

Klasifikasi Kepribadian Berdasarkan Dimensi Ekstraversi Berbasis Data Mining Menggunakan *Extremely Randomized Trees*

Yanuardi ^{a,1,*}; Firdiansyah Firdaus Basri ^{b,2}; Muhammad Luthfi Aksani ^{c,3}

^{a,b,c} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang
Jl. Perintis Kemerdekaan I No.33, Cikokol, Kecamatan Tangerang, Kota Tangerang, Banten 15118

¹yanuardi99@gmail.com, ²firdi.ian@gmail.com, ³L13fie@gmail.com

Keywords:

Personality Classification, Extraversion Dimension, Data Mining, Extremely Randomized Trees, Machine Learning

Abstract

Personality is one of the fundamental aspects that distinguishes individual behavior, thought patterns, and interaction styles. The extraversion dimension, which is part of the Big Five Personality Traits framework, reflects an individual's tendency to engage in social interactions with two main poles, namely introvert and extrovert. Identifying personality based on this dimension has various applications, ranging from education to employee recruitment. This study aims to develop a personality classification model based on the extraversion dimension using the Extremely Randomized Trees (ERT) algorithm and to compare its performance with other algorithms, namely Decision Tree, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). The dataset used in this study was obtained from the Kaggle platform, consisting of 2,900 entries and including social behavior indicators represented by five numerical variables and two categorical variables. The research methodology involves data preprocessing, exploratory data analysis, model construction, and evaluation using confusion matrix, precision, recall, F1-score, accuracy, and ROC-AUC. The results indicate that ERT achieves the best performance compared to the other algorithms. The ERT model obtained an accuracy of 92.69%, an F1-score of 0.9269, and a ROC-AUC of 0.9429, outperforming SVM (F1 0.9173; AUC 0.9300), KNN (F1 0.9086; AUC 0.9146), and Decision Tree (F1 0.8879; AUC 0.8876). The superiority of ERT is supported by its tree-based ensemble mechanism with high randomization, which enhances generalization, reduces variance, and captures complex non-linear interactions among behavioral variables. Therefore, ERT is proven to be effective in consistently distinguishing introvert and extrovert tendencies.

Pendahuluan

Kepribadian merupakan salah satu aspek fundamental yang membedakan perilaku, pola pikir, dan gaya interaksi setiap individu. Dalam psikologi modern, teori *Big Five Personality Traits* atau model OCEAN menjadi salah satu kerangka yang paling banyak digunakan untuk mengkaji kepribadian manusia [1]. Salah satu dimensi penting dalam model tersebut adalah ekstraversi (*extraversion*) yang menggambarkan kecenderungan individu untuk bersosialisasi, aktif berinteraksi, serta mendapatkan energi dari lingkungan sosial [2]. Dimensi ekstraversi ini berada pada dua kutub ekstrem, yaitu *introvert* yang cenderung lebih tertutup, dan *extrovert* yang lebih terbuka terhadap interaksi sosial [3]. Identifikasi kepribadian berdasarkan dimensi ekstraversi memiliki beragam aplikasi, seperti dalam bidang pendidikan untuk memahami gaya belajar siswa, dalam rekrutmen karyawan untuk menilai kesesuaian kandidat, maupun dalam sistem rekomendasi yang menyesuaikan preferensi pengguna. Selama ini, pengukuran ekstraversi banyak dilakukan melalui metode psikotes konvensional yang membutuhkan waktu, biaya, dan sumber daya manusia yang memadai. Peningkatan ketersediaan indikator perilaku yang bersifat *data-driven* mendorong pemanfaatan *data mining* untuk pengklasifikasian dimensi kepribadian secara terukur, skalabel, dan relatif tidak intrusif dibanding asesmen manual.

Penelitian terdahulu telah menggunakan berbagai algoritma untuk klasifikasi kepribadian. Salah satunya menerapkan *Naïve Bayes Classifier* dengan teknik SMOTE pada data Twitter untuk klasifikasi MBTI, dan hasilnya menunjukkan bahwa *Logistic Regression* mampu mencapai akurasi tertinggi sekitar 80%, meskipun masih lemah pada kelas minoritas [4]. Penelitian lain mengembangkan sistem klasifikasi sidik jari berbasis citra digital

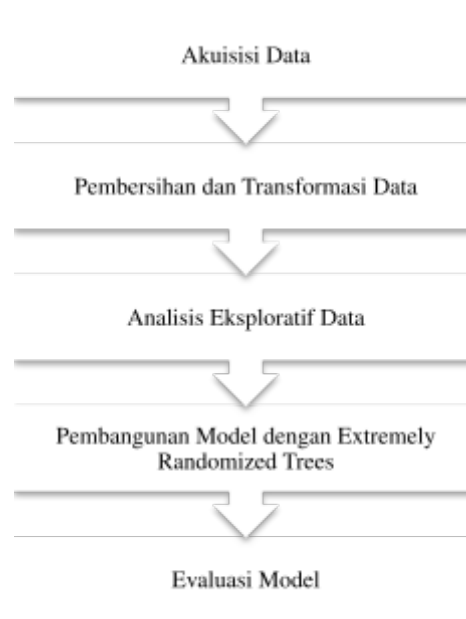
menggunakan *Support Vector Machine* (SVM) kernel RBF, dan membuktikan bahwa pola sidik jari (*arch*, *loop*, *whorl*) memiliki keterkaitan dengan kecenderungan kepribadian dengan akurasi 89.25% [5]. Selanjutnya, penelitian lain melakukan prediksi kepribadian berdasarkan status Facebook menggunakan *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) dengan model *Big Five Personality Traits* [6]. Hasil penelitian menunjukkan bahwa *Naïve Bayes* mencapai akurasi 98.80% pada dataset tertentu, sementara KNN hanya 58.96%. Penelitian lainnya membandingkan algoritma C4.5 dengan *Naïve Bayes* untuk menentukan kepribadian *introvert* dan *extrovert* pada siswa SMK, dengan hasil akurasi C4.5 sebesar 60.69% dan *Naïve Bayes* 60.00% [7].

Algoritma seperti *Decision Tree*, SVM, dan KNN cukup populer karena sederhana dan dapat bekerja baik dengan data tabular. Namun, masing-masing memiliki keterbatasan: *Decision Tree* rawan *Decision Tree* [8], KNN sensitif terhadap skala data [9], dan SVM membutuhkan pemilihan kernel yang tepat agar optimal [10]. Di sisi lain, algoritma *Extremely Randomized Trees* (ERT) atau bisa disebut juga *Extra Trees* sebagai varian dari *ensemble* dari *Decision Tree* relatif jarang dieksplorasi dalam konteks klasifikasi ekstrasversi, padahal memiliki beberapa keunggulan mendasar. *Extra Trees* menggunakan proses pembentukan pohon dengan tingkat randomisasi lebih tinggi dibanding *Random Forest*, baik pada pemilihan fitur maupun titik *split* [11]. Hal ini membuat *Extra Trees* lebih efisien secara komputasi, tahan terhadap *Decision Tree*, serta mampu memberikan generalisasi yang lebih baik pada data tabular berskala kecil hingga menengah [12].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi kepribadian berdasarkan dimensi ekstrasversi menggunakan algoritma *Extremely Randomized Trees* (ERT), serta membandingkan kinerjanya dengan beberapa algoritma lain seperti *Decision Tree*, KNN, dan SVM. Kontribusi utama penelitian ini adalah memberikan analisis komparatif performa *Extra Trees* terhadap algoritma lain dalam konteks klasifikasi ekstrasversi, sekaligus mengidentifikasi fitur-fitur yang paling berpengaruh dalam menentukan kecenderungan *introvert* maupun *extrovert*. Dengan demikian, penelitian ini diharapkan dapat memperkaya literatur sekaligus menawarkan pendekatan praktis bagi aplikasi berbasis kepribadian.

Metode penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun model klasifikasi kepribadian berdasarkan dimensi ekstrasversi. Pendekatan ini melibatkan serangkaian tahapan sistematis, mulai dari akuisisi data, pembersihan dan transformasi data, analisis eksploratif data, pembangunan model menggunakan algoritma *Extremely Randomized Trees* (ERT), hingga evaluasi model. Tahapan-tahapan tersebut divisualisasikan dalam Gambar 1.



Gambar 1. Prosedur penelitian yang diterapkan

A. Akuisisi Data

Dataset yang digunakan dalam penelitian ini diperoleh dari *platform* Kaggle dengan nama "*Extrovert vs Introvert Behavior Data*" [13]. Dataset ini berisi 2.900 entri dengan 8 atribut yang berkaitan dengan perilaku sosial individu. Fitur yang tersedia mencakup indikator numerik seperti *Time_spent_Alone* (waktu yang dihabiskan sendirian), *Social_event_attendance* (frekuensi menghadiri acara sosial), *Going_outside* (kecenderungan untuk keluar rumah), *Friends_circle_size* (ukuran lingkaran pertemanan), dan *Post_frequency* (frekuensi unggahan di media sosial). Selain itu, terdapat fitur kategorikal berupa *Stage_fear* (ketakutan tampil di depan umum) dan *Drained_after_socializing* (kondisi energi setelah bersosialisasi). Label target pada dataset ini adalah kepribadian berdasarkan dimensi ekstrasversi, yang terdiri dari dua kategori, yaitu *Introvert* dan *Extrovert*. Dataset ini dipilih karena menyajikan indikator perilaku yang relevan dengan kecenderungan ekstrasversi, sekaligus memiliki ukuran data yang cukup besar untuk melatih dan menguji model klasifikasi secara andal.

B. Pembersihan dan Transformasi Data

Tahapan ini mencakup konversi nilai kategorikal menjadi numerik (misalnya *Yes/No* menjadi 1/0), pemeriksaan nilai kosong, serta pemisahan data menjadi data latih dan data uji. Pemisahan data dilakukan dengan rasio 80% untuk pelatihan dan 20% untuk pengujian secara stratifikasi, karena pembagian ini dianggap seimbang antara menyediakan data yang cukup banyak untuk melatih model sekaligus menyisakan data yang representatif untuk menguji kinerja secara objektif. *Output* dari tahap ini adalah dua subset data yang bersih dan terstruktur, yaitu data latih yang siap digunakan untuk membangun model dan data uji yang disiapkan untuk mengevaluasi performa model.

C. Analisis Eksploratif Data

Tahap eksplorasi data dilakukan untuk memahami distribusi variabel numerik maupun kategorikal, serta hubungan antara fitur dengan label target [14]. Visualisasi distribusi kelas, histogram variabel numerik, dan heatmap korelasi digunakan untuk menemukan pola awal yang dapat mendukung pemodelan. Visualisasi distribusi kelas digunakan untuk memastikan keseimbangan jumlah data pada masing-masing kategori kepribadian sehingga evaluasi model tidak bias terhadap kelas mayoritas. Histogram variabel numerik digunakan untuk melihat pola sebaran nilai dan perbedaan distribusi antar kelas, yang dapat memberikan indikasi awal mengenai fitur yang paling relevan dalam memprediksi ekstrasversi. Sementara itu, heatmap korelasi digunakan untuk mengidentifikasi hubungan antar variabel numerik, sehingga dapat diketahui adanya potensi multikolinearitas atau keterkaitan fitur yang kuat. Hasil visualisasi ini memberikan gambaran awal yang penting dalam mendukung pemodelan dan interpretasi hasil klasifikasi.

D. Pembangunan Model dengan *Extremely Randomized Trees* (ERT)

Algoritma utama yang digunakan dalam penelitian ini adalah *Extremely Randomized Trees* (ERT) atau biasa disebut dengan *Extra Trees*, sebuah metode *ensemble* berbasis pohon keputusan. Berbeda dengan *Random Forest* yang menggunakan pemilihan atribut terbaik berdasarkan suatu kriteria, *Extra Trees* menggunakan pemilihan atribut dan titik pemisahan secara acak untuk membentuk pohon keputusan [15]. Strategi ini meningkatkan keragaman antar pohon sekaligus mengurangi risiko *overfitting* [16].

Secara umum, model *Extra Trees* dibangun dari himpunan pohon keputusan yang masing-masing dilatih pada data pelatihan [17]. Prediksi akhir untuk data uji ditentukan melalui mekanisme voting mayoritas untuk klasifikasi [18]. Misalkan terdapat T pohon, maka hasil prediksi $\hat{y}$ untuk sebuah sampel x diberikan oleh persamaan (1).

$$\hat{y} = \arg \arg \sum_{t=1}^T I(h_t(x) = c) \quad (1)$$

di mana C adalah himpunan kelas (*introvert*, *extrovert*), $h_t(x)$ adalah hasil prediksi pohon ke- t , dan I adalah fungsi indikator.

Kelebihan utama algoritma ERT terletak pada kemampuannya membangun model yang efisien, tahan terhadap varians tinggi, dan tetap stabil ketika digunakan pada data tabular berskala menengah [19]. Karakteristik ini membuat model mampu memberikan prediksi yang konsisten, meminimalkan risiko *Decision Tree*, serta memiliki tingkat generalisasi yang baik pada berbagai jenis dataset. Dengan demikian, *Extra Trees* sangat sesuai untuk klasifikasi kepribadian berbasis data perilaku, yang secara alami memiliki variabilitas tinggi antar individu.

E. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja klasifikasi berdasarkan data uji [20]. Tahap awal dimulai dengan *confusion matrix* yang memperlihatkan distribusi prediksi benar dan salah melalui empat komponen utama: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Dari matriks ini diturunkan berbagai metrik evaluasi, antara lain akurasi yang menunjukkan proporsi prediksi benar, *precision* yang menilai ketepatan prediksi positif, *recall* yang mengukur kemampuan menemukan seluruh data positif, serta *F1-score* yang

merepresentasikan keseimbangan antara *precision* dan *recall* [21]. Untuk analisis lebih lanjut digunakan nilai ROC-AUC sebagai ukuran kuantitatif yang merepresentasikan seberapa baik model mampu membedakan antar kelas [22]. Dalam penelitian ini, evaluasi tidak hanya difokuskan pada algoritma ERT, tetapi juga membandingkannya dengan *Decision Tree*, K-Nearest Neighbor (KNN), dan *Support Vector Machine* (SVM). Perbandingan berdasarkan akurasi, *precision*, *recall*, *F1-score*, serta AUC memberikan gambaran komprehensif mengenai kinerja dan keunggulan relatif masing-masing algoritma dalam klasifikasi kepribadian berbasis dimensi ekstraversi.

Hasil dan diskusi

Dalam pengembangan model klasifikasi kepribadian berdasarkan dimensi ekstraversi berbasis data mining menggunakan *Extremely Randomized Trees* (ERT) atau *Extra Trees*, langkah pertama yang dilakukan adalah menyiapkan dataset untuk proses pelatihan dan pengujian. Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan nama "*Extrovert vs Introvert Behavior Data*" [13]. Dataset ini berjumlah 2.900 entri yang terdiri atas 7 fitur prediktor (5 numerik dan 2 kategorikal) serta 1 label target (Personality). Karakteristik dataset secara rinci ditunjukkan pada Tabel 1.

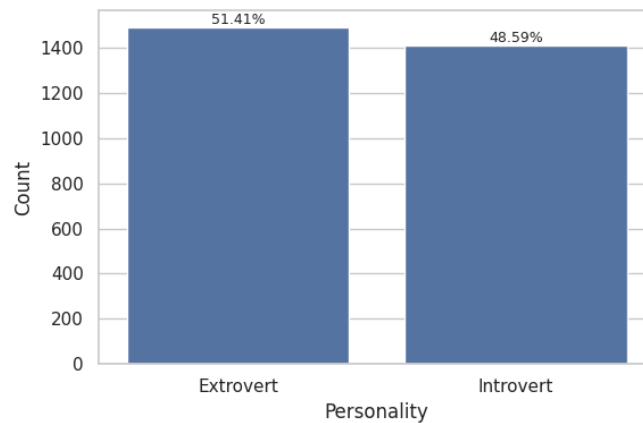
Tabel 1. Karakteristik dataset

Fitur	Tipe Data	Jumlah Nilai Unik	Contoh Nilai
<i>Time_spent_Alone</i>	float64	13	4.0, 9.0, 0.0, 3.0, 1.0
<i>Stage_fear</i>	object	2	No, Yes
<i>Social_event_attendance</i>	float64	12	4.0, 0.0, 1.0, 6.0, 9.0
<i>Going_outside</i>	float64	8	6.0, 0.0, 2.0, 7.0, 4.0
<i>Drained_after_socializing</i>	object	2	No, Yes
<i>Friends_circle_size</i>	float64	17	13.0, 0.0, 5.0, 14.0, 8.0
<i>Post_frequency</i>	float64	12	5.0, 3.0, 2.0, 8.0, 6.0
<i>Personality</i>	object	2	<i>Extrovert, Introvert</i>

Tabel 1 menunjukkan bahwa dataset memiliki kombinasi fitur numerik dan kategorikal yang mencerminkan perilaku sosial individu, seperti waktu yang dihabiskan sendirian, frekuensi menghadiri acara sosial, kecenderungan keluar rumah, ukuran lingkaran pertemanan, frekuensi unggahan, ketakutan tampil (*stage fear*), serta kondisi energi setelah bersosialisasi. Label target berupa kepribadian dibagi menjadi dua kategori, yaitu *introvert* dan *extrovert*.

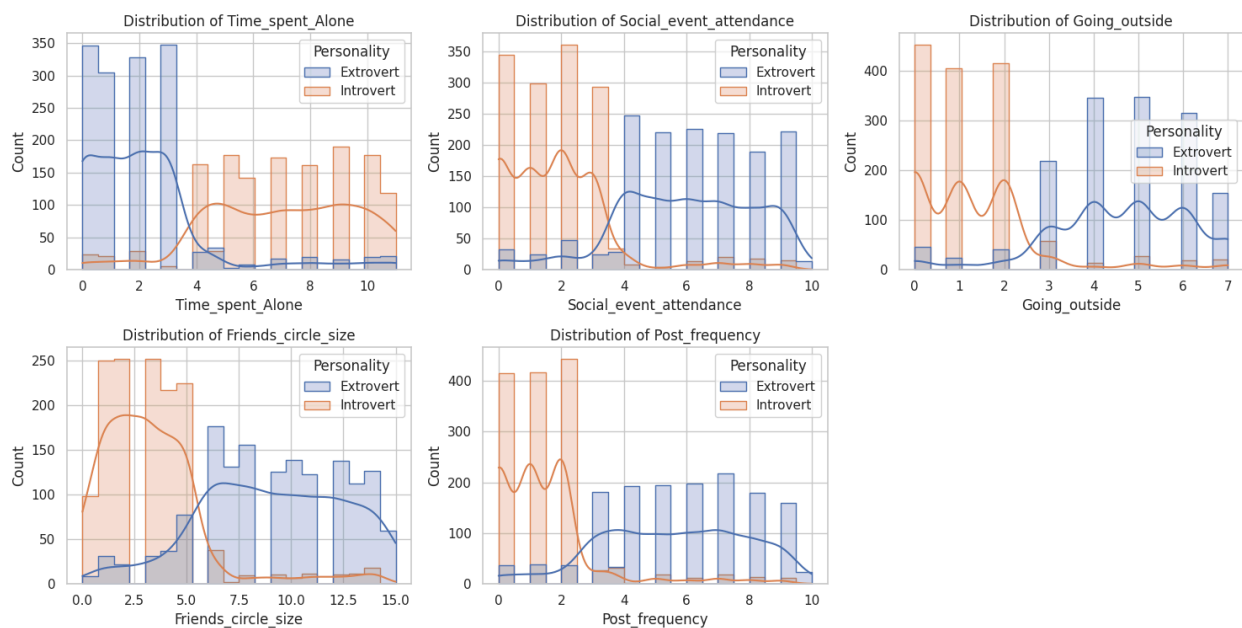
Setelah dataset tersedia, langkah berikutnya adalah melakukan pembersihan dan transformasi data untuk memastikan dataset siap digunakan dalam pemodelan. Proses ini mencakup pemeriksaan nilai kosong, yang menunjukkan bahwa dataset tidak memiliki data hilang sehingga tidak memerlukan imputasi tambahan. Nilai kategorikal seperti *Stage_fear* dan *Drained_after_socializing* dikonversi menjadi bentuk numerik biner dengan representasi Yes = 1 dan No = 0, sedangkan label target *Personality* diubah menjadi *Introvert* = 0 dan *Extrovert* = 1. Tidak seperti algoritma berbasis jarak atau proyeksi ruang vektor, *Extremely Randomized Trees* (ERT) tidak memerlukan normalisasi atau standarisasi fitur numerik. Hal ini karena mekanisme utama algoritma ini adalah melakukan pemisahan node pada pohon keputusan berdasarkan nilai ambang (*threshold*) pada masing-masing fitur, bukan menghitung jarak antar data. Selanjutnya, dataset dibagi menjadi 80% data latih dan 20% data uji secara stratifikasi, sehingga proporsi *introvert* dan *extrovert* tetap seimbang di kedua subset.

Tahap berikutnya adalah eksplorasi data untuk memahami distribusi variabel numerik maupun kategorikal, serta hubungan antara fitur dengan label target. Analisis awal dilakukan pada distribusi kelas untuk memastikan keseimbangan jumlah data pada masing-masing kategori kepribadian sehingga evaluasi model tidak bias terhadap kelas mayoritas. Distribusi data target dapat dilihat pada Gambar 2.



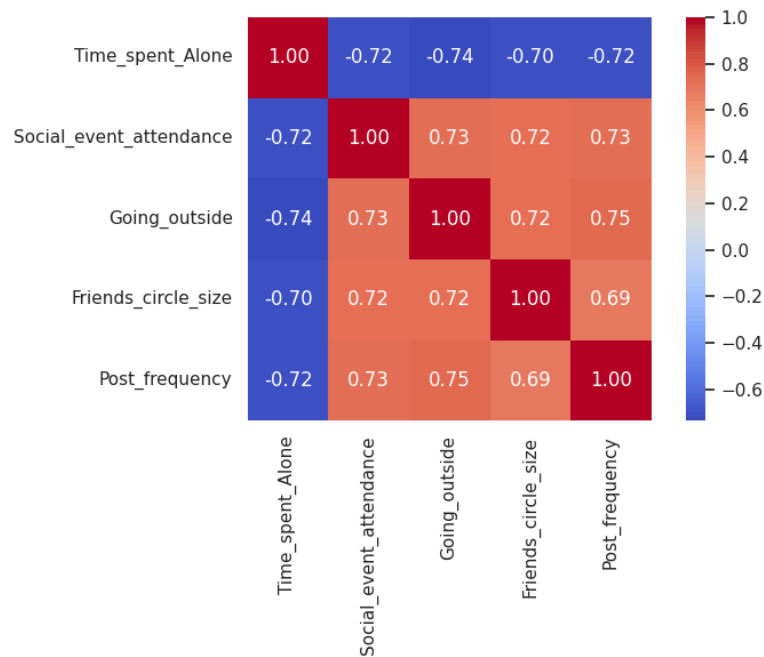
Gambar 2. Distribusi data target berdasarkan dimensi ekstrasversi

Gambar 2 menunjukkan bahwa kelas *extrovert* memiliki proporsi sebesar 51.41%, sedangkan kelas *introvert* sebesar 48.59%. Perbedaan distribusi yang relatif kecil ini mengindikasikan bahwa dataset cukup seimbang, sehingga setiap model klasifikasi yang dibangun dapat dievaluasi secara adil tanpa perlu menerapkan teknik penyeimbangan kelas tambahan. Eksplorasi data berikutnya dilakukan pada variabel numerik untuk melihat pola sebaran nilai dan perbedaan distribusi antar kelas, yang dapat memberikan indikasi awal mengenai fitur-fitur yang paling relevan dalam memprediksi ekstrasversi. Visualisasi variabel numerik ditunjukkan pada Gambar 3.



Gambar 3. Distribusi variabel numerik berdasarkan kelas kepribadian

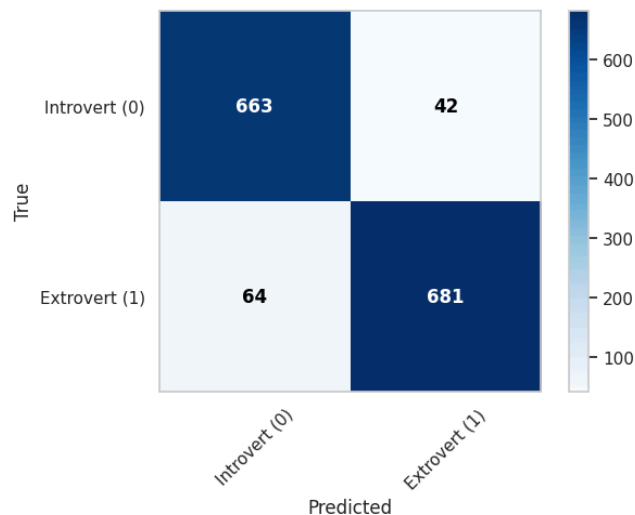
Gambar 3 menunjukkan bahwa terdapat pola distribusi yang berbeda antara *introvert* dan *extrovert* pada sebagian besar variabel numerik. Variabel *Time_spent_Alone* cenderung lebih tinggi pada kelompok *introvert*, sedangkan *extrovert* lebih dominan pada *Social_event_attendance* dan *Going_outside*. Demikian pula, *extrovert* memiliki lingkaran pertemanan (*Friends_circle_size*) yang lebih besar serta frekuensi unggahan (*Post_frequency*) yang lebih tinggi dibandingkan *introvert*. Pola-pola ini mengindikasikan bahwa fitur numerik tersebut memiliki peranan penting dalam membedakan kecenderungan kepribadian berdasarkan dimensi ekstrasversi. Berikutnya dilakukan eksplorasi data korelasi untuk mengidentifikasi hubungan antar variabel numerik, sehingga dapat diketahui adanya potensi multikolinearitas maupun keterkaitan fitur yang kuat. Visualisasi korelasi ditunjukkan pada Gambar 4.



Gambar 4. Peta korelasi antar variabel numerik menggunakan *Pearson Correlation*

Gambar 4 menunjukkan bahwa sebagian besar variabel numerik memiliki korelasi yang tinggi satu sama lain. Variabel *Time_spent_Alone* berkorelasi negatif cukup kuat dengan variabel lain, khususnya *Going_outside* (-0.74), yang sesuai dengan logika bahwa semakin banyak waktu yang dihabiskan sendirian maka semakin jarang seseorang keluar rumah. Sebaliknya, variabel *Social_event_attendance*, *Going_outside*, *Friends_circle_size*, dan *Post_frequency* memiliki korelasi positif yang cukup tinggi (0.69–0.75), menunjukkan bahwa individu yang sering menghadiri acara sosial cenderung memiliki lingkaran pertemanan lebih besar dan lebih aktif dalam aktivitas sosial maupun media. Pola korelasi ini mengindikasikan adanya keterkaitan perilaku yang konsisten dengan kecenderungan ekstrasversi, sekaligus menjadi indikasi awal fitur-fitur yang relevan untuk klasifikasi kepribadian.

Proses selanjutnya adalah membangun model klasifikasi kepribadian berdasarkan dimensi ekstrasversi menggunakan algoritma *Extremely Randomized Trees* (ERT). Setelah dataset dipersiapkan melalui tahap pra-pemrosesan, data dibagi ke dalam dua subset, yaitu data latih (80%) dan data uji (20%), menggunakan metode stratifikasi untuk menjaga proporsi kelas *introvert* dan *extrovert* tetap konsisten pada kedua subset. Model ERT dibangun menggunakan parameter dasar dan dilatih pada data latih. Berbeda dengan algoritma berbasis pohon tunggal, ERT bekerja dengan membangun sejumlah pohon keputusan yang dipisahkan secara acak baik pada pemilihan fitur maupun titik pemisahan, sehingga menghasilkan model yang lebih stabil, efisien, dan tahan terhadap *Decision Tree*. Setelah model selesai dilatih, prediksi dilakukan pada data uji dan hasilnya dievaluasi untuk mengukur kinerja klasifikasi. Evaluasi diawali dengan penggunaan *confusion matrix* untuk menganalisis distribusi prediksi yang benar dan salah pada masing-masing kelas. *Confusion matrix* yang dihasilkan ditampilkan pada Gambar 5.



Gambar 4. Hasil evaluasi klasifikasi menggunakan *confusion matrix*

Gambar 5 menunjukkan bahwa model ERT berhasil mengklasifikasikan mayoritas data dengan benar, di mana terdapat 663 data *introvert* dan 681 data *extrovert* yang terprediksi tepat. Kesalahan klasifikasi relatif kecil, yaitu 42 data *introvert* yang salah diprediksi sebagai *extrovert* dan 64 data *extrovert* yang salah diprediksi sebagai *introvert*. Hasil ini mengindikasikan bahwa model ERT memiliki kemampuan generalisasi yang baik dalam membedakan kecenderungan *introvert* dan *extrovert* berdasarkan fitur perilaku yang tersedia.

Berdasarkan *confusion matrix* yang telah diperoleh, selanjutnya dihitung metrik evaluasi berupa *precision*, *recall*, *F1-score*, akurasi, dan ROC-AUC untuk masing-masing kelas. Hasil evaluasi ditunjukkan pada Tabel 2.

Tabel 2. Hasil evaluasi performa model ERT berdasarkan metrik klasifikasi

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>	ROC-AUC
<i>Introvert</i>	0.9120	0.9404	0.9260	0.9269	0.9429
<i>Extrovert</i>	0.9419	0.9141	0.9278		

Tabel 2 menunjukkan bahwa model ERT mampu memberikan performa yang seimbang pada kedua kelas. Untuk kelas *introvert*, *precision* mencapai 0.9120 dengan *recall* 0.9404 dan *F1-score* 0.9260. Sementara itu, pada kelas *extrovert*, *precision* mencapai 0.9419 dengan *recall* 0.9141 dan *F1-score* 0.9278. Secara keseluruhan, model menghasilkan akurasi 0.9269 dan nilai ROC-AUC 0.9429, yang menandakan kemampuan model cukup tinggi dalam membedakan *introvert* dan *extrovert*. Hasil ini mengindikasikan bahwa ERT mampu mengklasifikasikan kedua kategori dengan tingkat konsistensi yang baik, serta memiliki keseimbangan antara ketepatan (*precision*) dan kelengkapan (*recall*).

Berikutnya dilakukan evaluasi dengan membandingkan model ERT yang telah dibangun dengan *Decision Tree*, KNN, dan SVM. Perbandingan ini dipilih karena keempatnya mewakili *inductive bias* yang berbeda. *Decision Tree* merepresentasikan model pohon tunggal yang sederhana dan mudah diinterpretasikan, KNN merupakan algoritma berbasis jarak yang melakukan klasifikasi berdasarkan kedekatan antar instance, SVM mengandalkan prinsip margin maksimum dengan kernel non-linier, sedangkan ERT merupakan *ensemble* dari banyak pohon dengan tingkat randomisasi yang tinggi. Hasil evaluasi perbandingan ditampilkan pada Tabel 3.

Tabel 3. Perbandingan kinerja model klasifikasi (*weighted average, split 80:20*)

Model	<i>Precision (Weinghted)</i>	<i>Recall (Weinghted)</i>	<i>F1-Score (Weinghted)</i>	<i>Accuracy</i>	ROC-AUC
DecisionTree	0.8886	0.8879	0.8879	0.8879	0.8876
KNN	0.9089	0.9086	0.9086	0.9086	0.9146
SVM	0.9181	0.9172	0.9173	0.9172	0.9300
ERT	0.9274	0.9269	0.9269	0.9269	0.9429

Tabel 3 menunjukkan bahwa ERT memberikan kinerja terbaik pada seluruh metrik ringkas, dengan nilai *F1-score* sebesar 0.9269 dan ROC-AUC sebesar 0.9429. SVM menempati posisi kedua dengan *F1-score* 0.9173 dan AUC 0.9300, diikuti oleh KNN dengan *F1-score* 0.9086 dan AUC 0.9146. *Decision Tree* sebagai baseline berada di posisi terakhir dengan *F1-score* 0.8879 dan AUC 0.8876. Margin keunggulan ERT terhadap SVM mencapai sekitar 0.95 poin pada *F1-score* dan 1.29 poin pada AUC, sedangkan dibandingkan dengan *Decision Tree* selisihnya lebih signifikan, yaitu sekitar 3.9 poin pada *F1-score* dan 5.53 poin pada AUC.

Secara keseluruhan, ERT konsisten unggul baik pada metrik *threshold-dependent* (*precision*, *recall*, *F1*, *accuracy*) maupun *threshold-independent* (AUC). Hal ini menandakan bahwa ERT tidak hanya berkinerja baik pada ambang klasifikasi tertentu, tetapi juga mampu membedakan kelas dengan stabil di berbagai ambang keputusan. Keunggulan ini didukung oleh mekanisme *ensemble* dan randomisasi tinggi dalam pembentukan pohon, yang mampu menurunkan korelasi antar pohon dan mengurangi varians, sehingga meningkatkan kemampuan generalisasi model. Selain itu, karena berbasis pemisahan *threshold*, ERT tidak memerlukan normalisasi skala fitur, sehingga dapat bekerja baik pada kombinasi data numerik dan kategorikal. Banyaknya pohon juga memungkinkan model menangkap interaksi non-linier yang kompleks antar variabel perilaku sosial, serta menunjukkan ketahanan terhadap outlier dengan kinerja yang sudah optimal meskipun menggunakan parameter standar.

Meski demikian, terdapat beberapa keterbatasan yang perlu diperhatikan. Skor probabilitas yang dihasilkan ERT sering belum terkalibrasi dengan baik, sehingga ke depan perlu diterapkan metode kalibrasi seperti *Platt scaling* atau *isotonic regression* untuk meningkatkan kualitas probabilitas prediksi. Selain itu, analisis *feature importance* berbasis impurity dapat menimbulkan bias pada fitur yang saling berkorelasi, sehingga interpretasi lebih reliabel dapat diperoleh melalui *permutation importance* atau SHAP. Dari sisi optimasi, performa dapat ditingkatkan lebih lanjut dengan tuning parameter seperti *max_depth*, *min_samples_leaf*, dan *max_features* menggunakan pendekatan *nested cross-validation* atau optimasi berbasis *Bayesian*.

Kesimpulan

Penelitian ini berhasil mengembangkan model klasifikasi kepribadian berdasarkan dimensi ekstrasversi menggunakan algoritma *Extremely Randomized Trees* (ERT) dengan memanfaatkan indikator perilaku sosial sebagai fitur. Hasil penelitian menunjukkan bahwa ERT mampu memberikan performa terbaik dibandingkan tiga algoritma pembandingan, yaitu *Decision Tree*, KNN, dan SVM. Model ERT mencapai akurasi sebesar 92.69% dengan nilai *F1-score* 0.9269 dan ROC-AUC 0.9429, melampaui algoritma lain baik pada metrik *threshold-dependent* (*precision*, *recall*, *F1*, *accuracy*) maupun *threshold-independent* (AUC). Keunggulan ERT didukung oleh mekanisme *ensemble* berbasis pohon yang menggabungkan randomisasi tinggi pada pemilihan fitur dan titik *split*, sehingga menghasilkan model yang lebih stabil, tahan terhadap varians, dan mampu menangkap interaksi non-linier antar variabel perilaku sosial. Dengan demikian, ERT terbukti efektif dalam membedakan kepribadian *introvert* dan *extrovert* secara konsisten. Meski demikian, penelitian ini memiliki keterbatasan pada aspek interpretabilitas dan kalibrasi probabilitas. Oleh karena itu, penelitian selanjutnya dapat diarahkan pada peningkatan interpretabilitas menggunakan teknik seperti SHAP atau *permutation importance*, penerapan kalibrasi probabilitas, serta pengujian pada dataset eksternal untuk memperkuat klaim generalisasi. Selain itu, eksplorasi algoritma *ensemble* modern seperti XGBoost, LightGBM, dan CatBoost juga dapat dilakukan untuk memperkaya hasil komparasi.

Referensi

- [1] Z. Zubaidah, F. F. Triana, G. Ananta, R. D. Sadewa, and R. Arkhan, "Konsep Dasar Tes Five Big Personality Traits pada Kepribadian Siswa," in *SENJA KKN (Seminar dalam Jaringan Konseling Kearifan Nusantara)*, 2024, pp. 280–290.
- [2] M. M. Karundeng, S. L. Mandey, and R. N. Taroreh, "Pengaruh Kepribadian Ekstrasversi dan Gaya Kepemimpinan Transformasional Terhadap Kinerja Pegawai di Kecamatan Ranowulu Kota Bitung," *J. EMBA*, vol. 10, no. 1, pp. 1030–1040, 2022.
- [3] N. F. Ainun, N. Nurhikmah, and A. M. Aditya, "Hubungan Antara Kecenderungan Tipe Kepribadian Extraversion dengan Cyberloafing pada Mahasiswa di Kota Makassar," *J. Psikol. Karakter*, vol. 4, no. 2, pp. 420–427, 2024, doi: 10.56326/jpk.v4i2.3725.
- [4] M. P. Pulungan, A. Purnomo, and A. Kurniasih, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 1033–1042, 2024, doi: 10.25126/jtiik.2024117989.
- [5] P. J. Aliffiyah and N. Pratiwi, "Deteksi Tipe Sidik Jari Untuk Mengenali Kepribadian Menggunakan Metode Support Vector Machine (SVM)," *Metik J.*, vol. 9, no. 2, pp. 375–384, 2025, doi: 10.47002/metik.v9i2.1073.
- [6] A. Oktafiqurrahman, K. Kusrini, and A. Nasiri, "Prediksi Kepribadian Berdasarkan Status Sosial Media Facebook Menggunakan Metode Naive Bayes dan KNN," *J. Teknol. Inf. dan Komun.*, vol. 11, no. 2, pp. 30–34, 2023, doi: 10.30646/tikomsin.v11i2.747.

- [7] A. Subtinanda and N. Yuliana, "Kepribadian Ekstrovert dan Introvert dalam Konteks Komunikasi Antarpribadi Mahasiswa Ilmu Komunikasi UNTIRTA," *J. Pendidik. Non Form.*, vol. 1, no. 2, p. 15, 2023, doi: 10.47134/jpn.v1i2.187.
- [8] M. Prasetio, H. Sulistiani, O. Y. Inonu, K. Magda, and B. Santosa, "Klasifikasi Tingkat Risiko Gempa Menggunakan Pola Spasial dan Temporal Berbasis Decision Tree Mugi," *Bull. Comput. Sci. Res.*, vol. 5, no. 5, pp. 1059–1066, 2025, doi: 10.47065/bulletinesr.v5i5.624.
- [9] M. Rizky, P. Soewarno, R. Ardianto, R. Suryani, R. R. Al-hakim, and I. Wahyudi, "Analisis Perbandingan Algoritma KNN dan SVM untuk Prediksi Risiko Kesehatan Ibu," *J. Kolaborasi Ris. Sarj.*, vol. 2, no. 3, pp. 1–8, 2025.
- [10] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bull. Inf. Technol.*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.
- [11] D. Al Mahkya, K. A. Notodiputro, and B. Sartono, "Extra Trees Method for Stock Price Forecasting With Rolling Origin Accuracy Evaluation," *Media Stat.*, vol. 15, no. 1, pp. 36–47, 2022, doi: 10.14710/medstat.15.1.36-47.
- [12] L. Nur Aina, V. R. S. Nastiti, and C. S. K. Aditya, "Implementasi Extra Trees Classifier dengan Optimasi Grid Search CV pada Prediksi Tingkat Adaptasi," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 9, no. 1, pp. 78–88, 2024, [Online]. Available: <https://doi.org/10.26760/mindjournal.v9i1.78-88>
- [13] R. Kapilavayi, "Extrovert vs. Introvert Behavior Data," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data>
- [14] A. Sah, C. Niesa, R. R. Jafar, and M. Muharrom, "Analisis Model Prediksi Penyakit Jantung Menggunakan Adaptive Boosting, Gradient Boosting, dan Extreme Gradient Boosting," *J. Ilm. FIFO*, vol. 17, no. 1, pp. 46–56, 2025, doi: 10.22441/fifo.2025.v17i1.006.
- [15] W. Aprilita, Junadhi, Agustin, and H. Asnal, "Analisis Sentimen Layanan Hotel Menggunakan Algoritma Extra Trees: Studi Kasus pada Ulasan Pelanggan," *Indones. J. Comput. Sci.*, vol. 13, no. 3, pp. 4642–4653, 2024, doi: 10.33022/ijcs.v13i3.4014.
- [16] A. Aminifar, M. Shokri, F. Rabbi, V. K. I. Pun, and Y. Lamo, "Extremely Randomized Trees With Privacy Preservation for Distributed Structured Health Data," *IEEE Access*, vol. 10, pp. 6010–6027, 2022, doi: 10.1109/ACCESS.2022.3141709.
- [17] R. K. Rachmansyah and R. Astriratma, "Implementasi Algoritma Extra Trees Untuk Klasifikasi Cuaca Provinsi DKI Jakarta Dengan Oversampling SMOTE," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2023, pp. 461–472.
- [18] U. Saeed, S. U. Jan, Y.-D. Lee, and I. Koo, "Fault diagnosis based on extremely randomized trees in wireless sensor networks," *Reliab. Eng. Syst. Saf.*, vol. 205, no. 107284, 2021, doi: <https://doi.org/10.1016/j.res.2020.107284>.
- [19] A. Aminifar, F. Rabbi, K. I. Pun, and Y. Lamo, "Privacy preserving distributed extremely randomized trees," in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, in SAC '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1102–1105. doi: 10.1145/3412841.3442110.
- [20] Parjito, I. Ahmad, R. I. Borman, A. D. Alexander, and Y. Jusman, "Combining Extreme Learning Machine and Linear Discriminant Analysis for Optimized Apple Leaf Disease Classification," in *International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, IEEE, 2024, pp. 138–143. doi: 10.1109/ICE3IS62977.2024.10775844.
- [21] R. Rusliyawati, K. Karnadi, A. M. Tanniewa, A. C. Widyawati, Y. Jusman, and R. I. Borman, "Detection of Pepper Leaf Diseases Through Image Analysis Using Radial Basis Function Neural Networks," in *BIO Web of Conferences*, 2024, pp. 1–10. doi: 10.1051/bioconf/202414401005.
- [22] F. S. Nahm, "Receiver operating characteristic curve: overview and practical use for clinicians," *Korean J. Anesthesiol.*, vol. 75, no. 1, pp. 25–36, 2022, doi: 10.4097/kja.21209.