Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari

OPEN ACCESS

https://doi.org/10.35870/jikti.v3i2.2205

Klasifikasi Tipe Kepribadian Mbti Mahasiswa Menggunakan Algoritma *Random Forest*

Nurhidayati [1], Reza Ade Putra [2], Armansyah [3*]

[1,2,3*] Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Kota Meden, Sumatera Utara, Indonesia.

Corresponding Email: nurhidayati0701222122@uinsu.ac.id [1], rezaadeputra@uinsu.ac.id [2], armansyah@uinsu.ac.id [3*]

Histori Artikel:
Dikirim 19 Oktober 2026; *Diterima dalam bentuk revisi* 10 Oktober 2026; *Diterima* 10 November 2026; *Diterbitkan* 27 Januari 2027. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Kepribadian memengaruhi cara seseorang berpikir, berperilaku, dan berinteraksi. Penelitian ini bertujuan menerapkan algoritma *Random Forest* untuk mengklasifikasikan dimensi *Extrovert–Introvert* (E–I) pada *Myers-Briggs Type Indicator* (MBTI), bukan keseluruhan 16 tipe kepribadian. Data diperoleh melalui kuesioner terhadap 250 mahasiswa Universitas Islam Negeri Sumatera Utara (UINSU). Dataset dibagi menjadi 80% data latih (200 data) dan 20% data uji (50 data). Proses klasifikasi meliputi *bootstrap sampling*, pemilihan fitur secara acak, pembentukan pohon keputusan, dan *majority voting*. Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik *accuracy, precision, recall,* dan *F1-score*. Hasil pengujian menunjukkan akurasi sebesar 96%. Kelas 0 memperoleh nilai *precision, recall,* dan *F1-score* sebesar 0,90%, sedangkan kelas 1 memperoleh nilai 0,97%. Hasil ini menunjukkan potensi *Random Forest* dalam membantu klasifikasi hasil kuesioner secara sistematis. Namun, hasil tersebut terbatas pada *dataset* yang digunakan dan perlu diuji lebih lanjut menggunakan data yang lebih beragam.

Kata Kunci: *Random Forest*, MBTI, Kepribadian.

Abstract

Personality influences how an individual thinks, behaves, and interacts. This study aims to apply the Random Forest algorithm to classify the Extrovert–Introvert (E–I) dimension of the Myers-Briggs Type Indicator (MBTI), rather than all 16 personality types. Data were collected via a questionnaire administered to 250 students at the State Islamic University of North Sumatra (UINSU). The dataset was divided into 80% training data (200 samples) and 20% testing data (50 samples). The classification process involved bootstrap sampling, random feature selection, the construction of decision trees, and majority voting. Evaluation was performed using a confusion matrix, with metrics including accuracy, precision, recall, and F1-score. The test results yielded an accuracy of 96%. Class 0 achieved a precision, recall, and F1-score of 0.90, while Class 1 achieved a score of 0.97. These findings demonstrate the potential of the Random Forest algorithm in assisting with the systematic classification of questionnaire results. However, these results are specific to the dataset used and require further validation using more diverse data.

Keyword: Sales; Random Forest, MBTI, Personality.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

1. Pendahuluan

Kepribadian merupakan aspek psikologis yang berperan penting dalam membentuk pola pikir, sikap, dan perilaku individu dalam berbagai situasi. Karakteristik kepribadian yang relatif konsisten menjadi salah satu faktor yang membedakan individu dalam berinteraksi dan merespons lingkungannya. Dalam konteks pendidikan tinggi, pemahaman terhadap kepribadian mahasiswa penting untuk mendukung proses adaptasi terhadap lingkungan akademik dan sosial. Mahasiswa pada fase dewasa awal menghadapi berbagai tuntutan, termasuk penyelesaian tugas perkuliahan, skripsi, dan penyesuaian terhadap lingkungan sosial, yang dapat memengaruhi kondisi psikologis mereka. Penelitian oleh Setyanto *et al.* (2023) terhadap 227 mahasiswa menunjukkan bahwa 26,9% mahasiswa mengalami depresi ringan, 18,5% mengalami depresi sedang, dan 9,3% mengalami depresi berat atau ekstrem. Selain itu, sebanyak 86,8% mahasiswa dilaporkan mengalami tingkat kecemasan yang tinggi. Sementara itu, Azim *et al.* (2025) menemukan adanya variasi distribusi tipe kepribadian mahasiswa, dengan proporsi dimensi Introvert mencapai lebih dari 70% dan dimensi Extrovert lebih dari 80% pada kelompok tertentu. Temuan tersebut menunjukkan pentingnya memahami karakteristik mahasiswa, termasuk kecenderungan kepribadiannya, dalam konteks pendidikan tinggi. Dalam proses pembelajaran, perbedaan kecenderungan kepribadian dapat berkaitan dengan preferensi mahasiswa dalam mengikuti diskusi, menyampaikan pendapat, bekerja dalam kelompok, dan mempelajari materi secara mandiri. Mahasiswa dengan kecenderungan Extrovert umumnya lebih menyukai aktivitas yang melibatkan interaksi langsung, sedangkan mahasiswa dengan kecenderungan Introvert dapat lebih menyukai situasi yang memberikan kesempatan untuk berpikir dan melakukan refleksi secara mandiri. Dalam interaksi sosial, kecenderungan tersebut juga dapat berkaitan dengan pola komunikasi, partisipasi dalam kegiatan kelompok, serta cara mahasiswa menyesuaikan diri dengan lingkungan kampus. Meskipun demikian, kecenderungan kepribadian bukan satu-satunya faktor yang menentukan keberhasilan belajar maupun kualitas interaksi sosial karena keduanya juga dipengaruhi oleh faktor lingkungan, pengalaman, dan karakteristik individu lainnya. Oleh karena itu, pemahaman terhadap kecenderungan kepribadian dapat menjadi salah satu informasi yang dipertimbangkan untuk memahami perbedaan karakteristik mahasiswa dalam lingkungan akademik dan sosial. Salah satu instrumen yang digunakan untuk mengidentifikasi tipe kepribadian adalah Myers-Briggs Type Indicator (MBTI). Instrumen ini mengelompokkan kecenderungan kepribadian berdasarkan empat dimensi, yaitu Extrovert–Introvert (E–I), Sensing–Intuition (S–N), Thinking–Feeling (T–F), dan Judging–Perceiving (Fitri, 2024). Penelitian ini berfokus pada dimensi Extrovert–Introvert (E–I), yang menggambarkan kecenderungan individu dalam mengarahkan perhatian dan berinteraksi dengan lingkungan.

Individu dengan kecenderungan Extrovert umumnya lebih menyukai interaksi sosial, sedangkan individu dengan kecenderungan Introvert cenderung lebih nyaman dalam refleksi internal dan aktivitas individual (Utomo *et al.*, 2025). Identifikasi dimensi tersebut dapat dilakukan melalui kuesioner, tetapi interpretasi hasil secara manual berpotensi membutuhkan waktu dan menimbulkan ketidakkonsistenan dalam proses klasifikasi. Oleh karena itu, diperlukan pendekatan komputasional yang mampu mengolah hasil kuesioner secara sistematis dan terukur. Sejumlah penelitian terdahulu telah menerapkan algoritma pembelajaran mesin untuk mengklasifikasikan tipe kepribadian. Azhari (2025) menunjukkan bahwa kombinasi algoritma genetika dan Random Forest mampu meningkatkan akurasi klasifikasi sebesar 0,2%. Penelitian oleh Erkamim *et al.* (2025) juga membandingkan algoritma Random Forest, Naïve Bayes, dan K-Nearest Neighbor untuk klasifikasi kepribadian Introvert dan Extrovert. Hasil penelitian tersebut menunjukkan bahwa Random Forest menghasilkan akurasi sebesar 90,86%, sedangkan K-Nearest Neighbor memperoleh akurasi tertinggi sebesar 92,59%. Selain itu, penelitian oleh Kasidi dan Rasyidi (2021) memanfaatkan fitur visual dan fitur berbasis konten dari Instagram dalam klasifikasi kepribadian. Penelitian oleh Azim *et al.* (2025) dan Erkamim *et al.* (2025) menggunakan atribut perilaku sosial, sedangkan Maulidah (2023), Harahap dan Muslim (2022), Anggraini *et al.* (2025), serta Haziqah *et al.* (2020) menggunakan data berbasis teks. Berbagai penelitian tersebut memperlihatkan bahwa klasifikasi kepribadian dapat

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

dilakukan dengan memanfaatkan beragam representasi data dan pendekatan komputasional. Perbedaan karakteristik data, atribut yang digunakan, serta metode klasifikasi menjadi aspek yang perlu diperhatikan ketika menerapkan model pembelajaran mesin pada konteks penelitian yang berbeda. Meskipun penelitian terdahulu telah menerapkan algoritma Random Forest untuk klasifikasi kepribadian Introvert dan Extrovert, penelitian ini memiliki perbedaan pada sumber data, representasi fitur, dan konteks penerapannya. Sementara itu, penelitian ini menggunakan data primer hasil kuesioner dari 250 mahasiswa Universitas Islam Negeri Sumatera Utara (UINSU), dengan 15 indikator yang direpresentasikan dalam bentuk biner, yaitu 1 untuk jawaban "Ya" dan 0 untuk jawaban "Tidak". Fokus klasifikasi dibatasi pada dimensi Extrovert–Introvert (E–I) dalam MBTI. Dengan demikian, penelitian ini menitikberatkan pada penerapan dan evaluasi algoritma Random Forest terhadap data kuesioner mahasiswa dengan karakteristik fitur tersebut. Kinerja model dievaluasi menggunakan confusion matrix melalui metrik accuracy, precision, recall, dan F1-score. Perbedaan sumber dan representasi data ini menjadi dasar untuk mengkaji kinerja model dalam konteks penelitian yang dilakukan, sedangkan perbandingan kinerja dengan penelitian terdahulu perlu mempertimbangkan perbedaan dataset, fitur, dan prosedur pengujian yang digunakan. Algoritma Random Forest dipilih karena merupakan metode pembelajaran mesin berbasis ansambel yang menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi. Pembentukan model melibatkan proses bootstrap sampling, yaitu pengambilan sampel data latih dengan pengembalian, serta pemilihan fitur secara acak dalam proses pembentukan pohon.

Prediksi dari setiap pohon kemudian digabungkan melalui mekanisme majority voting untuk menentukan kelas akhir. Penggabungan sejumlah pohon tersebut memungkinkan model menghasilkan prediksi berdasarkan hasil klasifikasi dari beberapa pohon keputusan. Pada penelitian ini, kinerja model dievaluasi menggunakan confusion matrix melalui metrik accuracy, precision, recall, dan F1-score. Penggunaan beberapa metrik tersebut bertujuan memberikan gambaran mengenai ketepatan prediksi secara keseluruhan maupun kemampuan model dalam mengklasifikasikan setiap kelas. Evaluasi ini penting untuk menilai kinerja model pada data pengujian dan mengetahui kesesuaian hasil klasifikasi dengan label yang digunakan dalam penelitian. Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan algoritma Random Forest untuk mengklasifikasikan kecenderungan kepribadian mahasiswa Universitas Islam Negeri Sumatera Utara berdasarkan dimensi Extrovert–Introvert (E–I) pada MBTI menggunakan data primer hasil kuesioner yang terdiri atas 15 indikator. Penelitian ini juga bertujuan mengevaluasi kinerja model menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian diharapkan dapat memberikan gambaran mengenai kinerja algoritma Random Forest dalam mengklasifikasikan data kuesioner mahasiswa berdasarkan dimensi E–I serta menjadi alternatif pendekatan komputasional untuk membantu pengolahan hasil kuesioner secara sistematis.

2. Metode Penelitian

2.1 Data dan Perangkat Penelitian

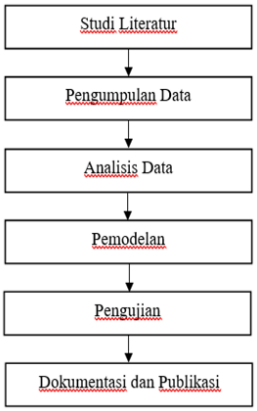
Penelitian ini menggunakan data primer yang diperoleh melalui penyebaran kuesioner kepada 250 mahasiswa Universitas Islam Negeri Sumatera Utara (UINSU). Kuesioner difokuskan pada dimensi *Extrovert–Introvert* (E–I) dalam *Myers-Briggs Type Indicator* (MBTI), bukan untuk mengklasifikasikan keseluruhan 16 tipe kepribadian MBTI. Instrumen penelitian terdiri atas 15 indikator pernyataan yang berkaitan dengan kecenderungan interaksi sosial, komunikasi, aktivitas bersama orang lain, ekspresi diri, dan respons terhadap situasi sosial. Setiap indikator diukur menggunakan pilihan jawaban "Ya" dan "Tidak". Jawaban "Ya" diberi nilai 1, sedangkan jawaban "Tidak" diberi nilai 0. Dengan demikian, setiap responden direpresentasikan oleh 15 fitur numerik yang digunakan sebagai masukan dalam proses klasifikasi.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

Penelitian dilakukan menggunakan laptop Dell Latitude E7440 dengan prosesor Intel Core i3-4030U berkecepatan 1,90 GHz, RAM 4 GB, dan SSD 128 GB. Perangkat lunak yang digunakan meliputi Windows 10 64-bit, Google Chrome, Google Colab, Python, Microsoft Excel 2021, dan Microsoft Word 2021. Google Colab digunakan untuk pengolahan data, pembangunan model *Random Forest*, dan evaluasi hasil klasifikasi. Microsoft Excel digunakan untuk pengelolaan data dan pemeriksaan perhitungan, sedangkan Microsoft Word digunakan untuk penyusunan laporan penelitian.

2.2 Tahapan Penelitian

Penelitian dilaksanakan melalui tahapan studi literatur, pengumpulan data, analisis dan pengujian instrumen, prapemrosesan data, pembagian *dataset*, pemodelan menggunakan algoritma *Random Forest*, pengujian model, serta dokumentasi hasil penelitian. Setiap tahapan dilakukan secara berurutan untuk menjaga keteraturan proses pengolahan data dan memastikan kesesuaian antara prosedur penelitian dengan tujuan klasifikasi.



Gambar 2.1 Kerangka Penelitian

2.3 Studi literatur

Studi literatur dilakukan dengan mengkaji buku, artikel ilmiah, dan penelitian terdahulu yang berkaitan dengan MBTI, dimensi *Extrovert–Introvert*, instrumen kuesioner, pengujian validitas dan reliabilitas, metode klasifikasi, algoritma *Random Forest*, serta teknik evaluasi model. Hasil kajian digunakan sebagai landasan dalam menentukan fokus penelitian, memahami karakteristik indikator yang digunakan, memilih metode klasifikasi, dan menetapkan metrik evaluasi.

2.4 Pengumpulan data

Pengumpulan data dilakukan melalui penyebaran kuesioner kepada mahasiswa UINSU. Kuesioner terdiri atas 15 pernyataan dengan pilihan jawaban "Ya" dan "Tidak". Pernyataan dirancang untuk menggambarkan kecenderungan responden dalam berinteraksi dengan orang lain, bekerja dalam kelompok, memulai percakapan, mengekspresikan diri, serta menghadapi situasi sosial. Data yang terkumpul selanjutnya diperiksa untuk mengetahui kelengkapan jawaban dan kesesuaian format sebelum digunakan dalam proses analisis.

2.5 Analisis data dan Pengukuran Fitur

Analisis data dilakukan untuk mengetahui karakteristik data, jumlah responden, struktur atribut, serta kesesuaian data dengan kebutuhan pemodelan. Fitur penelitian ditetapkan berdasarkan 15 indikator kuesioner yang berkaitan dengan dimensi E–I. Setiap indikator dipertahankan sebagai fitur masukan karena mewakili aspek perilaku yang diukur melalui instrumen penelitian. Dengan demikian, fitur tidak dipilih berdasarkan hasil prediksi model, melainkan ditentukan berdasarkan rancangan indikator kuesioner.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

Pengukuran setiap fitur dilakukan dengan mengonversi jawaban responden menjadi nilai biner. Jawaban "Ya" diberi nilai 1, sedangkan jawaban "Tidak" diberi nilai 0. Hasil konversi tersebut menghasilkan representasi numerik yang seragam untuk seluruh responden. Sebelum pemodelan, data diperiksa untuk memastikan bahwa nilai setiap fitur sesuai dengan skema pengukuran yang telah ditetapkan. Dalam algoritma *Random Forest*, pemilihan fitur secara acak merupakan bagian dari mekanisme pembentukan pohon keputusan. Proses tersebut berbeda dari penetapan 15 fitur penelitian. Seluruh fitur yang telah ditentukan dari kuesioner menjadi kumpulan fitur masukan, sedangkan algoritma memilih subset fitur secara acak pada proses pembentukan pohon sesuai dengan konfigurasi model. Oleh karena itu, penetapan fitur berdasarkan indikator kuesioner dan pemilihan fitur secara acak oleh algoritma merupakan dua proses yang berbeda. Label kelas ditentukan menggunakan aturan penskoran kuesioner yang ditetapkan dalam penelitian. Jika total skor minimal 8 digunakan untuk menetapkan kelas *Extrovert* dan total skor di bawah 8 untuk kelas *Introvert*, aturan tersebut diterapkan secara konsisten pada seluruh data. Aturan penskoran dan proses pembentukan label perlu dijelaskan secara transparan karena berpengaruh terhadap interpretasi hasil evaluasi model.

2.6 Uji Validitas Instrumen

Uji validitas dilakukan untuk mengevaluasi sejauh mana setiap butir pernyataan pada kuesioner berkaitan dengan konstruk yang hendak diukur. Pengujian validitas butir dapat dilakukan menggunakan korelasi *Pearson Product Moment* antara skor setiap butir dan skor total instrumen. Dalam pengujian ini, skor butir dinyatakan sebagai variabel X, skor total sebagai variabel Y, dan jumlah responden sebagai n. Rumus korelasi Pearson dinyatakan sebagai berikut:

$$r_{hitung} = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{n(\sum X^2 - (\sum X)^2\,(n(\sum X^2 - (\sum X)^2)}}$$

Keterangan :

r_{hitung} = Koefisien korelasi
X = Variabel bebas
Y = Variabel terikat
n = Banyak responden

Nilai korelasi yang diperoleh kemudian dibandingkan dengan nilai (r) tabel pada taraf signifikansi 5% dengan derajat kebebasan (df=n-2). Butir dinyatakan memenuhi kriteria validitas apabila nilai (r) hitung lebih besar daripada (r) tabel sesuai dengan kriteria yang digunakan dalam penelitian (Widodo *et al.*, 2023). Hasil pengujian setiap butir perlu disajikan dalam tabel yang memuat nomor indikator, nilai korelasi, nilai (r) tabel, dan keputusan validitas. Karena jawaban kuesioner berbentuk biner, metode korelasi dan cara menghitung skor total perlu diterapkan secara konsisten. Apabila korelasi Pearson digunakan, skor total sebaiknya dihitung sesuai dengan prosedur instrumen dan tidak memasukkan skor butir yang sedang diuji apabila penelitian menggunakan korelasi butir-total terkoreksi. Pemilihan prosedur tersebut perlu disesuaikan dengan perhitungan validitas yang benar-benar dilakukan.

2.7 Uji Reliabilitas Instrumen

Uji reliabilitas dilakukan untuk mengetahui konsistensi internal antarbutir dalam instrumen kuesioner. Pengujian dilakukan menggunakan koefisien *Cronbach's Alpha*. Rumus perhitungan reliabilitas dinyatakan sebagai berikut :

$$r_i = \frac{k}{k-1}\left\{1 - \frac{\sum s_i^2}{s_t^2}\right\}$$

Keterangan :

r_i = Koefisien reliabilitas *Cronbach Alpha*

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

k = Jumlah item soal
$\sum s_i^2$ = Jumlah varians skor tiap item
s_t^2 = Varians total

Pengujian dilakukan menggunakan skor jawaban dari responden yang sama dengan data yang digunakan dalam pengujian instrumen. Berdasarkan kriteria yang dirujuk dalam penelitian, instrumen dinyatakan memenuhi kriteria reliabilitas apabila nilai *Cronbach's Alpha* lebih besar dari 0,60 dan lebih kecil dari 0,90 (Yusup, 2021). Nilai koefisien yang diperoleh harus dilaporkan berdasarkan hasil perhitungan aktual, disertai kriteria interpretasi yang digunakan. Hasil uji validitas dan reliabilitas digunakan untuk memberikan informasi mengenai kualitas pengukuran instrumen. Kedua pengujian tersebut memiliki fungsi berbeda: validitas butir menilai keterkaitan butir dengan konstruk atau skor yang dijadikan pembanding, sedangkan reliabilitas menilai konsistensi internal antarbutir. Oleh karena itu, hasil pengujian perlu disajikan secara terpisah dan tidak digunakan sebagai pengganti evaluasi kinerja model klasifikasi.

2.8 Prapemrosesan Data dan Pembagian *Dataset*

Prapemrosesan dilakukan untuk memastikan data berada dalam kondisi yang sesuai sebelum digunakan dalam pemodelan. Tahapan ini meliputi pemeriksaan kelengkapan data, pemeriksaan duplikasi, penyesuaian format, konversi jawaban menjadi nilai numerik, serta penetapan fitur dan label kelas. Jumlah data sebelum dan sesudah prapemrosesan dicatat untuk memastikan perubahan jumlah observasi dapat ditelusuri. Setelah prapemrosesan, *dataset* dibagi menjadi data latih dan data uji dengan rasio 80:20. Dari total 250 responden, sebanyak 200 data digunakan sebagai data latih dan 50 data digunakan sebagai data uji. Pembagian dilakukan menggunakan fungsi train_test_split dengan *parameter test_size=0.20, random_state=10, dan stratify=y. Parameter stratify=y* digunakan untuk mempertahankan proporsi kelas pada data latih dan data uji sedekat mungkin dengan distribusi kelas pada dataset awal. Data latih digunakan untuk membangun model *Random Forest*, sedangkan data uji digunakan untuk mengevaluasi kinerja model terhadap data yang tidak digunakan dalam pelatihan. Penggunaan *random_state=10* ditujukan untuk membuat pembagian data dapat direproduksi selama dataset dan konfigurasi eksperimen tidak berubah.

2.9 Pemodelan Menggunakan Random Forest

Algoritma *Random Forest* merupakan metode klasifikasi berbasis ansambel yang menggabungkan prediksi sejumlah pohon keputusan. Pembentukan model melibatkan pengambilan sampel data latih dan pembentukan pohon berdasarkan fitur yang tersedia. Prediksi dari setiap pohon selanjutnya digabungkan menggunakan mekanisme *majority voting* untuk menentukan kelas akhir. Penelitian ini menggunakan 100 pohon keputusan dengan kriteria pemisahan entropy, parameter max_features=None, dan random_state=10. Parameter max_features=None berarti seluruh fitur yang tersedia dipertimbangkan pada setiap pemisahan simpul. Dengan demikian, konfigurasi ini tidak menggunakan pembatasan jumlah fitur kandidat pada setiap simpul. Hal tersebut perlu dibedakan dari mekanisme pemilihan subset fitur secara acak yang lazim digunakan dalam konfigurasi *Random Forest*. Setiap pohon menghasilkan prediksi kelas berdasarkan pola pada data latih. Prediksi dari 100 pohon kemudian digabungkan melalui *majority voting*, yaitu penetapan kelas berdasarkan jumlah suara terbanyak dari seluruh pohon. Kelas dengan suara terbanyak menjadi hasil klasifikasi akhir model.

2.10 Pengujian dan Evaluasi Model

Pengujian dilakukan menggunakan 50 data uji dengan membandingkan hasil prediksi model terhadap label kelas sebenarnya. Kinerja klasifikasi diukur menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion matrix* digunakan untuk mengetahui jumlah prediksi benar dan salah pada setiap kelas, yang terdiri atas *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN).

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

Accuracy mengukur proporsi seluruh prediksi yang benar terhadap jumlah data uji. Perhitungan *accuracy* dinyatakan dengan persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + RN + FP}$$

Precision mengukur proporsi prediksi positif yang benar dari seluruh data yang diprediksi sebagai kelas positif. Nilai *precision* dihitung menggunakan persamaan berikut:

$$Precision = \frac{TP}{TP + FP}$$

Recall mengukur proporsi data positif aktual yang berhasil dikenali dengan benar oleh model. Perhitungan *recall* dinyatakan sebagai berikut:

$$Recall = \frac{TP}{TP + FN}$$

F1-score merupakan rata-rata harmonis antara *precision* dan *recall*. Metrik ini membantu memberikan gambaran keseimbangan antara ketepatan prediksi kelas positif dan kemampuan model dalam mengenali kelas positif yang sebenarnya. Persamaan *F1-score* adalah sebagai berikut:

$$F1\ Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

Keterangan :
True Positive (*TP*) : Data positif yang diprediksi benar sebagai *positive*
True Negative (*TN*) : Data negatif yang diprediksi benar sebagai *negative*
False Positive (*FP*) : Data negatif yang salah diprediksi sebagai *positive*
False Negative (*FN*) : Data positif yang salah diprediksi sebagai *negative*

Dalam penerapannya, kelas positif harus ditentukan secara konsisten ketika menghitung *precision*, *recall*, dan *F1-score*. Karena penelitian melibatkan kelas *Extrovert* dan *Introvert*, nilai metrik untuk setiap kelas sebaiknya dilaporkan secara terpisah beserta jumlah data aktual pada masing-masing kelas. Pelaporan tersebut diperlukan untuk mengetahui apakah kinerja model berbeda antara kedua kelas, terutama jika distribusi kelas pada data uji tidak seimbang. Hasil evaluasi digunakan untuk menggambarkan kinerja model *Random Forest* dalam memprediksi label kelas pada data uji. Interpretasi hasil perlu mempertimbangkan karakteristik data, distribusi kelas, dan aturan pembentukan label. Jika label berasal dari skor fitur yang sama dengan yang digunakan sebagai masukan, nilai evaluasi harus ditafsirkan sebagai kinerja model dalam mereproduksi aturan pelabelan tersebut, bukan sebagai bukti tunggal bahwa model mampu mengidentifikasi kepribadian berdasarkan standar independen.

2.11 Dokumentasi dan Publikasi

Tahap akhir penelitian dilakukan dengan mendokumentasikan data yang digunakan, prosedur prapemrosesan, hasil pengujian instrumen, konfigurasi model, proses klasifikasi, dan evaluasi kinerja algoritma *Random Forest*. Dokumentasi hasil mencakup hasil uji validitas dan reliabilitas, pembagian data latih dan data uji, konfigurasi model, *confusion matrix*, nilai *accuracy*, *precision*, *recall*, dan *F1-score*, serta pembahasan mengenai kemampuan model dalam memprediksi kelas berdasarkan dimensi E–I. Seluruh hasil dilaporkan berdasarkan keluaran pengolahan data dan eksperimen yang benar-benar dilakukan agar proses penelitian dapat ditelusuri dan direproduksi.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Hasil Pengolahan Data

Penelitian ini menggunakan 250 data kuesioner mahasiswa Universitas Islam Negeri Sumatera Utara (UINSU) yang berfokus pada dimensi *Extrovert–Introvert* (E–I) pada *Myers-Briggs Type Indicator* (MBTI). Instrumen penelitian terdiri atas 15 indikator pernyataan yang merepresentasikan kecenderungan individu dalam berinteraksi dengan orang lain, berkomunikasi, melakukan aktivitas bersama, mengekspresikan diri, dan menghadapi situasi sosial. Setiap jawaban dikonversi menjadi nilai biner, yaitu 1 untuk jawaban "Ya" dan 0 untuk jawaban "Tidak". Hasil konversi tersebut digunakan untuk membentuk fitur X1–X15 sebagai masukan dalam proses klasifikasi. Berdasarkan pemeriksaan awal, dataset tidak memiliki nilai kosong maupun data duplikat sehingga seluruh 250 data dapat diproses pada tahapan berikutnya.

3.1.2 Hasil Uji Validitas Instrumen

Uji Validitas dilakukan untuk mengetahui kemampuan setiap item pernyataan dalam mengukur variabel penelitian. Pengujian Validitas dilakukan menggunakan korelasi *Pearson Product Moment* dengan membandingkan nilai jika r_{hitung} dan r_{tabel}. Suatu item pernyataan dinyatakan *Valid* apabila jika $r_{hitung} > r_{tabel}$. Hasil uji Validitas instrumen penelitian dapat dilihat pada Tabel 3.1.

Tabel 3. 1Hasil Uji Validitas.

No	r_{hitung}	r_{tabel}	Keterangan
1	0.6691	0.1241	Valid
2	0.5230	0.1241	Valid
3	0.5062	0.1241	Valid
4	0.4449	0.1241	Valid
5	0.5178	0.1241	Valid
6	0.4614	0.1241	Valid
7	0.5354	0.1241	Valid
8	0.2472	0.1241	Valid
9	0.3617	0.1241	Valid
10	0.5192	0.1241	Valid
11	0.4607	0.1241	Valid
12	0.5221	0.1241	Valid
13	0.5056	0.1241	Valid
14	0.6380	0.1241	Valid
15	0.5285	0.1241	Valid

Berdasarkan Tabel 3.1, nilai korelasi seluruh butir berada pada rentang 0,2472–0,6691 dan lebih besar dari pada nilai r_{tabel} sebesar 0,1241. Dengan demikian, seluruh 15 butir memenuhi kriteria validitas berdasarkan metode dan ambang yang digunakan dalam penelitian ini. Nilai korelasi tertinggi terdapat pada butir pertama sebesar 0,6691, sedangkan nilai terendah terdapat pada butir kedelapan sebesar 0,2472. Meskipun memiliki nilai korelasi paling rendah, butir kedelapan tetap memenuhi kriteria perbandingan yang digunakan. Hasil tersebut menunjukkan bahwa setiap butir memiliki keterkaitan dengan skor total instrumen berdasarkan pengujian yang dilakukan. Namun, hasil korelasi butir-total perlu ditafsirkan sesuai dengan prosedur perhitungan yang digunakan dan tidak secara otomatis membuktikan seluruh aspek validitas konstruk instrumen. Oleh karena itu, hasil uji validitas ini menjadi salah satu dasar untuk menjelaskan kualitas pengukuran indikator sebelum data digunakan dalam pemodelan klasifikasi.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

3.1.3 Hasil Uji Reliabilitas Instrumen

Uji reliabilitas dilakukan untuk mengetahui konsistensi internal dari 15 butir pernyataan yang digunakan dalam kuesioner. Pengujian dilakukan menggunakan metode *Cronbach's Alpha*. Dalam penelitian ini, instrumen dinyatakan memenuhi kriteria reliabilitas apabila nilai *Cronbach's Alpha* lebih besar dari 0,60. Hasil pengujian reliabilitas disajikan pada Tabel 3.2.

Tabel 3.2 Hasil Uji Reliabilitas.

Jumlah Item	Varians Item	Jumlah Varians Item	Varians Total	Cronbach's Alpha	Keterangan
15	0.222667 0.224 0.227807 0.183133 0.217012 0.178892 0.224 0.145606 0.140337 0.23351 0.22657 0.245799 0.209221 0.251004 0.176723	11.67846	3.106281	0.786446	Reliabel

Berdasarkan Tabel 3.2, pengujian terhadap 15 butir pernyataan menghasilkan nilai *Cronbach's Alpha* sebesar 0,786446. Nilai tersebut lebih besar dari pada ambang 0,60 yang digunakan dalam penelitian sehingga instrumen memenuhi kriteria reliabilitas yang ditetapkan. Hasil ini menunjukkan adanya konsistensi internal antar butir berdasarkan pengujian yang dilakukan terhadap data responden. Hasil uji validitas dan reliabilitas memberikan informasi mengenai kualitas pengukuran instrumen sebelum digunakan dalam proses klasifikasi. Uji validitas digunakan untuk mengevaluasi keterkaitan setiap butir dengan skor total, sedangkan uji reliabilitas digunakan untuk menilai konsistensi internal instrumen. Meskipun demikian, kedua hasil tersebut tidak secara langsung menunjukkan tingkat akurasi algoritma *Random Forest*, karena kinerja model dinilai secara terpisah menggunakan data uji dan metrik evaluasi klasifikasi.

3.1.4 Representasi Data dan Pemilihan Fitur

Representasi data dilakukan dengan mengonversi jawaban kuesioner dari 250 responden menjadi nilai numerik. Jawaban "Ya" diberi nilai 1, sedangkan jawaban "Tidak" diberi nilai 0. Proses ini menghasilkan 15 fitur, yaitu X1–X15, yang digunakan untuk menggambarkan jawaban responden terhadap indikator kuesioner. Pemilihan fitur didasarkan pada rancangan instrumen penelitian dan keterkaitannya dengan dimensi E–I, bukan pada hasil prediksi model. Setiap fitur merepresentasikan satu pernyataan yang mengukur aspek kecenderungan perilaku tertentu. Rincian indikator dan cara pengukurannya disajikan pada Tabel 3.3.

Tabel 3.3 Proses Representasi Data.

Kode Atribut	Indikator Yang Diukur	Pengukuran
$X1$	Saya merasa berenergi ketika berada di sekitar banyak orang	Ya = 1 ; Tidak = 0
$X2$	Saya lebih suka bekerja dalam tim daripada sendiri	Ya = 1 ; Tidak = 0
$X3$	Saya mudah memulai percakapan dengan orang yang baru saya kenal	Ya = 1 ; Tidak = 0
$X4$	Saya sering berbagi pikiran dan perasaan dengan orang lain	Ya = 1 ; Tidak = 0

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

$X5$	Saya merasa nyaman menjadi pusat perhatian	Ya = 1 ; Tidak = 0
$X6$	Saya lebih suka menghabiskan waktu luang dengan bertemu teman-teman	Ya = 1 ; Tidak = 0
$X7$	Saya cepat beradaptasi dengan lingkungan sosial yang baru	Ya = 1 ; Tidak = 0
$X8$	Saya sering berpikir sambil berbicara (*thinking out loud*)	Ya = 1 ; Tidak = 0
$X9$	Saya lebih suka memproses informasi dengan mendiskusikannya bersama orang lain	Ya = 1 ; Tidak = 0
$X10$	Saya merasa terstimulasi oleh aktivitas dan keramaian	Ya = 1 ; Tidak = 0
$X11$	Saya mudah mengekspresikan emosi dan antusiasme saya	Ya = 1 ; Tidak = 0
$X12$	Saya lebih produktif ketika ada orang lain di sekitar	Ya = 1 ; Tidak = 0
$X13$	Saya sering mencari interaksi sosial untuk mengatasi stress	Ya = 1 ; Tidak = 0
$X14$	Saya merasa lebih percaya diri dalam situasi sosial yang ramai	Ya = 1 ; Tidak = 0
$X15$	Saya lebih suka berkomunikasi secara verbal daripada tertulis	Ya = 1 ; Tidak = 0

Pengukuran pada Tabel 3.3 menggunakan skala biner untuk menyamakan format seluruh jawaban sehingga dapat diproses oleh algoritma klasifikasi. Nilai 1 menunjukkan bahwa responden memilih jawaban "Ya" pada pernyataan yang bersangkutan, sedangkan nilai 0 menunjukkan jawaban "Tidak". Nilai tersebut merupakan pengodean jawaban responden dan tidak secara terpisah menunjukkan tingkat kepribadian yang telah tervalidasi. Penetapan 15 fitur dilakukan berdasarkan indikator yang telah ditentukan dalam instrumen. Seluruh fitur tersebut digunakan sebagai variabel masukan dalam model *Random Forest*. Sementara itu, label kelas dibentuk berdasarkan aturan penskoran kuesioner, yaitu kelas *Introvert* dengan label 0 untuk skor total kurang dari 8 dan kelas *Extrovert* dengan label 1 untuk skor total minimal 8. Dengan demikian, fitur menggambarkan jawaban pada setiap indikator, sedangkan label menunjukkan kategori yang diperoleh berdasarkan aturan penskoran tersebut. Perlu diperhatikan bahwa label kelas dalam penelitian ini berasal dari total skor 15 indikator yang sama dengan fitur masukan. Oleh karena itu, terdapat hubungan langsung antara fitur dan label. Kondisi ini perlu dipertimbangkan dalam menafsirkan hasil klasifikasi, karena model dapat mempelajari pola aturan penskoran yang digunakan untuk membentuk label. Hasil akurasi yang tinggi dalam kondisi tersebut tidak dengan sendirinya membuktikan kemampuan model mengidentifikasi kepribadian berdasarkan ukuran atau penilaian independen.

3.1.5 Pembagian *Dataset*

Setelah proses representasi data, dataset dibagi menjadi data latih dan data uji menggunakan metode *hold-out* dengan proporsi 80:20. Dari total 250 data responden, sebanyak 200 data digunakan sebagai data latih dan 50 data sebagai data uji. Pembagian dilakukan menggunakan fungsi train_test_split() dengan random_state=10 sesuai dengan konfigurasi eksperimen yang digunakan. Data latih digunakan untuk membangun model *Random Forest*, sedangkan data uji digunakan untuk membandingkan hasil prediksi model dengan label kelas pada data yang tidak digunakan dalam proses pelatihan. Pemisahan ini bertujuan memberikan penilaian terhadap hasil klasifikasi pada data uji. Namun, karena evaluasi dilakukan menggunakan satu pembagian data, hasilnya perlu ditafsirkan dengan mempertimbangkan komposisi sampel pengujian.

3.1.6 Penerapan Algoritma Random Forest

Dataset yang digunakan dalam pemodelan terdiri atas 250 data responden dengan 15 fitur kuesioner dan satu label kelas. Setiap fitur merepresentasikan jawaban biner pada indikator X1–X15, sedangkan label menunjukkan kategori *Introvert* (0) atau *Extrovert* (1) berdasarkan aturan penskoran yang digunakan. Tampilan dataset hasil representasi data disajikan pada Gambar 3.1.

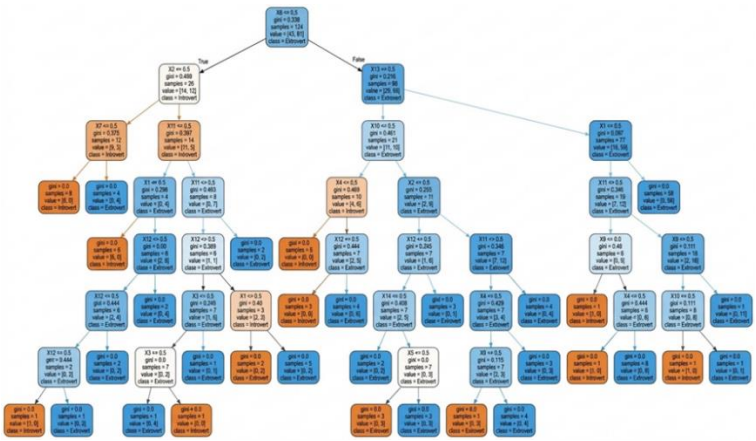
Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

	Cap waktu	Nama Pengguna	Nama	Program Studi	Jenis Kelamin	Semester	Usia	X1	X2	X3	...	X8	X9	X10	X11	X12	X13	X14	X15	Total	Hasil
0	2026/01/23 9:27:11 PM GMT+7	juwitasari1001104@gmail.com	Juwita Sari	Ilmu Komputer	Perempuan	7	22 Tahun	0	0	0	...	1	1	0	1	0	0	0	1	5	0
1	2026/01/23 9:31:04 PM GMT+7	intanwidyasaputrinst@gmail.com	Intan Widya Saputri Nst	Ilmu Komputer	Perempuan	7	23 Tahun	0	1	0	...	1	1	0	0	0	0	1	0	5	0
2	2026/01/23 10:05:15 PM GMT+7	adindatarisyshhsb@gmail.com	Adinda Tarisya Hsb	Ilmu Komputer	Perempuan	7	22 Tahun	0	0	0	...	1	1	1	0	0	0	0	1	4	0
3	2026/01/23 10:07:35 PM GMT+7	gimagim1224@gmail.com	Agim	Ilmu Komputer	Laki-laki	7	22 Tahun	0	0	1	...	0	0	1	0	0	1	0	1	5	0
4	2026/01/23 10:07:49 PM GMT+7	hasti07012222113@ulnsu.su.id	Hasti Fadillah	Ilmu Komputer	Perempuan	7	22 Tahun	1	1	0	...	1	1	0	0	0	0	0	1	7	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
245	2026/05/03 4:30:14 PM GMT+7	amnsep491@gmail.com	Sev	Ilmu Komputer	Perempuan	8	22 Tahun	1	1	1	...	0	1	1	1	1	1	0	11	11	1
248	2026/05/04 7:14:59 PM GMT+7	nisitarahmi.k25@gmail.com	Nisita Rahmi Karunanda	Ilmu Komputer	Perempuan	6	21 Tahun	1	0	1	...	0	0	1	1	1	0	0	1	7	0
247	2026/05/04 7:17:38 AM GMT+7	zhritaull@gmail.com	ZA	Manajemen	Perempuan	6	21 Tahun	1	1	0	...	0	1	0	0	1	0	1	1	8	1
248	2026/05/04 7:20:18 AM GMT+7	zahraa.growthup@gmail.com	mff	sosiologi	Perempuan	8	21 Tahun	1	1	1	...	1	0	1	1	0	1	1	1	11	1
249	2026/05/04 7:31:50 AM GMT+7	Fatmzahra901@gmail.com	Zara	SI	Perempuan	8	20 Tahun	0	0	0	...	0	1	1	1	0	1	0	1	7	0

250 rows × 24 columns

Gambar 3.1 Tampilan *Dataset* Hasil Representasi Data.

Model *Random Forest* dibangun melalui proses *bootstrap sampling*, pemilihan fitur secara acak, pembentukan pohon keputusan, dan penggabungan prediksi menggunakan *majority voting*. Pada tahap *bootstrap sampling*, data latih diambil dengan pengembalian sehingga satu observasi dapat terpilih lebih dari satu kali dalam sampel pembentukan suatu pohon. Proses tersebut menghasilkan variasi sampel pelatihan yang digunakan oleh pohon-pohon dalam model. Tahap pemilihan fitur secara acak dilakukan terhadap 15 fitur yang telah ditetapkan dari indikator kuesioner. Berdasarkan konfigurasi yang digunakan dalam penelitian, empat fitur dipertimbangkan sebagai kandidat pada proses pemisahan simpul. Pemilihan kandidat fitur secara acak bertujuan menghasilkan keragaman dalam pembentukan pohon keputusan. Perlu dibedakan bahwa ke-15 fitur tetap menjadi kumpulan fitur masukan model, sedangkan empat fitur merupakan kandidat yang dipertimbangkan pada proses pemisahan sesuai dengan konfigurasi tersebut. Selanjutnya, setiap pohon keputusan mempelajari pola hubungan antara fitur dan label kelas dari sampel pelatihan yang digunakan. Prediksi dari seluruh pohon kemudian digabungkan melalui mekanisme *majority voting*. Kelas yang memperoleh jumlah suara terbanyak ditetapkan sebagai prediksi akhir model. Visualisasi salah satu pohon keputusan dari model disajikan pada Gambar 3.2.



Gambar 3.2 Visualisasi Salah Satu Pohon Keputusan dalam *Random Forest* .

Visualisasi pohon keputusan memperlihatkan proses pemisahan data berdasarkan fitur tertentu hingga menghasilkan prediksi kelas pada simpul akhir. Pohon yang ditampilkan merupakan salah satu pohon dalam model *Random Forest*, sehingga visualisasi tersebut tidak mewakili keseluruhan mekanisme pengambilan keputusan model. Hasil klasifikasi akhir tetap ditentukan melalui penggabungan prediksi seluruh pohon menggunakan *majority voting*. Penerapan *Random Forest* memungkinkan proses klasifikasi dilakukan berdasarkan gabungan prediksi dari sejumlah pohon keputusan. Namun, kemampuan model dalam memprediksi label kelas tetap perlu dinilai menggunakan data uji yang terpisah dari proses pelatihan. Evaluasi dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

3.1.7 Hasil Evaluasi Model

Hasil pengujian terhadap 50 data uji menunjukkan bahwa model *Random Forest* memperoleh akurasi sebesar 96%. Berdasarkan *classification report*, kelas 0 (*Introvert*) memperoleh nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,90 dengan *support* sebanyak 10 data. Sementara itu, kelas 1 (*Extrovert*) memperoleh nilai *precision* sebesar 0,97, *recall* sebesar 0,97, dan *F1-score* sebesar 0,97 dengan *support* sebanyak 40 data. Nilai *macro average* sebesar 0,94 dan *weighted average* sebesar 0,96 melengkapi hasil evaluasi model. Hasil tersebut ditampilkan pada Gambar 3.3.

```
============================================================
                    HASIL EVALUASI MODEL
============================================================

Akurasi Model: 96.00%

Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.90      0.90        10
           1       0.97      0.97      0.97        40

    accuracy                           0.96        50
   macro avg       0.94      0.94      0.94        50
weighted avg       0.96      0.96      0.96        50
```

Gambar 3.3 Hasil Evaluasi Model *Random Forest*.

Untuk memperjelas hasil klasifikasi, evaluasi dilakukan menggunakan *confusion matrix* sebagaimana ditampilkan pada Tabel 3.4.

Tabel 3.4 Hasil *Confusion Matrix*.

Kelas Aktual	Prediksi *Introvert*	Prediksi *Extrovert*	Total
Introvert	9	1	10
Extrovert	1	39	40
Total	10	40	50

Berdasarkan Tabel 3.4, model berhasil mengklasifikasikan 9 data *Introvert* dan 39 data *Extrovert* dengan benar, sedangkan 2 data lainnya mengalami kesalahan klasifikasi. Hasil ini menghasilkan akurasi sebesar 96% dari 50 data uji. Distribusi kelas yang terdiri atas 10 data *Introvert* dan 40 data *Extrovert* menunjukkan adanya ketidakseimbangan jumlah data, sehingga evaluasi juga mempertimbangkan nilai *precision*, *recall*, dan *F1-score* pada setiap kelas. Meskipun memperoleh akurasi tinggi, hasil tersebut masih terbatas pada pembagian data uji yang digunakan. Selain itu, label kelas dibentuk dari total skor 15 indikator yang juga digunakan sebagai fitur masukan. Oleh karena itu, model dapat mempelajari kembali aturan penskoran tersebut, sehingga akurasi yang diperoleh belum membuktikan kemampuan identifikasi kepribadian secara independen.

3.2 Pembahasan

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest dapat digunakan secara efektif untuk mengklasifikasikan kecenderungan kepribadian mahasiswa Universitas Islam Negeri Sumatera Utara berdasarkan dimensi Extrovert–Introvert (E–I) dalam MBTI. Dengan menggunakan data primer dari kuesioner yang mencakup 15 indikator, model ini berhasil memberikan akurasi yang memuaskan, yang sejalan dengan temuan Azhari (2025) yang menunjukkan bahwa penggunaan algoritma pembelajaran mesin, khususnya kombinasi algoritma genetika dan Random Forest, mampu meningkatkan akurasi klasifikasi kepribadian. Penelitian ini juga sejalan dengan studi Erkamim *et al.* (2025), yang membandingkan beberapa algoritma klasifikasi dan menemukan bahwa Random Forest memberikan hasil yang kompetitif, dengan akurasi mencapai 90,86%. Lebih lanjut, hasil penelitian ini memperkuat pemahaman bahwa karakteristik kepribadian mahasiswa, terutama kecenderungan Extrovert dan Introvert, memiliki pengaruh signifikan terhadap cara mereka berinteraksi dalam lingkungan akademik. Hal ini konsisten dengan temuan Azim *et al.* (2025) yang menunjukkan bahwa proporsi mahasiswa dengan dimensi Introvert dan Extrovert dapat memengaruhi partisipasi mereka dalam kegiatan sosial dan akademik. Dalam konteks pembelajaran, mahasiswa dengan kecenderungan Extrovert cenderung lebih aktif dalam

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

diskusi kelompok dan interaksi sosial, sedangkan mahasiswa Introvert mungkin lebih memilih metode belajar yang lebih mandiri dan reflektif. Penelitian oleh Maulidah (2023) juga menyoroti pentingnya memahami perbedaan ini untuk mendukung strategi pengajaran yang lebih efektif dan inklusif. Selain itu, penelitian ini menegaskan perlunya pendekatan komputasional dalam pengolahan hasil kuesioner untuk mengurangi ketidakkonsistenan dan waktu yang dibutuhkan dalam interpretasi manual. Hal ini sejalan dengan gagasan yang diusulkan oleh Fitri (2024), yang menekankan bahwa penerapan teknologi dalam pengolahan data psikologis dapat meningkatkan efisiensi dan akurasi dalam identifikasi kepribadian. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada pemahaman kepribadian mahasiswa, tetapi juga membuka jalan bagi penelitian lebih lanjut yang mengintegrasikan teknologi dalam bidang psikologi dan pendidikan.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian, algoritma Random Forest menunjukkan kinerja yang baik dalam mengklasifikasikan dimensi kepribadian Introvert dan Extrovert mahasiswa berdasarkan 15 indikator kuesioner. Pengujian terhadap 50 data testing menghasilkan akurasi sebesar 96%, dengan nilai precision, recall, dan F1-score masing-masing sebesar 0,90 untuk kelas Introvert dan 0,97 untuk kelas Extrovert. Hasil ini menunjukkan bahwa model mampu melakukan klasifikasi dengan baik pada dataset pengujian yang digunakan, sejalan dengan temuan sebelumnya oleh Erkamim *et al.* (2025), yang juga melaporkan akurasi tinggi dalam klasifikasi kepribadian menggunakan model pembelajaran mesin. Namun, perlu dicatat bahwa nilai akurasi tersebut belum dapat digeneralisasikan secara langsung pada populasi mahasiswa yang lebih luas karena pengujian masih terbatas pada dataset dan pembagian data yang digunakan dalam penelitian ini. Model yang dikembangkan berpotensi menjadi alat bantu awal dalam mengolah hasil kuesioner dan mengenali kecenderungan kepribadian mahasiswa. Hal ini dapat mendukung pengembangan potensi diri, pendampingan akademik, serta layanan bimbingan dan konseling di perguruan tinggi. Temuan ini konsisten dengan penelitian oleh Azim *et al.* (2025) yang menekankan pentingnya pemahaman karakteristik kepribadian untuk meningkatkan pengalaman belajar mahasiswa. Meskipun demikian, penelitian ini memiliki keterbatasan, yaitu jumlah responden yang terbatas pada 250 mahasiswa, penggunaan satu kali pembagian data latih dan data uji, serta penggunaan indikator kuesioner yang sama untuk membentuk fitur masukan dan menentukan label kelas. Kondisi terakhir ini berpotensi menyebabkan model mempelajari pola penentuan label berdasarkan skor kuesioner, sehingga nilai akurasi yang diperoleh belum sepenuhnya menunjukkan kemampuan model dalam mengidentifikasi kepribadian berdasarkan pengukuran yang independen. Untuk penelitian selanjutnya, disarankan untuk menggunakan dataset yang lebih besar dan beragam, baik dari segi program studi, semester, maupun perguruan tinggi, agar kemampuan generalisasi model dapat dievaluasi secara lebih menyeluruh. Pengujian juga dapat dilakukan menggunakan k-fold cross-validation untuk memperoleh evaluasi yang lebih stabil, seperti yang disarankan oleh Fitri (2024), serta membandingkan kinerja Random Forest dengan algoritma lain, seperti Decision Tree, Support Vector Machine (SVM), dan XGBoost. Selain itu, penelitian selanjutnya perlu memastikan bahwa fitur masukan tidak secara langsung menentukan label target, misalnya dengan menggunakan label hasil penilaian instrumen yang independen dari fitur prediktor. Penerapan hyperparameter tuning dan evaluasi menggunakan metrik accuracy, precision, recall, dan F1-score juga dapat dipertimbangkan untuk memperoleh model yang lebih optimal dan hasil evaluasi yang lebih objektif. Dengan langkah-langkah ini, diharapkan penelitian mendatang dapat memberikan kontribusi yang lebih signifikan dalam pengembangan alat bantu untuk memahami kepribadian mahasiswa dan mendukung proses pembelajaran di perguruan tinggi.

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

5. Daftar Pustaka

Abidin, N. H. Z., Remli, M. A., Ali, N. M., Phon, D. N. E., Yusoff, N., Adli, H. K., & Busalim, A. H. (2020). Improving intelligent personality prediction using Myers-Briggs type indicator and random forest classifier. *International Journal of Advanced Computer Science and Applications*, *11*(11).

Aini, N., Arif, M., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika. Jurnal Informatika, 11(1), 11–16. https://doi.org/10.31294/inf.v11i1.20637.

Anggraini, S., & Abidin, M. (2025). Sudut pandang individu introvert dan ekstrovert dalam berinteraksi sosial. *JSHP: Jurnal Sosial Humaniora Dan Pendidikan*, *9*(1), 105-117. https://doi.org/10.32487/jshp.v9i1.2156.

Anggraini, S., & Abidin, M. (2025). Sudut pandang individu introvert dan ekstrovert dalam berinteraksi sosial. *JSHP: Jurnal Sosial Humaniora Dan Pendidikan*, *9*(1), 105-117. https://doi.org/10.32487/jshp.v9i1.2156.

Aufa, M. J., & Qoiriah, A. (2023). Analisis Sentimen Pengguna Platform Belajar Online Coursera menggunakan Random Forest dengan Metode Ekstraksi Fitur Word2vec. Journal of Informatics and Computer Science (Jinacs), 244–255. https://doi.org/10.26740/jinacs.v4n02.p244-255.

Azim, F., Sihaloho, D. I. R., Rizki, M., Alghifary, A. H., Habibullah, H. Q., Kusuma, H. J., & Gunawan, R. (2025). Penerapan Random Forest untuk Identifikasi Tipe Kepribadian: Studi Kasus Klasifikasi Introvert dan Extrovert. *Coding: Journal of Computing and Software Engineering*, 66-73. https://doi.org/10.25181/coding.v1i2.4368.

Edman Priyono, S. (2021). *PREDIKSI KEPRIBADIAN PENGGUNA INSTAGRAM BERDASARKAN MBTI MENGGUNAKAN ALGORITMA RANDOM FOREST* (Doctoral dissertation, Universitas International Semen Indonesia).

Erkamim, M., Nurhayati, N., Heriyani, N., & Riyanto, U. Klasifikasi Kepribadian Introvert dan Extrovert Menggunakan Random Forest, Naïve Bayes, dan K-Nearest Neighbor. *Jurnal Ilmiah Fifo*, *17*(2), 202-211. https://doi.org/10.22441/fifo.2025.v17i2.009.

Harahap, R. N., & Muslim, K. (2020). Peningkatan Akurasi Pada Prediksi Kepribadian MBTI Pengguna Twitter Menggunakan Augmentasi Data. *J. Teknol. Inf. dan Ilmu Komput*, *7*(4), 815.

Irawan, I., Qisthiano, M. R., & Jakak, P. M. (2023). Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO. Jurnal Pengembangan Sistem Informasi Dan Informatika, 4(4), 26–35. https://doi.org/10.47747/jpsii.v4i4.1374.

Maulidah, M. (2023). Klasifikasi Kepribadian Menggunakan Algoritma Machine Learning.

Putri, R. A., & Fatonah, N. S. (2022). Perbandingan Metode Klasifikasi serta Analisis Faktor Akademis Pola Kelulusan Mahasiswa di Perguruan Tinggi. Jurnal Informatika Jurnal Pengembangan It, 7(2), 109–117. https://doi.org/10.30591/jpit.v7i2.3082.

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2205

Saleh, J., Adrian, A., & Sanger, J. B. (2022). SISTEM KLASIFIKASI KELULUSAN MAHASISWA DENGAN ALGORITMA RANDOM FOREST. Jurnal Ilmiah Realtech, 18(1), 10–14. https://doi.org/10.52159/realtech.v18i1.4.

Setyanto, A. T. (2023). Deteksi dini prevalensi gangguan kesehatan mental mahasiswa di Perguruan Tinggi. *Wacana*, *15*(1), 66-78. https://doi.org/10.20961/wacana.v15i1.69548.

Sudarmadji, P. W., Fallo, N., & Peli, Y. S. (2023). KOMPARASI ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI KELULUSAN MAHASISWA PROGRAM STUDI TEKNIK KOMPUTER JARINGAN. Jurnal Ilmiah Flash, 8(2), 109. https://doi.org/10.32511/flash.v8i2.998.

Utomo, F. S., Kusuma, V. S., Baihaqi, W. M., & Subarkah, P. (2025). Optimasi Klasifikasi Gaya Belajar Mahasiswa Inklusif Berdasarkan Model VAK dengan Stratified Split dan Multilayer Perceptron. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *12*(5), 1163-1172. https://doi.org/10.25126/jtiik.2025125.

Widodo, S., Ladyani, F., Lestari, S. M. P., Wijayanti, D. R., Devrianya, A., Hidayat, A., ... & Widya, N. (2023). Buku ajar metode penelitian.

Yusuf, B., Qalbi, M., Basrul, B., Dwitawati, I., Malahayati, M., & Ellyadi, M. (2020). IMPLEMENTASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST DALAM MEMPREDIKSI PRESTASI AKADEMIK MAHASISWA UNIVERSITAS ISLAM NEGERI AR-RANIRY BANDA ACEH. Cyberspace Jurnal Pendidikan Teknologi Informasi, 4(1), 50. https://doi.org/10.22373/cj.v4i1.7247.