A. Hidayah, "METODE RANDOM FOREST UNTUK MEMUDAHKAN KLASIFIKASI DIAGNOSIS PENYAKIT MENTAL," Jurnal Informatika Medis (J-INFORMED), vol. 2, no. 1, pp. 1–4, Jun. 2024, doi: ` 10.52060/im.v2i1.2119.

N. Rohmah, N. Aminudin, A. Wantoro, D. Yana, and A. Andini, "Analisis Sentimen Kesehatan Mental di TikTok pada Generasi Milenial, Gen Z, dan Alpha Menggunakan SVM dan Random Forest," Jurnal Rekayasa Perangkat Lunak, vol. 04, no. 2, 2025, [Online]. Available: https://journal.aisyahuniversity.ac.id/index.php/J-Rapa

S. Octarina, F. M. Puspita, E. Yuliza, and I. Indrawati, "PENDAMPINGAN PENGGUNAAN GOOGLE COLAB PADA PEMBELAJARAN PYTHON DAN MACHINE LEARNING BAGI DOSEN MATEMATIKA DI PALEMBANG," Jurnal Pepadu, vol. 6, no. 1, pp. 56–66, Mar. 2025, doi: 10.29303/pepadu.v6i1.6457.

A. Gumelar and S. Moeslim, "Big Data Analysis of Indonesian Health Q&A for Identifying Dominant Disease Topics and Mitigation Information Patterns Using the Random Forest Algorithm Powered by Apache Spark on Google Colab", doi: 10.3390/xxxxx.