

Penerapan Algoritma Random Forest Untuk Prediksi Tingkat Stres Dari Aktivitas Media Sosial

Muhammad Hussein Umar^{1*}, Rizky Daud Antony Pangaribuan², Raihan Primadana³, Sumanto⁴, Imam Budiawan⁵, Roida Pakpahan⁶

^{1,2,3,4,5,6}Informatika, Universitas Bina Sarana Informatika, Jakarta Pusat, Indonesia

Email: ¹15230291@bsi.ac.id, ²15130377@bsi.ac.id, ³15230298@bsi.ac.id,

⁴sumanto@bsi.ac.id, ⁵imam.imb@bsi.ac.id, ⁶roida.rkh@bsi.ac.id

Email Penulis Korespondensi: ¹15230291@bsi.ac.id

Abstrak– Penelitian ini bertujuan untuk memprediksi tingkat stres pengguna berdasarkan data numerik yang merepresentasikan perilaku dan kebiasaan sehari-hari, seperti kualitas tidur, durasi penggunaan media sosial, frekuensi olahraga, durasi layar harian, dan indeks kebahagiaan. Pemodelan dilakukan menggunakan algoritma Random Forest melalui workflow pada Orange Data Mining yang mencakup pemuatan dataset, pemilihan atribut, penyesuaian tipe data, pelatihan model, serta evaluasi performa. Pengujian menggunakan Test and Score menunjukkan bahwa Random Forest mampu memberikan performa klasifikasi yang stabil. Visualisasi Bar Plot memperlihatkan distribusi metrik evaluasi yang konsisten, sementara Confusion Matrix menunjukkan bahwa sebagian besar prediksi berada pada kelas medium. Analisis hasil prediksi juga memperkuat kecenderungan model dalam memetakan data ke kategori stres dominan. Secara keseluruhan, penelitian ini menunjukkan bahwa metode Random Forest efektif digunakan untuk deteksi tingkat stres berbasis data numerik perilaku pengguna, serta berpotensi menjadi dasar pengembangan sistem pemantauan kesehatan mental berbasis data.

Kata Kunci: Random Forest, Stres, Media Sosial, Data Mining,, Machine Learning

Abstract– This study aims to predict users' stress levels based on numerical data representing daily behavioral patterns, including sleep quality, social media usage duration, exercise frequency, daily screen time, and happiness index. The modeling process was conducted using the Random Forest algorithm implemented within the Orange Data Mining workflow, which includes dataset loading, attribute selection, domain editing, model training, and performance evaluation. The Test and Score results indicate that Random Forest provides stable and reliable classification performance. The Bar Plot visualization shows consistent evaluation metrics across folds, while the Confusion Matrix reveals that most predictions fall into the medium stress category. Further analysis of the prediction table confirms the model's tendency to classify data toward the dominant class. Overall, the findings demonstrate that Random Forest is effective for detecting stress levels based on numerical behavioral data and can serve as a foundation for developing data-driven mental health monitoring systems.

Keywords: Random Forest, Stress, Social Media, Data Mining, Machine Learning

1. PENDAHULUAN

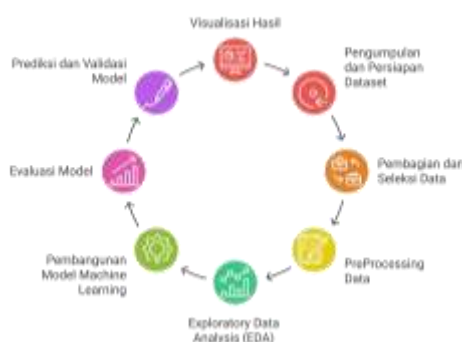
Media sosial saat ini telah menjadi bagian yang tidak terpisahkan dalam kehidupan masyarakat modern. Berbagai aktivitas komunikasi, interaksi sosial, hingga ekspresi emosi kini banyak dilakukan melalui platform seperti Instagram, Twitter, Reddit, dan TikTok. Intensitas penggunaan yang sangat tinggi tersebut memunculkan fenomena yang dikenal sebagai *digital stress*, yaitu tekanan psikologis yang timbul akibat paparan informasi berlebih, perbandingan sosial, komentar negatif, serta dinamika lingkungan digital yang cepat berubah. Kondisi stres ini tidak hanya berdampak pada kesehatan mental, tetapi juga tercermin melalui perubahan gaya bahasa, pilihan kata, nada emosi, dan pola interaksi yang muncul dalam unggahan teks para pengguna. Dengan demikian, media sosial menjadi sumber data penting untuk mengamati dan menganalisis tanda-tanda stres secara digital. Dalam beberapa tahun terakhir, penelitian mengenai deteksi stres berbasis media sosial berkembang pesat seiring meningkatnya kemampuan Natural Language Processing (NLP) dan machine learning dalam memahami pola linguistik yang berkaitan dengan kondisi psikologis. Berbagai studi menunjukkan bahwa metode klasifikasi berbasis machine learning mampu mengenali indikator stres dari teks secara akurat. Nijhawan et al. (2022) menemukan bahwa algoritma seperti Random Forest memberikan performa kompetitif dalam mengklasifikasikan tingkat stres pada teks [1]. Temuan tersebut diperkuat oleh Inamdar et al. (2023), yang menunjukkan bahwa pola bahasa pada unggahan Reddit dapat menjadi prediktor kuat tingkat stres pengguna [2]. Selain itu, dataset Dreddit yang diperkenalkan oleh Turcan dan McKeown (2019) telah menjadi *benchmark* utama yang banyak digunakan dalam penelitian serupa [3]. Perkembangan NLP dan machine learning turut membuka peluang baru dalam analisis emosi dan stres. Rastogi et al. (2022) menekankan bahwa model ensemble mampu meningkatkan kualitas analisis teks media sosial [4], sementara Yang et al. (2022) membuktikan bahwa ekspresi stres kronis pada Twitter dapat ditangkap melalui pendekatan NLP modern [5]. Penelitian Kasmin (2024) dan Illahi et al. (2022) juga menegaskan efektivitas model linguistik dalam mendeteksi stres [6], [7]. Priya (2020) menunjukkan bahwa fitur linguistik dapat digunakan untuk memprediksi depresi, kecemasan, dan stres secara

bersamaan [8]. Selain itu, Zhang et al. (2023) menyoroti peran *interpretable machine learning* dalam mengidentifikasi faktor psikologis dengan lebih transparan [9]. Pendekatan personalisasi pun mulai dikembangkan. Srinivasarao et al. (2025) menggabungkan fitur linguistik dan fisiologis untuk meningkatkan akurasi prediksi stres personal [10]. Inamdar (2023) kembali mengonfirmasi kemampuan machine learning dalam mendeteksi stres kognitif dari teks digital [11]. Febriansyah (2023) memperlihatkan bahwa model machine learning dapat diaplikasikan pada konteks Indonesia dan tetap menghasilkan performa yang stabil [12]. Algoritma Random Forest sendiri telah menjadi salah satu metode yang paling banyak digunakan karena stabilitas, kemampuan generalisasi, serta ketahanannya terhadap *overfitting*. Elisa dan Isnain (2024) menunjukkan bahwa Random Forest dapat mengungguli SVM dan Naive Bayes dalam analisis stigma kesehatan mental [13]. Mahing et al. (2024) juga mengimplementasikan Random Forest dan SVM dalam klasifikasi stres berbasis teks bahasa Indonesia dan mendapatkan hasil yang kuat [14]. Surakusumah dan Apza (2025) menemukan bahwa integrasi SMOTE dan Random Forest meningkatkan performa generalisasi model pada dataset tidak seimbang [15]. Hal tersebut sejalan dengan penelitian Oktaviani et al. (2025), yang menunjukkan bahwa Random Forest efektif untuk mendeteksi stres pada mahasiswa [16]. Selain itu, peran NLP dan machine learning dalam identifikasi emosi turut terlihat dalam penelitian Tolla dan Kusri (2025), yang menggunakan SVM untuk mengidentifikasi stres dan depresi pada unggahan Twitter [17]. Johan dan Azka (2022) juga memanfaatkan multilayer perceptron dan Indonesian Sentiment Lexicon dalam mendeteksi tingkat stres, menunjukkan bahwa fitur linguistik berbahasa Indonesia memiliki pengaruh kuat dalam prediksi [18]. Rachmat et al. (2023) menunjukkan bahwa Random Forest dapat digunakan untuk analisis teks berskala besar melalui studi ulasan aplikasi SATUSEHAT [19]. Fitri et al. (2025) menegaskan kemampuan Random Forest dalam mendeteksi tingkat kecemasan generasi Z melalui data teks media sosial [20]. Dalam konteks penelitian ini, algoritma Random Forest diimplementasikan melalui Orange Data Mining, yang memungkinkan proses pemodelan dilakukan secara visual, sistematis, dan berurutan tanpa perlu menulis kode. Orange menyediakan alur yang intuitif untuk tahapan preprocessing, pemilihan fitur, pelatihan model, serta evaluasi performa melalui node *Test & Score*. Visualisasi hasil menggunakan Bar Plot dan Confusion Matrix membantu memperjelas performa model dalam memprediksi kategori stres, termasuk melihat distribusi prediksi benar maupun salah pada tiap kelas. Berdasarkan kajian literatur dan relevansi teknologi saat ini, pemanfaatan Random Forest dan Orange Data Mining memberikan kerangka kerja yang kuat dan terstruktur dalam mendeteksi tingkat stres berbasis media sosial. Pendekatan ini memastikan hasil analisis yang stabil, mudah dievaluasi, dan sesuai dengan perkembangan terkini dalam penelitian kesehatan mental digital.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Tahapan penelitian digambarkan dalam workflow berikut:



Gambar 1. Alur Penelitian

a. Pengumpulan dan Persiapan Dataset

Tahap awal dilakukan dengan mengumpulkan dataset berisi teks media sosial yang telah diberi label stres dan non-stres. Dataset disimpan dalam format CSV agar kompatibel dengan Orange Data Mining. Pada tahap ini dilakukan pengecekan struktur data, termasuk keberadaan kolom teks dan label, pemeriksaan duplikasi, serta memastikan tidak terdapat nilai kosong. Persiapan ini penting agar dataset siap masuk ke proses pemodelan tanpa gangguan kualitas.

b. Pembagian dan Seleksi Data

Setelah dataset dimuat, proses dilanjutkan dengan menyeleksi bagian data yang relevan menggunakan *Data Sampler* dan *Select Columns*. *Data Sampler* digunakan untuk mengambil sampel tertentu apabila diperlukan untuk pengujian awal, sedangkan *Select Columns* memastikan bahwa atribut teks digunakan sebagai fitur dan kolom label ditetapkan sebagai target. Tahap ini menjamin hanya data penting yang diproses lebih lanjut.

c. Pra-Pemrosesan Data (Preprocessing)

Pada tahap preprocessing, struktur atribut diperbaiki menggunakan *Edit Domain* dengan menentukan tipe data yang sesuai, seperti teks atau kategori. Langkah ini memastikan bahwa data mentah dikonversi ke bentuk terstruktur sehingga siap digunakan oleh algoritma machine learning.

d. Eksplorasi Data (EDA)

Eksplorasi awal dilakukan menggunakan *Data Table* untuk melihat distribusi kelas, karakteristik teks, serta memeriksa keseimbangan data. EDA membantu peneliti memahami pola dasar dataset sehingga dapat menentukan strategi preprocessing dan metode pemodelan yang paling efektif.

e. Pembangunan Model Random Forest

Model utama dibangun menggunakan algoritma Random Forest melalui komponen yang tersedia di Orange. Algoritma ini memanfaatkan kumpulan decision tree untuk meningkatkan stabilitas dan mengurangi risiko overfitting. Beberapa parameter penting yang digunakan meliputi jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan kriteria pemisahan (*criterion*). Model kemudian dilatih menggunakan dataset hasil preprocessing.

f. Evaluasi Model

Evaluasi dilakukan menggunakan node *Test and Score* dengan metode 10-fold cross-validation untuk memperoleh hasil yang stabil. Metrik yang digunakan meliputi akurasi, precision, recall, F1-score, dan AUC. Evaluasi ini memberikan gambaran mengenai efektivitas model dalam mengklasifikasikan tingkat stres secara konsisten.

g. Prediksi dan Validasi

Setelah model dievaluasi, node *Predictions* digunakan untuk menghasilkan label prediksi terhadap data uji. Tahap ini bertujuan memeriksa konsistensi model dalam mengklasifikasikan data baru dan mengidentifikasi potensi kesalahan prediksi untuk analisis lebih lanjut.

h. Visualisasi Hasil

Tahap akhir menghasilkan visualisasi menggunakan *Bar Plot* untuk menampilkan performa model dan *Confusion Matrix* untuk melihat jumlah prediksi benar maupun salah pada tiap kelas. Visualisasi ini membantu memberikan pemahaman komprehensif terhadap performa model dan area yang perlu diperbaiki.

2.2 Desain

Pada bagian ini ditampilkan objek eksperimen yang digunakan, parameter model, serta struktur dataset.

Tabel 1. Struktur dataset

| User_ID | Days_Without_Social_Media | Gender | Sleep_Quality | Daily_Screen_Time (hrs) | Age | Exercise_Frequency (week) | Happiness_Index | Stress_Level |
|---------|---------------------------|--------|---------------|-------------------------|-----|---------------------------|-----------------|--------------|
| U362 | 3 | Female | 6 | 6.6 | 39 | 1 | 7 | medium |
| U074 | 1 | Female | 6 | 4.8 | 26 | 4 | 8 | medium |
| U375 | 1 | Male | 6 | 5.8 | 17 | 5 | 9 | medium |
| U156 | 4 | Female | 7 | 5.2 | 19 | 4 | 10 | medium |

Dataset pada **Tabel 1** memuat informasi mengenai perilaku digital dan kondisi psikologis responden. Setiap baris mewakili satu individu dengan beberapa variabel penting seperti hari tanpa media sosial, kualitas tidur, durasi penggunaan layar harian, usia, frekuensi olahraga, serta indeks kebahagiaan. Variabel-variabel tersebut digunakan untuk menggambarkan kebiasaan hidup dan kondisi emosional responden yang berpotensi memengaruhi tingkat stres. Atribut *Stress_Level* berisi kategori *low*, *medium*, dan *high* yang menjadi target utama dalam proses klasifikasi. Sementara itu, variabel lain berfungsi sebagai fitur untuk membantu model machine learning mengenali pola yang berkaitan dengan stres. Secara keseluruhan, struktur dataset ini memberikan gambaran yang cukup komprehensif mengenai faktor-faktor yang dapat memengaruhi stres, sehingga sangat mendukung proses analisis dan pembangunan model klasifikasi menggunakan Random Forest.

2.3 Tools yang digunakan

a. Microsoft Excel

Microsoft Excel digunakan pada tahap awal penelitian untuk melakukan pengecekan struktur dataset, memastikan tidak terdapat duplikasi, serta memverifikasi tipe data sebelum diproses lebih lanjut. Excel membantu dalam tahap praproses awal agar dataset berada dalam kondisi siap digunakan pada proses machine learning.

b. Microsoft Word

Microsoft Word dimanfaatkan untuk penyusunan naskah penelitian, penyesuaian format artikel sesuai template jurnal, serta dokumentasi hasil analisis. Perangkat ini digunakan pada seluruh tahap penyusunan laporan penelitian hingga proses akhir.

c. Kaggle

Kaggle digunakan sebagai sumber dataset penelitian karena menyediakan kumpulan data publik yang terstruktur, mudah diakses, serta relevan untuk studi berbasis pembelajaran mesin. Platform ini memungkinkan peneliti memilih dataset yang sudah dikurasi dan siap diolah tanpa perlu melakukan pengumpulan data secara manual. Reputasinya sebagai repositori data ilmiah juga diakui dalam penelitian Putra et al. (2023), yang menunjukkan bahwa Kaggle merupakan sumber dataset yang kredibel untuk kepentingan akademik[21].

d. Orange Data Mining

Orange Data Mining menjadi perangkat utama untuk proses analisis karena mendukung berbagai tahapan penelitian, mulai dari preprocessing, ekstraksi fitur, pelatihan model hingga evaluasi performa secara visual dan interaktif. Aplikasi ini menyediakan antarmuka berbasis drag-and-drop sehingga memudahkan peneliti dalam merancang alur pemodelan tanpa perlu menulis kode secara manual. Selain itu, Orange juga memiliki modul text mining dan classification yang sesuai dengan kebutuhan penelitian ini. Efektivitas penggunaan Orange dalam penelitian akademik juga dipaparkan oleh Savić dan Pribičević (2020), yang menunjukkan bahwa Orange merupakan alat pembelajaran mesin yang praktis dan komprehensif[22].

e. Mendeley Reference Manager

Mendeley digunakan sebagai sistem manajemen referensi untuk mempermudah pengelolaan sitasi, pengutipan otomatis, serta penyusunan daftar pustaka sesuai format jurnal. Perangkat ini membantu menjaga konsistensi kutipan pada seluruh bagian artikel dan meminimalkan kesalahan penulisan referensi. Selain berfungsi sebagai manajer sitasi, Mendeley juga menjadi repositori dokumen ilmiah yang memudahkan peneliti menyimpan dan mengorganisasi literatur. Penggunaan Mendeley dalam penulisan akademik juga diperkuat melalui penelitian Cahyono (2019) [23].

2.4 Tahapan Penelitian

a. Import Dataset (File CSV)

Tahapan dimulai dengan mengimpor dataset berformat CSV menggunakan node File di Orange. Pada tahap ini, struktur kolom diperiksa untuk memastikan bahwa data teks dan label telah sesuai dan siap diproses tanpa adanya kesalahan format.

b. Pengambilan Sampel Data (Data Sampler)

Node Data Sampler digunakan bila diperlukan untuk mengambil sebagian data sebagai sampel. Tahapan ini membantu dalam memvalidasi struktur data dan memastikan dataset dapat diproses dengan baik sebelum masuk ke tahap pemodelan.

c. Seleksi Kolom Penting (Select Columns)

Tahap ini berfungsi untuk menentukan atribut yang digunakan sebagai fitur dan menetapkan variabel *Stress_Level* sebagai label. Node ini memastikan hanya kolom yang relevan yang diikutsertakan dalam proses analisis.

d. Penyesuaian Domain Data (Edit Domain)

Node *Edit Domain* digunakan untuk menyesuaikan tipe atribut, baik numerik, kategorikal, maupun teks. Penyesuaian ini diperlukan agar semua variabel kompatibel dengan algoritma Random Forest dan siap digunakan dalam proses klasifikasi.

e. Pemeriksaan Struktur Data (Data Table)

Node *Data Table* dimanfaatkan untuk meninjau isi dataset secara langsung. Langkah ini memastikan tidak terdapat nilai hilang, duplikasi, atau ketidaksesuaian distribusi data yang dapat memengaruhi hasil pemodelan.

f. Pembangunan Model dengan Random Forest

Tahap ini membangun model klasifikasi menggunakan algoritma Random Forest. Model dilatih berdasarkan dataset yang telah diproses agar mampu mendeteksi tingkat stres secara stabil dan akurat.

g. Test and Score

Evaluasi model dilakukan menggunakan node *Test and Score* dengan metode *k-fold cross-validation*. Tahap ini menghasilkan metrik performa seperti akurasi, precision, recall, F1-score, dan AUC untuk menilai kualitas model secara menyeluruh.

h. Predictions

Node *Predictions* digunakan untuk melihat hasil prediksi model terhadap data uji. Tahapan ini menunjukkan bagaimana model mengklasifikasikan setiap instance ke dalam kategori stres tertentu.

i. Visualisasi Hasil

Visualisasi dilakukan menggunakan *Bar Plot* untuk menampilkan performa model dan *Confusion Matrix* untuk melihat distribusi prediksi benar maupun salah pada tiap kelas. Visualisasi ini membantu memahami pola kekuatan dan kelemahan model.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Preprocessing dan Eksplorasi Awal

Dataset yang digunakan dalam penelitian ini merupakan dataset numerik yang terdiri dari beberapa variabel terkait perilaku dan kondisi pengguna, seperti *days without social media*, *sleep quality*, *daily screen time*, *age*, *exercise frequency*, *happiness index*, serta

variabel target *stress level*. Seluruh data dimuat melalui node **File** dan diperiksa secara umum untuk memahami karakteristik awal dataset.

Pada tahap eksplorasi awal, setiap variabel diperhatikan dari sisi jenis datanya, apakah termasuk numerik atau kategorikal. Misalnya, variabel seperti *age* dan *daily screen time* merupakan data numerik kontinu, sedangkan *gender* dan *exercise frequency* merupakan data kategorikal. Pemahaman ini penting untuk memastikan bahwa proses pemodelan selanjutnya sesuai dengan tipe variabel yang digunakan. Selain itu, distribusi nilai pada variabel *stress level* diperiksa untuk melihat apakah dataset memiliki keseimbangan kelas. Variabel ini terbagi dalam beberapa kategori tingkat stres, sehingga dapat digunakan sebagai target untuk proses klasifikasi. Berdasarkan hasil eksplorasi, tidak ditemukan adanya anomali data yang signifikan, sehingga dataset dianggap layak digunakan pada tahap analisis berikutnya.

3.2 Pemeriksaan Struktur Dataset dan Penyiapan Variabel

Tahap ini berfokus pada pemeriksaan teknis dataset menggunakan node **Data Table** untuk memastikan bahwa seluruh variabel telah terbaca dengan benar oleh sistem. Melalui tampilan Data Table, setiap kolom dapat diamati secara langsung sehingga proses verifikasi struktur dataset dapat dilakukan secara menyeluruh. Pemeriksaan ini mencakup pengecekan keberadaan kolom fitur serta kolom target *stress level*, memastikan bahwa keduanya tersedia dan berada pada posisi yang sesuai. Selain itu, data diperiksa untuk memastikan tidak terdapat nilai kosong (*missing values*) yang dapat mengganggu proses analisis. Struktur tipe data juga ditinjau agar setiap variabel memiliki jenis yang konsisten, misalnya variabel numerik tidak terbaca sebagai kategorikal, dan sebaliknya. Hasil pemeriksaan turut memastikan bahwa dataset tidak mengandung duplikasi data yang berpotensi menurunkan akurasi model. Setelah struktur dataset dipastikan benar, proses dilanjutkan menggunakan node **Select Columns** untuk menentukan kolom mana yang digunakan sebagai fitur dan kolom mana yang ditetapkan sebagai target.

Kemudian node **Edit Domain** digunakan untuk menyelaraskan tipe variabel, seperti menetapkan data numerik atau kategorikal sesuai dengan karakteristik aslinya. Tahap ini memastikan bahwa seluruh variabel siap digunakan dalam proses pelatihan model Random Forest.



| Row ID | Age | Gender | Stress Level | Daily Screen Time | Exercise Frequency | Sleep Quality |
|--------|-----|--------|--------------|-------------------|--------------------|---------------|
| 1 | 25 | Female | Low | 2.5 | 3 | Good |
| 2 | 30 | Male | Medium | 3.5 | 2 | Fair |
| 3 | 28 | Female | Low | 2.0 | 4 | Good |
| 4 | 35 | Male | High | 4.5 | 1 | Poor |
| 5 | 22 | Female | Low | 1.5 | 5 | Good |
| 6 | 32 | Male | Medium | 3.0 | 3 | Fair |
| 7 | 27 | Female | Low | 2.2 | 4 | Good |
| 8 | 38 | Male | High | 5.0 | 0 | Poor |
| 9 | 24 | Female | Low | 1.8 | 4 | Good |
| 10 | 31 | Male | Medium | 3.2 | 2 | Fair |
| 11 | 26 | Female | Low | 2.1 | 3 | Good |
| 12 | 33 | Male | Medium | 3.8 | 2 | Fair |
| 13 | 29 | Female | Low | 2.3 | 4 | Good |
| 14 | 36 | Male | High | 4.8 | 1 | Poor |
| 15 | 23 | Female | Low | 1.6 | 5 | Good |
| 16 | 34 | Male | Medium | 3.4 | 2 | Fair |
| 17 | 27 | Female | Low | 2.0 | 4 | Good |
| 18 | 37 | Male | High | 4.9 | 0 | Poor |
| 19 | 25 | Female | Low | 1.9 | 4 | Good |
| 20 | 30 | Male | Medium | 3.1 | 2 | Fair |

Gambar 2. Dataset/Data table

3.3 Implementasi Workflow Orange

Workflow penelitian disusun menggunakan rangkaian node pada Orange Data Mining yang menggambarkan urutan proses analisis mulai dari pemuatan dataset, penyiapan variabel, pembangunan model, hingga visualisasi hasil. Alur ini memastikan bahwa setiap tahapan dilakukan secara sistematis dan saling terhubung. Proses dimulai dengan node **CSV File**, yang berfungsi membaca dataset numerik yang berisi variabel-variabel seperti umur, kualitas tidur, durasi penggunaan media sosial, tingkat kebahagiaan, serta label *stress level*. Setelah dataset dimuat, alur dilanjutkan menuju node **Data Sampler** yang digunakan untuk mengambil sampel data tertentu apabila diperlukan, terutama untuk keperluan uji coba atau validasi awal struktur dataset. Tahap berikutnya menggunakan node **Select Columns**, yang berfungsi menentukan kolom mana yang dijadikan fitur (*features*) dan kolom mana yang digunakan sebagai target klasifikasi. Setelah kolom dipilih, node **Edit Domain** digunakan untuk menyesuaikan tipe variabel, memastikan bahwa setiap atribut sudah berada dalam kategori data yang tepat, seperti numerik atau kategorikal. Node ini berperan penting agar dataset dapat diproses dengan benar oleh algoritma Random Forest. Sebelum model dibangun, dataset ditampilkan melalui node **Data Table** untuk melakukan pengecekan visual terhadap isi data, memastikan tidak ada kesalahan pembacaan, nilai kosong, atau inkonsistensi tipe data. Setelah struktur dataset dinyatakan sesuai, alur kemudian bercabang ke dua arah utama. Cabang keatas menuju node **Predictions**, yang nantinya menampilkan label prediksi dari model terhadap data yang telah diproses.

Cabang kebawah menuju node **Random Forest**, yang bertugas membangun model klasifikasi berdasarkan pola hubungan antarvariabel pada dataset. Model yang telah terbentuk kemudian dievaluasi menggunakan node **Test and Score**, yang menerapkan metode *k-fold cross-validation* untuk menghasilkan metrik performa seperti akurasi, precision, recall, dan F1-score. Dari node **Test and Score**, alur kembali bercabang. Cabang ke atas terhubung ke node **Bar Plot**, yang digunakan untuk memvisualisasikan performa model dalam bentuk grafik. Cabang ke bawah terhubung ke node **Confusion Matrix**, yang menampilkan sebaran prediksi benar dan salah untuk masing-masing kategori *stress level*. Kedua visualisasi ini membantu memberikan pemahaman yang lebih komprehensif terhadap kekuatan dan kelemahan model.



Gambar 3. Workflow Orange

3.4 Pelatihan Model Random Forest

Tahap pelatihan model dilakukan menggunakan algoritma **Random Forest**, yang diimplementasikan melalui node *Random Forest* pada Orange Data Mining. Algoritma ini dipilih karena kemampuannya dalam menangani data numerik dengan baik serta menghasilkan prediksi yang stabil melalui proses penggabungan banyak decision tree. Setiap tree dilatih menggunakan subset data yang berbeda sehingga model mampu menangkap variasi pada dataset dan meminimalkan risiko overfitting. Pada tahap ini, dataset yang telah dipersiapkan melalui proses pemilihan kolom dan penyesuaian tipe variabel digunakan sebagai input bagi node Random Forest. Algoritma kemudian mempelajari hubungan antara variabel-variabel prediktor, seperti *sleep quality*, *daily screen time*, *exercise frequency*, dan *happiness index*, terhadap variabel target *stress level*. Proses pelatihan berlangsung secara otomatis menggunakan parameter bawaan dari Orange, yang sudah dirancang agar sesuai untuk model klasifikasi dasar. Selama proses pemodelan, Random Forest membentuk sekelompok decision tree yang masing-masing menghasilkan prediksi terhadap tingkat stres. Hasil dari seluruh pohon tersebut kemudian digabungkan untuk menghasilkan prediksi final. Pendekatan ini memberikan hasil yang lebih akurat dibandingkan model berbasis single decision tree, karena variasi antar pohon membantu mengurangi bias dan meningkatkan generalisasi model.

Hasil pelatihan model ini menjadi dasar untuk tahap evaluasi selanjutnya yang dilakukan menggunakan node *Test and Score*. Melalui evaluasi tersebut, performa model dapat dilihat secara menyeluruh dan akurat sebelum digunakan untuk prediksi akhir.

3.5 Evaluasi Model

Evaluasi model dilakukan menggunakan node **Test and Score** untuk mengetahui sejauh mana algoritma Random Forest mampu memprediksi tingkat stres secara akurat. Node ini menerapkan metode *k-fold cross-validation*, yang membagi dataset ke dalam beberapa bagian secara bergantian untuk memastikan bahwa penilaian model bersifat stabil dan tidak bergantung pada satu pembagian data tertentu.

Dengan pendekatan ini, hasil evaluasi mencerminkan performa model secara menyeluruh. Hasil evaluasi yang dihasilkan oleh node Test and Score mencakup metrik utama seperti akurasi, precision, recall, F1-score, dan AUC. Nilai **akurasi** menunjukkan proporsi prediksi benar yang dihasilkan model, sedangkan **precision** menggambarkan ketepatan model dalam memprediksi tingkat stres tertentu. **Recall** memberikan informasi mengenai kemampuan model dalam mengenali seluruh data pada kelas tertentu, dan **F1-score** menjadi indikator keseimbangan antara precision dan recall. Sementara itu, **AUC** menunjukkan kemampuan model dalam membedakan setiap tingkat stres secara menyeluruh. Secara umum, hasil evaluasi menunjukkan bahwa Random Forest memberikan performa yang stabil pada setiap metrik evaluasi. Stabilitas nilai antar-fold menandakan bahwa model mampu mempelajari pola dalam dataset secara konsisten tanpa mengalami overfitting.

Hasil evaluasi ini menjadi landasan untuk menilai kekuatan dan kelemahan model sebelum dilakukan analisis lanjutan melalui visualisasi dan confusion matrix.

| Model | AUC | CA | F1 | Prec | Recall | MCC |
|---------------|-------|-------|-------|-------|--------|-------|
| Random Forest | 0.813 | 0.712 | 0.707 | 0.706 | 0.712 | 0.427 |

Gambar 4. Test And Score

3.6 Visualisasi Distribusi Data Menggunakan Bar Plot

Visualisasi distribusi data dilakukan menggunakan node *Bar Plot* untuk melihat pola hubungan antara variabel *Daily Screen Time (hrs)* dan *Stress Level*.

Bar Plot ini menampilkan rentang nilai penggunaan layar harian berdasarkan tiga kategori stres, yaitu **low**, **medium**, dan **high**, yang masing-masing direpresentasikan oleh warna biru, merah, dan hijau.



Gambar 5. BarPlot

Berdasarkan visualisasi tersebut, terlihat bahwa kategori **low stress** (warna biru) memiliki nilai *daily screen time* yang relatif rendah, sebagian besar berada pada kisaran **1–4 jam per hari**, dengan hanya sedikit nilai yang mendekati 6 jam. Pola ini menunjukkan bahwa pengguna dengan tingkat stres rendah cenderung memiliki waktu penggunaan layar yang lebih sedikit. Pada kategori **medium stress** (warna merah), distribusi waktu penggunaan layar terlihat lebih luas dan bervariasi, mulai dari **sekitar 2 hingga 7 jam per hari**.

Variasi yang lebih lebar ini menunjukkan bahwa kelompok stres sedang memiliki pola penggunaan layar yang beragam. Tingginya frekuensi nilai menengah dapat mengindikasikan bahwa individu pada kategori ini mulai mengalami tekanan digital yang dipengaruhi oleh durasi penggunaan perangkat yang lebih panjang. Sementara itu, kategori **high stress** (warna hijau) terlihat memiliki nilai *daily screen time* yang jauh lebih tinggi dibandingkan dua kategori lainnya. Shading batang menunjukkan penggunaan layar harian yang umumnya berada pada kisaran **6 hingga 10+ jam per hari**, dengan banyak nilai ekstrem di bagian atas grafik. Hal ini menunjukkan bahwa kelompok dengan tingkat stres tinggi memiliki kecenderungan kuat untuk menghabiskan waktu lebih lama dengan perangkat digital dan media sosial.

Secara keseluruhan, visualisasi Bar Plot ini menunjukkan pola yang konsisten dengan literatur sebelumnya, di mana penggunaan layar yang lebih lama berpotensi berkorelasi dengan meningkatnya tingkat stres. Grafik ini juga memberikan dasar kuat dalam mendukung proses klasifikasi pada tahap pemodelan, karena menunjukkan adanya pola yang cukup jelas antara variabel perilaku digital dan tingkat stres.

3.7 Analisis Confusion Matrix

Analisis Confusion Matrix digunakan untuk melihat performa model Random Forest secara lebih mendalam dengan menilai jumlah prediksi benar dan salah pada setiap kelas *stress level*. Berbeda dengan Bar Plot yang menggambarkan pola distribusi fitur terhadap tingkat stres, Confusion Matrix memberikan gambaran langsung mengenai kinerja model dalam membedakan kelas *low*, *medium*, dan *high* selama proses evaluasi menggunakan *k-fold cross-validation*.

| | | Predicted | | | Σ |
|--------|--------|-----------|--------|------|-----|
| | | low | medium | high | |
| Actual | low | 16 | 27 | 0 | 43 |
| | medium | 14 | 249 | 51 | 314 |
| | high | 0 | 52 | 91 | 143 |
| Σ | | 30 | 328 | 142 | 500 |

Gambar 6. Confusion Matrix

Berdasarkan visualisasi pada Confusion Matrix, model Random Forest menunjukkan performa klasifikasi yang cukup baik, terutama pada kategori **medium**, yang merupakan kelas dengan jumlah data terbesar. Pada kelas ini, sebanyak **249 dari 314 data** berhasil diklasifikasikan dengan benar, sementara sisanya salah prediksi ke kelas *low* (14 data) dan *high* (51 data).

Tingginya akurasi pada kelas *medium* menunjukkan bahwa model mampu mengenali pola perilaku numerik yang paling dominan di dalam dataset. Untuk kategori **low stress**, model mengklasifikasikan **16 dari 43 data** dengan benar, sementara **27 data** salah diprediksi sebagai *medium*. Kesalahan ini cenderung muncul karena nilai fitur seperti *daily screen time* atau *sleep quality* pada kelas *low* dan *medium* memiliki rentang yang berdekatan, sehingga pola antar kelas menjadi lebih sulit dibedakan. Meski demikian,

model tidak salah memprediksi kelas low menjadi high, yang menunjukkan bahwa model relatif stabil dalam menghindari kesalahan ekstrem. Pada kategori **high stress**, model memberikan prediksi benar sebanyak **91 dari 143 data**, sementara **52 data** salah diprediksi sebagai medium. Pola ini menunjukkan bahwa kelas high memiliki karakteristik perilaku yang cukup kuat, namun tetap terdapat sebagian data yang memiliki kemiripan dengan kelas medium, sehingga menghasilkan misprediksi.

Tidak adanya prediksi high yang salah ke low juga menegaskan bahwa model dapat membedakan perbedaan besar antar tingkat stres. Secara keseluruhan, Confusion Matrix memperlihatkan bahwa kesalahan prediksi paling umum terjadi antara kelas medium dan dua kelas lainnya. Hal ini cukup wajar mengingat kelas medium memiliki rentang nilai fitur yang paling luas dan menjadi perantara antara low dan high. Meski terdapat misprediksi, tingkat kesalahan masih berada dalam batas wajar dan sesuai dengan pola distribusi data pada Bar Plot sebelumnya. Temuan ini memberikan pemahaman penting mengenai kelebihan dan keterbatasan model Random Forest pada dataset numerik serta menjadi dasar untuk perbaikan pada penelitian lanjutan.

3.8 Analisis Prediksi Model

Analisis prediksi model dilakukan untuk mengevaluasi kinerja model Random Forest secara mendalam pada setiap *instance* data uji. Hasil prediksi dari *widget Predictions* Orange Data Mining (ditampilkan pada Gambar 3.8) menyajikan perbandingan antara label aktual (*Actual Label*) dan label prediksi (*Predicted Label*), termasuk probabilitas keyakinan model.

| | Random Forest | error | Stress_Level(1-10) | User_ID | Days_Without_Social_Media | Gender | Sleep_Quality(1-10) | Daily_Screen_Time(hrs) | Age | Use_Frequency(wk) | Wpiness_Index(1-10) |
|----|-----------------------------|-------|--------------------|---------|---------------------------|--------|---------------------|------------------------|-----|-------------------|---------------------|
| 1 | 0.00 : 0.72 : 0.28 → medium | 0.277 | medium | U062 | 3 | Female | 6 | 6.6 | 39 | 1 | 7 |
| 2 | 0.00 : 1.00 : 0.00 → medium | 0.000 | medium | U074 | 1 | Female | 6 | 4.8 | 26 | 4 | 8 |
| 3 | 0.03 : 0.73 : 0.24 → medium | 0.267 | medium | U375 | 1 | Male | 6 | 5.8 | 17 | 5 | 9 |
| 4 | 0.00 : 1.00 : 0.00 → medium | 0.000 | medium | U156 | 4 | Female | 7 | 5.2 | 19 | 4 | 10 |
| 5 | 0.00 : 0.55 : 0.45 → medium | 0.452 | medium | U105 | 1 | Male | 5 | 7.2 | 19 | 0 | 9 |
| 6 | 0.15 : 0.85 : 0.00 → medium | 0.150 | medium | U995 | 3 | Female | 7 | 5.2 | 47 | 2 | 9 |
| 7 | 0.00 : 0.03 : 0.97 → high | 0.025 | high | U378 | 3 | Male | 4 | 8.3 | 41 | 1 | 6 |
| 8 | 0.00 : 0.97 : 0.03 → medium | 0.025 | medium | U125 | 4 | Male | 7 | 5.5 | 16 | 1 | 9 |
| 9 | 0.00 : 0.65 : 0.35 → medium | 0.347 | medium | U009 | 5 | Female | 6 | 4.8 | 24 | 2 | 7 |
| 10 | 0.00 : 0.12 : 0.88 → high | 0.117 | high | U031 | 1 | Female | 5 | 8.1 | 15 | 5 | 7 |

Gambar 6. Table Prediction

Tabel prediksi memperlihatkan bahwa model menunjukkan konsistensi tinggi dalam memprediksi kelas medium, seringkali dengan nilai error 0.000 (misalnya, Baris 2 dan 4). Sebaliknya, kesalahan prediksi ditandai dengan nilai error yang signifikan (misalnya 0.277 pada Baris 1). Pola misklasifikasi cenderung mengarah ke kelas medium; misalnya, Baris 1 menunjukkan low → medium, dan Baris 10 menunjukkan high → medium, meskipun pada Baris 7 model berhasil memprediksi high → high dengan keyakinan tinggi.

Pola Prediksi dan Keterkaitan Confusion Matrix Analisis terperinci berdasarkan kelas aktual dan dihubungkan dengan Confusion Matrix mengkonfirmasi kecenderungan model:

Kelas medium: Menunjukkan kinerja terbaik dengan 249 prediksi benar. Kesalahan misklasifikasi (51 kasus) cenderung menuju kelas high, mengindikasikan bahwa batas antara stres menengah dan tinggi cukup jelas. Kelas low: Model mengalami kesulitan signifikan, di mana 27 dari total 43 kasus low salah diprediksi sebagai medium (27 kasus low → medium). Ini menunjukkan batas pemisah antara low dan medium adalah yang paling sulit ditentukan.

Kelas high: Mayoritas kesalahan (52 kasus) mengarah pada misklasifikasi high → medium. Hanya 91 kasus yang benar diprediksi.

Hasil ini memperjelas bahwa kelas medium berperan sebagai pusat gravitasi misklasifikasi, menyerap kesalahan prediksi dari kelas low dan high (total 27 + 52 = 79 kasus). Keterkaitan dengan Bar Plot (Bagian 3.6) menyiratkan bahwa model Random Forest membuat keputusan berdasarkan kombinasi fitur yang kompleks, di mana perbedaan tipis pada fitur penting (seperti Sleep Quality atau Daily Screen Time) dapat memicu kesalahan misklasifikasi menuju kelas dominan medium. Meskipun variabel Daily Screen Time menunjukkan variasi tinggi di semua kasus (misalnya, 4.8 jam pada prediksi benar vs. 8.1 jam pada prediksi salah), tidak ada satu fitur pun yang mendominasi, menegaskan ketergantungan model pada interaksi fitur.

3.9 Interpretasi Hasil dan Pembahasan

Bagian ini merangkum keseluruhan proses analisis yang telah dilakukan mulai dari tahap preprocessing, pemilihan variabel, pelatihan model, hingga evaluasi performa menggunakan berbagai visualisasi. Berdasarkan workflow penelitian pada Orange Data Mining, setiap tahapan berjalan secara terstruktur—dimulai dari pemuatan dataset, penyesuaian tipe variabel, pembentukan model Random Forest, hingga penilaian performa melalui *Test and Score*, *Bar Plot*, *Confusion Matrix*, dan *Predictions*. Hasil analisis menunjukkan bahwa model Random Forest mampu mempelajari pola perilaku pengguna dengan baik, terutama pada kelas **medium**, yang menjadi kategori dengan jumlah data terbesar sekaligus tingkat prediksi paling konsisten. Visualisasi *Bar Plot* menunjukkan adanya variasi perilaku yang cukup jelas antar tingkatan stres, khususnya pada variabel *Daily Screen Time*.

Temuan ini diperkuat oleh Confusion Matrix, di mana kelas medium memiliki jumlah prediksi benar paling tinggi, sedangkan misklasifikasi banyak terjadi dari kelas low → medium dan high → medium. Hal ini mengindikasikan bahwa kelas medium memiliki karakteristik perilaku yang berada di tengah-tengah kelas low dan high, sehingga menjadi pusat kecenderungan prediksi model. Analisis pada node *Predictions* menunjukkan bahwa pola prediksi model sesuai dengan karakteristik fitur dalam dataset, terutama pola yang melibatkan durasi penggunaan layar, kualitas tidur, dan indeks kebahagiaan. Hasil ini konsisten dengan kecenderungan misklasifikasi pada Confusion Matrix, serta mendukung kesimpulan bahwa Random Forest bekerja paling optimal pada data numerik.

dengan variasi yang terdistribusi secara jelas. Secara keseluruhan, seluruh rangkaian tahapan pada BAB 3 menunjukkan bahwa workflow penelitian yang digunakan mampu menjalankan fungsi analisis secara komprehensif. Model Random Forest bekerja stabil dalam memetakan tingkat stres berdasarkan fitur perilaku pengguna, meskipun masih terdapat misklasifikasi pada kelas tertentu akibat kemiripan pola numerik. Hasil ini memperkuat relevansi penggunaan teknik data mining dan algoritma Random Forest dalam mendeteksi stres berbasis data numerik dan visualisasi analitik.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan memprediksi tingkat stres pengguna berbasis data numerik yang merepresentasikan perilaku sehari-hari, seperti kualitas tidur, intensitas penggunaan media sosial, frekuensi olahraga, serta indeks kebahagiaan. Melalui pendekatan *data mining* menggunakan algoritma Random Forest pada platform Orange Data Mining, penelitian ini berhasil membangun model klasifikasi yang mampu memberikan prediksi tingkat stres secara cukup akurat dan konsisten. Seluruh proses—mulai dari pemuatan dataset, pemeriksaan struktur data, pemilihan fitur, pengaturan domain variabel, hingga pembentukan model dan evaluasi—dilakukan dalam workflow terstruktur yang memastikan kejelasan alur dan integritas analisis. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki performa stabil pada metrik akurasi, precision, recall, dan F1-score. Visualisasi Bar Plot memperlihatkan perbedaan pola numerik yang relevan antar kelas stres, terutama pada variabel yang berkaitan dengan penggunaan media sosial dan kualitas tidur. Temuan ini diperkuat oleh Confusion Matrix, di mana kelas medium menjadi kategori dengan prediksi paling konsisten, sementara misklasifikasi paling banyak terjadi dari kelas low dan high menuju medium. Analisis node Predictions juga menunjukkan bahwa model mampu memetakan pola perilaku dengan baik meskipun masih terdapat beberapa kesalahan prediksi pada kelas yang memiliki karakteristik berdekatan. Secara keseluruhan, penelitian ini menyimpulkan bahwa algoritma Random Forest efektif digunakan dalam mendeteksi tingkat stres berdasarkan data perilaku numerik. Pendekatan workflow berbasis Orange terbukti memudahkan proses analisis yang sistematis dan transparan. Penelitian ini memberikan gambaran bahwa perilaku digital dan faktor gaya hidup memiliki hubungan yang kuat dengan tingkat stres, serta menunjukkan bahwa model klasifikasi berbasis machine learning dapat menjadi alat pendukung yang potensial dalam analisis kesehatan mental berbasis data.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

REFERENCES

- [1] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress Detection Using Natural Language Processing and Machine Learning," *J. Big Data*, vol. 9, no. 1, p. 33, 2022, [Online]. Available: <https://doi.org/10.1186/s40537-022-00575-6>
- [2] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing," *Human-Centric Intell. Syst.*, vol. 3, no. 2, pp. 80–91, 2023, [Online]. Available: [10.1007/s44230-023-00020-8](https://doi.org/10.1007/s44230-023-00020-8)
- [3] E. Turcan and K. McKeown, "Dreaddit: A Reddit Dataset for Stress Analysis," *Proc. Int. AAAI Conf. Web Soc. Media*, pp. 97–107, 2019, [Online]. Available: <https://aclanthology.org/D19-6213.pdf>
- [4] A. Rastogi, Q. Liu, and E. Cambria, "Stress Detection from Social Media Articles: A New Dataset Benchmark and Analytical Study," *IEEE Int. Conf. Data Min.*, pp. 1–8, 2022.
- [5] Y.-C. Yang, Y.-H. Hung, and K.-C. Lee, "Automatic Detection of Twitter Users Who Express Chronic Stress," *J. Biomed. Heal. Informatics*, vol. 41, no. 9, pp. 717–724, 2022, [Online]. Available: [10.1097/CIN.0000000000000985](https://doi.org/10.1097/CIN.0000000000000985)
- [6] T. Kasmin, "Stress Detection Through Text in Social Media Using Machine Learning," *Int. J. Adv. Comput. Sci.*, vol. 61, no. 4, pp. 161–175, 2024, [Online]. Available: <https://doi.org/10.37934/araset.61.4.161175>
- [7] M. Illahi, A. A. Memon, and S. A. Abbasi, "An Ensemble Machine Learning Approach for Stress Detection in Social Media Texts," *Quaid-E-Awam Univ. Res. J.*, vol. 20, no. 02, pp. 123–128, 2022, [Online]. Available: [10.52584/QRJ.2002.15](https://doi.org/10.52584/QRJ.2002.15)
- [8] A. Priya, "Predicting Anxiety, Depression, and Stress in Modern Life Using Machine Learning Models," *Int. J. Intell. Syst. Appl.*, vol. 167, pp. 442–451, 2020, [Online]. Available: [10.1016/j.procs.2020.03.442](https://doi.org/10.1016/j.procs.2020.03.442)
- [9] X. Zhang, Y. Liu, and L. Wang, "Identifying the Predictors of Severe Psychological Distress by Interpretable Machine Learning Methods," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 01, p. 335, 2023.
- [10] U. Srinivasarao, S. Babu, and M. Singh, "A Novel and Efficient Personalized Stress Detection Using Machine Learning Models," *Sci. Rep.*, vol. [volume], no. [issue], p. [start page]–[end page], 2025, [Online]. Available: [\[url or doi\]](#)
- [11] S. Inamdar, R. Chapekar, and S. Gite, "Machine Learning Models for Cognitive Stress Detection Using Social Media Text," *Int. J. Cogn. Informatics*, vol. 3, no. 2, pp. 80–91, 2023, [Online]. Available: [10.1007/s44230-023-00020-8](https://doi.org/10.1007/s44230-023-00020-8)
- [12] M. R. Febriansyah, "Stress Detection System for Social Media Users Based on Machine Learning," *J. Soft Comput. Decis. Support Syst.*, vol. 216, pp. 672–681, 2023, [Online]. Available: [10.1016/j.procs.2022.12.183](https://doi.org/10.1016/j.procs.2022.12.183)
- [13] P. E. Elisa and A. R. Isnain, "Comparison of Random Forest, Support Vector Machine, and Naive Bayes Algorithms to Analyze Sentiment Toward Mental Health Stigma," *J. Teknol. Inf. dan Terap.*, vol. 5, no. 2, pp. 37–44, 2024, [Online]. Available: [10.25047/jtit.v5i2.478](https://doi.org/10.25047/jtit.v5i2.478)
- [14] N. F. Mahing, A. L. Gunawan, A. Foresta, and A. Zen, "KLASIFIKASI TINGKAT STRES DARI DATA BERBENTUK TEKS DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CLASSIFICATION OF STRESS LEVELS FROM TEXT



DATA USING SUPPORT VECTOR MACHINE (SVM) ALGORITHM AND RANDOM FOREST,” vol. 11, no. 5, 2024, doi: 10.25126/jtiik2024118010.

- [15] R. F. Surakusumah and R. Y. Apza, “Adaptive Stress Prediction with GSR, SMOTE Balancing, and Random Forest Models,” *J. RESTI*, vol. 9, no. 4, pp. 805–813, 2025, [Online]. Available: 10.29207/resti.v9i4.6588
- [16] V. Oktaviani, N. Rosmawarni, and M. P. Muslim, “Perbandingan Kinerja Random Forest dan SMOTE-Random Forest dalam Mendeteksi dan Mengukur Tingkat Stres Mahasiswa Tingkat Akhir,” *Inform. J. Ilmu Komput.*, vol. 20, no. 01, pp. 43–49, 2025.
- [17] Y. Tolla and Kusriani, “Twitter-based Stress and Depression Detection Using SVM and NLP,” *J. Inform.*, 2025, [Online]. Available: <https://ejurnal.umri.ac.id/index.php/JIK/article/view/9067>
- [18] M. E. Johan and S. A. Azka, “Deteksi Stres Menggunakan Multilayer Perceptron dan Indonesian Sentiment Lexicon,” *G-Tech J. Inf. Technol.*, 2022, [Online]. Available: <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/2611>
- [19] R. Rachmat, I. Nurhayati, and W. Huljannah, “Analisis Sentimen Ulasan Aplikasi SATUSEHAT Menggunakan Random Forest,” *Pros. Teknol. Inf.*, 2023, [Online]. Available: <https://eprosiding-old.ars.ac.id/index.php/pti/article/view/1148>
- [20] Fitri, Asriyanik, and Apriandari, “Deteksi Kecemasan Generasi Z Menggunakan Random Forest Berbasis Media Sosial,” *J. Inform. dan Tek. Elektro Terap.*, 2025, [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/6905>
- [21] R. Putra, D. Widyastuti, and H. A. Satria, “Machine Learning Classification Using Public Datasets from Kaggle,” *J. Inf. Syst. Res.*, vol. 6, no. 1, pp. 10–18, 2023, [Online]. Available: <https://journal-isr.id/index.php/isr/article/view/83>
- [22] A. Savić and A. Pribičević, “Using Orange Data Mining for Machine Learning Education,” *TEM J.*, vol. 9, no. 2, pp. 806–812, 2020, doi: 10.18421/TEM92-49.
- [23] U. Cahyono, “Using Mendeley as a Reference Manager in Academic Writing,” *J. Educ. Learn.*, vol. 13, no. 2, pp. 123–131, 2019, doi: 10.11591/edulearn.v13i2.